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

April 28, 2024

Abstract

Given a loss function and a set of policy objectives, how should we evaluate and compare the performances of policy institutions? In this work, we show that it is possible to evaluate policy makers with minimal assumptions on the underlying economic model. The Distance to Minimum Loss —the component of the loss that a policy institution can be held accountable for—, can be computed from well known and estimable sufficient statistics: the impulse responses to policy and non-policy shocks. We use our methodology to evaluate US monetary policy since 1879. We find no material improvement in performance over the first 100 years, and it is only in the last 30 years that we estimate large and uniform improvements in the conduct of monetary policy.

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

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

*We thank our discussant Frank Smets as well as Ryan Chahrour, Vasco Curdia, Davide Debortoli, Thomas Drechsel, John Friedman, Toru Kitagawa, Sylvain Leduc, Karel Mertens, Emi Nakamura, Adam Shapiro, Jón Steinsson, Christian Wolf and seminar participants at CEMFI, CREI, Paris School of Economics, Mannheim, UC Berkeley, Brown, Duke, Cornell, UCL, the NBER Summer Institute (ME, FEM), INSPER, PUC Rio, FGV, Bank of Italy, ECB, Riksbank and the International Francqui Chair Symposium in Brussels for helpful comments. Mesters acknowledges financial support from the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), funded by MCIN/AEI/10.13039/50110001103 and the 2021 ERC Starting Grant POLICYMETRICS. The views expressed in this paper are the sole responsibility of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

1 Introduction

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

A naive approach would consist in measuring performance based on realized macroeconomic outcomes; on the realized value of some loss function. For instance, we could assess a central banker based on average inflation and unemployment outcomes over her term. Unfortunately, that approach suffers from numerous confounding problems, as many factors are outside policy makers' control but affect performance: (i) different policy makers may face different initial conditions upon beginning their term, e.g. a central banker can inherit a strong or weak economy from her predecessor, and this will influence realized outcomes, (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 economies, e.g., a steeper or flatter Phillips curve will affect a central banker's ability to control inflation.

To control for the environment, one approach would be to use a structural model fitted to the data spanning a policy maker's term. Provided that the model is well specified, the structural model can be used to derive an optimal rule; a policy rule that delivers the minimum loss possible *given* the environment. A policy maker is best performing when she follows a rule that delivers that minimum loss, and we can rank policy makers based on their *distance to minimum loss* (DML) —the distance between their realized loss and the minimum loss possible given the environment—. Better performing policy makers will feature smaller distances to minimum loss. A possible risk with this approach however is one of model mis-specification: if the model is mis-specified, performance comparisons can be inaccurate.

In this paper, we show that it is possible to evaluate and rank policy makers with minimal assumptions on the underlying economic model. For a large class of linear forward looking macro models and quadratic loss functions, it is possible to measure the DML from well known and estimable sufficient statistics: the impulse responses (IRs) to policy and non-policy shocks.

To measure the DML with sufficient statistics, the key difficulty is to measure the minimum feasible loss; to characterize the optimal policy rule without having to rely on a specific underlying model. We achieve this thanks to two new results. First, an identification result: knowledge of the optimal reaction to structural shocks alone is sufficient to characterize the optimal policy rule, that is to construct a policy rule that minimizes the loss function given the environment. Second, a sufficient statistics result: the effects of counterfactual reactions to structural shocks can be computed from the IRs to policy and non-policy shocks.

Taken together, these results imply that the optimal reaction function can be characterized from simple regressions in "impulse response space": regressions of the impulse responses to the non-policy shock on the impulse responses to policy shocks. Each regression coefficient measures by how much more or less the policy maker should have responded to a given non-policy shock, and this *optimal reaction adjustment* (ORA)can be used to measure the distance to minimum loss conditional on a specific type of non-policy shock. Overall policy performance —the total distance to minimum loss— can then be computing by aggregating these distances across the different shocks.

With the DML depending only on impulse responses to shocks, the evaluation and comparison of policy makers reduces to a well-known econometric task: the estimation of structural impulse responses, and this realization opens a number of important avenues for policy evaluation, as one can draw on a large macro-econometric literature to evaluate policy institutions. See e.g., Ramey (2016) for a recent discussion of structural shock identification and Stock and Watson (2016), Kilian and Lütkepohl (2017) and Li, Plagborg-Møller and Wolf (2024) for recent approaches and comparisons for impulse response estimation.

Computing the total DML requires the identification of all types of policy and non-policy shocks, and in a dynamic setting, this includes news shocks at all possible horizons. While this data requirement is unlikely to be met in practice, we show how subset statistics, which only use a subset of all possible policy shocks, can be used to evaluate and rank policy makers. Intuitively, this property stems from our characterization of the optimal rule as a set of optimal reaction coefficients to shocks. Since each type of shock can be studied separately from the others, we can split the optimal policy problem into separate problems, and evaluate/compare policy makers separately for each type of shock.

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

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

Evaluating and comparing policy makers requires to take a stand on a set of objectives, i.e., on a loss function. In our empirical application, we consider a quadratic loss function with equal weights on inflation and unemployment.[1] Given that loss function, our main results are as follows: (i) we estimate large and uniform improvements in the conduct of monetary policy, but *only* in the last 30 years, (ii) we cannot reject that the Fed's reaction to recent financial shocks (notably the 2007-2008 financial crisis) was appropriate, in contrast to the "highly" sub-optimal reaction of the Fed during the Great Depression, (iii) despite much larger realized losses in the 1920s-1930s, the performance of the early Fed is no worse than the performance of the passive Gold Standard, and (iv) the Fed's reaction function during the 1960s-1970s is almost as sub-optimal as the reaction function of the early Fed.

Related literature

An early contribution is Fair (1978) who highlights the distortions stemming from different initial conditions and economic environments. His solution was to adopt optimal control methods to compare policy makers through the lens of a fully specified model. Modern versions of this approach include (e.g. Galı, López-Salido and Vallés, 2003; Gali and Gertler, 2007; Blanchard and Galí, 2007). Unfortunately, specifying the correct model for (i) the policy rule and (ii) the macroeconomic non-policy block is a very difficult task (e.g., Svensson,

[1] Our approach can accommodate other loss functions, for instance different loss functions across time periods, or even micro-founded welfare-based loss functions.

2003; Mishkin, 2010). A less structural approach has studied monetary performance through the lens of estimated policy rules —requiring only the specification of a policy rule—.[2] In particular, a number of studies compared the Fed in the pre- and post-Volcker periods by assessing whether the Taylor principle was satisfied. However, beyond the Taylor principle, that approach can say little about the optimality of reaction functions, and thus can only provide a coarse evaluation of reaction functions.

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

Closer to our work, the literature has proposed reduced-form methods to study policy rule counterfactuals (e.g., Sims and Zha, 2006; Bernanke et al., 1997; Leeper and Zha, 2003), though these approaches are not fully robust to the Lucas critique. Instead, our approach builds on recent work showing that robustness to the Lucas critique is possible in a large class of macroeconomic models (McKay and Wolf, 2023). When the coefficients of the non-policy block are independent of the coefficients of the policy block, it is possible to reproduce any policy rule counterfactual with an appropriate combination of policy news shocks at different horizons. Our work exploits a little studied yet attractive class of policy rule counterfactuals —counterfactual reactions to non-policy shocks—, which have some appealing properties: (i) the class is sufficient to characterize the optimal reaction function starting from any baseline reaction function and (ii) the class allows to split the optimal policy problem into computationally simple separate problems, allowing to evaluate/compare policy makers under subset identification.

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

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

that a sufficient statistics approach to rule evaluation is possible, but it requires a different set of statistics, and notably additional identifying restrictions: the identification of (at least some) non-policy shocks.

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

2 Illustrative example

Before formally describing our general framework, we first illustrate how it is possible to evaluate and compare policy makers without having access to the underlying economic model nor the policy rule. To illustrate the method, 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

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

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

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.[3] The parameters are collected in $\theta = (\kappa, \sigma, \sigma_\xi)'$. We can think of θ as capturing the economic "environment".

The policy maker sets the interest rate following the rule

$$i_t = \phi_\pi \pi_t + \sigma_\epsilon \epsilon_t \ , \tag{3}$$

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

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

unit variance.[4]

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.

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

The vectors Γ and $\mathcal{R}$ capture the impulse responses of the policy objectives to the structural shocks. Note that Γ and $\mathcal{R}$ depend on the environment parameters θ and the policy rule parameters ϕ, which include the standard deviations of the structural shocks.

Evaluating policy makers requires taking a stance on a performance metric. To that effect, we consider the unconditional loss function

$$\mathcal{L}(\phi;\theta) = \mathbb{E}Y_t^{'}Y_t \ , \qquad \text{which using (4) becomes} \ \ \mathcal{L}(\phi;\theta) = \Gamma'\Gamma + \mathcal{R}'\mathcal{R} \ . \tag{5}$$

Given this loss function, an "optimal reaction function" is defined as any ϕ that minimizes $\mathcal{L}(\phi;\theta)$ given the underlying structure of the economy, i.e., given equations (1)-(3). An optimal reaction function can thus be seen as a policy rule that best stabilizes (minimizes the sum-of-squares) the impulse responses to shocks.

In this example the optimal reaction function is unique and given by $\phi^{\text{opt}} = (\phi_\pi^{\text{opt}}, \sigma_\epsilon^{\text{opt}})' = (\kappa\sigma, 0)'$ (e.g. Galí, 2015). First, exogenous policy changes are not optimal, and an optimal policy features no policy shocks ($\sigma_\epsilon^{\text{opt}} = 0$, which implies $\mathcal{R} = 0$). Second, the optimal reaction coefficient ϕ_π^{opt} is the coefficient that minimizes the effects of cost-push shocks, i.e., that best stabilizes Γ. The minimum loss is then given by

$$\mathcal{L}^{\text{opt}} = \Gamma^{\text{opt}'}\Gamma^{\text{opt}} \ , \qquad \text{with} \qquad \Gamma^{\text{opt}} \equiv \Gamma(\phi^{\text{opt}}, \theta) \ , \tag{6}$$

capturing the minimal effect of cost-push shocks that a policy maker can achieve on average given the environment θ.

A naive approach to policy evaluation

Consider a policy maker with reaction function ϕ^0 during her term and associated loss $\mathcal{L}^0 = \mathcal{L}(\phi^0;\theta)$. How should we evaluate that policy maker?

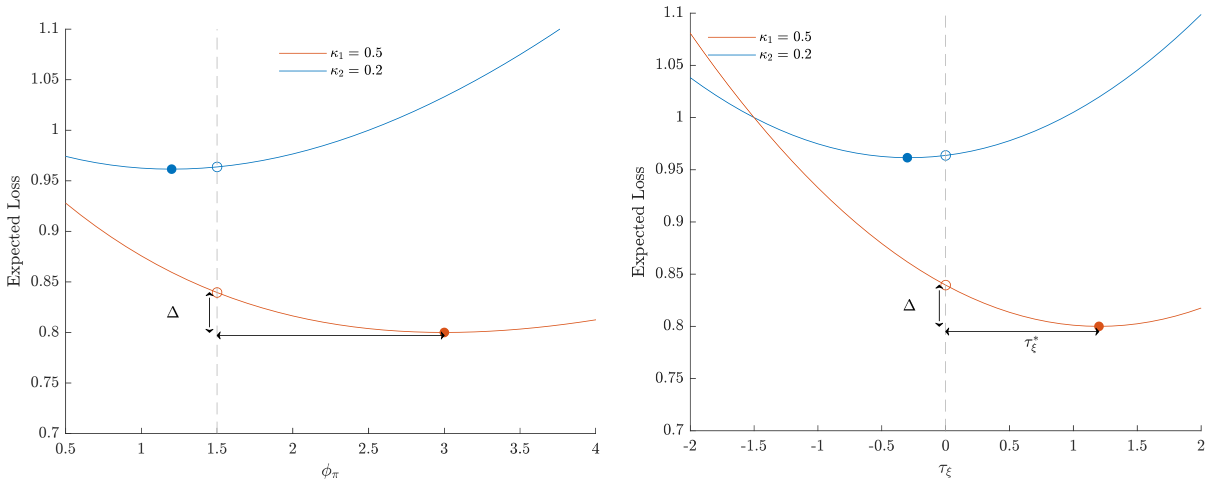
A naive approach would consist in comparing realized losses. Specifically, (i) compute the average loss during a policy maker's term, which provides an estimate of the loss $\mathcal{L}^0$,

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

and (ii) evaluate and rank policy makers based on that estimate. Policy makers with higher average loss would then be deemed less performant.

Unfortunately, the parameter vector θ that describes the economic environment is outside the control of the policy maker and acts as a confounder; by influencing the impulse responses to shocks and thus the loss, see (4) and (5).

Figure 1: Loss based Policy Evaluation



(a) Naive loss-based evaluation

(b) Using auxiliary loss to measure Δ

Notes: The optimal rule ϕ_π^{opt} is indicated by the filled dots. The empty dots indicated the policy maker's policy rule. The left panel depicts the original loss function $\mathcal{L}((\phi_\pi, 0); \theta)$, the right panel depicts the auxiliary loss function $\mathcal{L}((\tau_\xi, 0), ; (\phi_\pi^0, 0), \theta)$. Note how the levels of both loss functions are identical in two points: (i) at the baseline reaction function and (ii) at the optimum.

To give a concrete example of how the economic environment can distort policy evaluation, consider two policy makers —Red and Blue— following the same rule $\phi_\pi = 1.5$ but operating in different environments: one with a steeper Phillips curve ($\kappa = 0.5$) and the other with a flatter Phillips curve ($\kappa = 0.2$). Figure 1a plots the expected loss function as a function of the parameter ϕ_π for these two policy makers.[5]

The expected loss under Red —the policy maker under the steep Phillips curve— is lower than under Blue —the policy maker under the flat Phillips curve—: the empty dot is lower for Red than for Blue. A naive approach to policy evaluation would thus conclude that Red is not a better policy maker than Blue. However, it's the exact opposite: in this example, Red is further away from the optimal reaction function (filled dot) than Blue. In other words, Red performs less well than Blue. The reason for these different conclusions is the underlying environment: in the steep Phillips curve world of Red, it is *easier* to achieve a lower loss. No matter the value ϕ_π that Red picks, she will always face a lower loss than Blue.

[5]As illustrative calibration, we take $\beta = .99$, $\sigma = 1$, $\sigma_\epsilon = 0$ and $\sigma_\xi = 1$.

To properly compare Red and Blue, we must thus take into account the environment, or in other words, measure how far is a policy maker from the minimum loss possible *given* the environment. This is the *Distance to Minimum Loss* (DML) defined as

$$\Delta = \mathcal{L}^0 - \mathcal{L}^{\text{opt}} \ , \tag{7}$$

which is depicted in Figure 1a: the distance between the actual expected loss and the feasible minimum loss.

A sufficient statistics approach to policy evaluation

To measure the DML, one possible approach consists in specifying a structural model, fit that model to the data spanning the policy maker's term and then compute the minimum feasible loss $\mathcal{L}^{\text{opt}}$ from that model. In the context of the example this amounts to specifying the Phillips and IS curves and estimating the associated parameters θ. The risks with this approach are (i) model mis-specification: if the model does not capture the full complexity of the underlying environment, the policy assessment can be compromised, and (ii) identification: the parameters of forward looking macro models are typically hard to identify implying large confidence intervals (e.g. Canova and Sala, 2009).

In this paper, we propose a different approach to measure $\mathcal{L}^{\text{opt}}$ and Δ, an approach that requires minimal assumptions on the underlying economic model.

A class of policy rule counter-factuals

To measure the minimum attainable loss $\mathcal{L}^{\text{opt}}$, we characterize the optimal reaction function ϕ^{opt} in a different way. Instead of minimizing the loss with respect to the reaction coefficients in front of endogenous variables (as is common in the literature, e.g., Galí (2015)), we propose to optimize with respect to the reaction coefficients in front of structural shocks. While this class of rule counterfactuals could seem of little direct interest, they have two important, yet overlooked, properties: (i) the effects of counterfactual reaction to structural shocks can be computed with minimal assumptions on the underlying model, depending only on estimable sufficient statistics, and (ii) the optimal reaction coefficients to structural shocks is sufficient to fully characterize the optimal policy rule, and thus to compute the minimum attainable loss.

To see that, recall that $\phi^0 = (\phi^0_\pi, \sigma^0_\epsilon)$ is the policy maker's reaction function and consider the policy rule counter-factual

$$i_t = \phi^0_\pi \pi_t + \underbrace{\sigma^0_\epsilon(\tau_\xi \xi_t + \tau_\epsilon \epsilon_t)}_{\text{Reaction adjustment}} + \sigma^0_\epsilon \epsilon_t \ , \tag{8}$$

where $\tau = (\tau_\xi, \tau_\epsilon)'$ is a vector of responses to structural shocks. Unlike the original reaction function (3), the modified reaction function (8) fixes the reaction coefficients ϕ^0 at their baseline value.

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

$$Y_t = (\Gamma^0 + \mathcal{R}^0 \tau_\xi)\xi_t + (\mathcal{R}^0 + \mathcal{R}^0 \tau_\epsilon)\epsilon_t \ , \tag{9}$$

where $\Gamma^0 = \Gamma(\phi^0, \theta)$ and $\mathcal{R}^0 = \mathcal{R}(\phi^0, \theta)$.

From expression (9), we can see that $\Gamma^0 + \mathcal{R}^0 \tau_\xi$ is the impulse response to cost-push shocks *after* the reaction function adjustment τ_ξ. In other words, the adjustment τ_ξ modifies the impulse response to cost-push shocks from Γ^0 to $\Gamma^0 + \mathcal{R}^0 \tau_\xi$, which means that the "old" impulse responses Γ^0 and $\mathcal{R}^0$ are all we need to compute the effects of the policy rule counter-factual (8).

Optimal reaction adjustment

From (9), we can use Γ^0 and $\mathcal{R}^0$ to search for the optimal reaction coefficient to structural shocks. To that effect, it is helpful to define an auxiliary loss function that takes τ, the vector of reaction coefficients to shocks, as its arguments and holding ϕ^0 fixed.

$$\begin{aligned} \mathsf{L}(\tau; \phi^0, \theta) &= \mathbb{E} Y_t^{'} Y_t \\ &= \mathsf{L}_\xi(\tau_\xi; \phi^0, \theta) + \mathsf{L}_\epsilon(\tau_\epsilon; \phi^0, \theta) \ , \end{aligned} \tag{10}$$

with

$$\mathsf{L}_\xi(\tau_\xi; \phi^0, \theta) = \left(\Gamma^0 + \mathcal{R}^0 \tau_\xi\right)' \left(\Gamma^0 + \mathcal{R}^0 \tau_\xi\right) \quad \text{and} \quad \mathsf{L}_\epsilon(\tau_\epsilon; \phi^0, \theta) = \left(\mathcal{R}^0 + \mathcal{R}^0 \tau_\epsilon\right)' \left(\mathcal{R}^0 + \mathcal{R}^0 \tau_\epsilon\right) \ .$$

The first component L_ξ captures how a reaction adjustment τ_ξ affects the loss through its effect on the impulse responses to cost-push shock. Similarly, the second component L_ϵ captures how a reaction adjustment τ_ϵ affects the loss through its effect on the impulse responses to policy shocks.

Solving for the optimum reaction adjustment

$$\tau^* = \underset{\tau}{\operatorname{argmin}} \, \mathsf{L}(\tau; \phi^0, \theta) \ ,$$

we get

$$\tau_\xi^* = -(\mathcal{R}^{0'} \mathcal{R}^0)^{-1} \mathcal{R}^{0'} \Gamma^0 \qquad \text{and} \qquad \tau_\epsilon^* = -1 \ . \tag{11}$$

The statistic τ^* is the Optimal Reaction Adjustment (ORA):[6] (i) it adjusts the reaction to non-policy shocks ξ_t in order to minimize their effects, i.e., to reach Γ^{opt}; the minimal effect of cost-push shock that a policy maker can achieve given the environment θ, and (ii) it cancels monetary mistakes by setting policy shocks back to zero with $\tau^*_\epsilon = -1$.

Distance to minimum loss

While the ORA itself may not be of direct interest to policy makers—the ORA is a reaction coefficient to unobserved structural shocks instead of endogenous variables—, its importance comes from its one key property: it allows to fully characterize the optimal reaction function and thereby compute the minimum loss.

Indeed, we have[7]

$$\mathsf{L}(\tau^*; \phi^0, \theta) = \mathcal{L}^{\text{opt}} \ . \tag{12}$$

This means that the auxiliary loss function $\mathsf{L}(\tau; \phi^0, \theta)$ has the same minimum as the original loss function $\mathcal{L}(\phi; \theta)$.

Figure 1b traces out the auxiliary loss function (10) as a function of τ_ξ. While the auxiliary loss function is different from the original loss function $\mathcal{L}(\phi; \theta)$ (Figure 1a), note how the two functions coincide in two points: First, by construction when $\tau = 0$ the auxiliary loss function coincides with the original loss function taken at the policy makers choice ϕ^0. Second, and this is key, the auxiliary loss function has the same minimum as the original loss function.

From those two points, we can compute the distance to minimum loss Δ based on

$$\begin{aligned} \Delta &= \mathcal{L}^0 - \mathcal{L}^{\text{opt}} \\ &= \Delta_\xi + \Delta_\epsilon \ , \end{aligned} \tag{13}$$

[6] Note how τ^*_ξ is the coefficient of a regression of Γ^0 on $-\mathcal{R}^0$; a regression in impulse response space. Intuitively, Γ^0 (the impulse response to a cost-push shock) captures what the policy maker *did* on average to counteract cost-push shocks with his rule ϕ^0, while $\mathcal{R}^0$ (the impulse response to a monetary shock) captures what the policy maker *could have done* to counteract cost-push shocks —how reacting to ξ_t by τ could have better stabilized the effect of cost-push shocks by transforming Γ^0 into $\Gamma^0 + \tau\mathcal{R}^0$, see (9)—. A regression on $\mathcal{R}^0$ on Γ^0 precisely finds the τ that minimizes the sum-of-squares of that adjusted impulse response, i.e., that best cancels out the effects of non-policy shock. At an optimal policy rule, Γ^0 and $\mathcal{R}^0$ should be orthogonal.

[7] To see this plug in τ^* into the auxiliary loss function to obtain

$$\begin{aligned} \mathsf{L}(\tau^*; \phi^0, \theta) &= \Gamma^{0'}(I - \mathcal{R}^0(\mathcal{R}^{0'}\mathcal{R}^0)^{-1}\mathcal{R}^{0'})\Gamma^0 \\ &= \frac{\sigma^2_\xi}{1+\kappa^2} = \mathcal{L}(\phi^{\text{opt}}; \theta) \ , \end{aligned}$$

using the expressions for Γ^0 and $\mathcal{R}^0$.

where

$$\Delta_\xi = \Gamma^{0\prime}\mathcal{R}^0\left(\mathcal{R}^{0\prime}\mathcal{R}^0\right)^{-1}\mathcal{R}^{0\prime}\Gamma^0 \qquad \text{and} \qquad \Delta_\epsilon = \mathcal{R}^{0\prime}\mathcal{R}^0 \ . \tag{14}$$

The expression shows that the DML can be computed from the impulse responses to policy and non-policy shocks, i.e. $\mathcal{R}^0$ and Γ^0.

Figure 1 summarizes the main idea underlying our approach. Looking at both panels, the DML can be computed either from the original loss function or from the auxiliary loss function. By going through the auxiliary loss function (i.e., by studying how the loss depends on the policy maker's systematic reaction to structural shocks), we can compute the DML with minimal assumptions on the underlying model, as the derivatives of the auxiliary loss function depend only on estimable sufficient statistics: the impulse responses $\mathcal{R}^0$ and Γ^0. Figure 1b also illustrates how the ORA and DML can be seen as two sides of the same coin —two complementary ways to evaluate a policy maker—. The ORA measures how far is a policy maker's reaction coefficient from the optimal reaction coefficient, while the DML captures the "welfare" consequences of that sub-optimal reaction coefficient.

Case of multiple non-policy shocks

Before generalizing to a generic class of dynamic macro models, we briefly mention the case of multiple non-policy shocks. To that effect, consider adding a preference shock ν_t to the IS curve (2). With two types of non-policy shocks (ξ_t and ν_t), the minimum loss can be attained by a policy rule of the form $i_t = \phi_\pi \pi_t + \phi_x x_t + \epsilon_t$ (e.g. Galí, 2015). Proceeding exactly as above, we can compute the distance to minimum loss from the impulse responses to policy shocks ($\mathcal{R}$) and non-policy shocks (Γ_ξ and Γ_ν) and we have

$$\Delta = \Delta_\xi + \Delta_\nu + \Delta_\epsilon \ ,$$

where Δ_ξ and Δ_ν have the same expression as (14) but for $\Gamma = \Gamma^0_\xi$ and $\Gamma = \Gamma^0_\nu$, respectively. This illustrates one additional property of our DML decomposition: the total distance to minimum loss Δ is a sum of sub-distances to minimum loss, with each sub-distance capturing the distance to the minimum loss in response to a *different* type of structural shock, and each sub-distance (e.g., Δ_ξ or Δ_ν) can be computed independently of the other.

In sum, this example illustrates how we can (i) evaluate and (ii) compare policy makers based on their DML 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 general linear forward looking macro models.

3 Environment

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

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

$$\begin{array}{rcl} \mathcal{A}_{yy}\mathbf{Y} - \mathcal{A}_{yp}\mathbf{P} & = & \mathcal{B}_{y\xi}\boldsymbol{\Xi} \\ \mathcal{A}_{pp}\mathbf{P} - \mathcal{A}_{py}\mathbf{Y} & = & \mathcal{B}_{p\xi}\boldsymbol{\Xi} + \mathcal{B}_{p\epsilon}\boldsymbol{\epsilon} \end{array} , \tag{15}$$

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

We normalize all elements of $\boldsymbol{\Xi}$ and $\boldsymbol{\epsilon}$ to have mean zero and unit variance.[8] Also, we assume that they are serially and mutually uncorrelated, consistent with the common definition of structural shocks (e.g. Bernanke, 1986; Ramey, 2016).[9] 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 $\theta = \{\mathcal{A}_{yy}, \mathcal{A}_{yp}, \mathcal{B}_{y\xi}\}$ which the policy maker takes as given, and the reaction function $\phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{B}_{p\xi}, \mathcal{B}_{p\epsilon}\}$, which is under the control of the policy maker and we assume that $\mathcal{B}_{p\epsilon}$ is invertible. Further, 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 (15) implies a unique equilibrium, that is all ϕ for which

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

Many structural models found in the literature can be written in the form of (15); prominent

[8]The unit variance is without loss of generality as the diagonal elements of $\mathcal{B}_{..}$ are unrestricted.

[9]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 $\mathcal{B}_{p\epsilon}$ such that the equation residuals —the shocks— are uncorrelated. For example if $\text{var}(\boldsymbol{\Xi}) = \Sigma$, then redefine $\mathcal{B}_{y\xi}\boldsymbol{\Xi} = \tilde{\mathcal{B}}_{y\xi}\tilde{\boldsymbol{\Xi}}$ with $\tilde{\mathcal{B}}_{y\xi} = \mathcal{B}_{y\xi}\Sigma^{1/2}$ and $\tilde{\boldsymbol{\Xi}} = \Sigma^{-1/2}\boldsymbol{\Xi}$ such that $\tilde{\boldsymbol{\Xi}}$ is serially uncorrelated. The same can be done for $\mathcal{B}_{p\xi}\boldsymbol{\Xi}$ or $\mathcal{B}_{p\epsilon}\boldsymbol{\epsilon}$.

examples include New Keynesian models and heterogeneous agents models, see McKay and Wolf (2023) for a more in depth discussion.

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

Lemma 1. *Given the generic model* (15) *with* $\phi \in \Phi$*, we have*

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

The maps $\Gamma(\phi, \theta)$ and $\mathcal{R}(\phi, \theta)$ capture the causal effects of the structural shocks $\boldsymbol{\Xi}$ and $\boldsymbol{\epsilon}$ on the non-policy variables. Explicit characterizations for $\Gamma(\phi, \theta)$ and $\mathcal{R}(\phi, \theta)$ are given in the appendix. Note the similarity between (16) and (4), as the illustrative static NK model is a special case with only contemporaneous shocks.

Lemma 1 implies that the identification of the impulse responses requires observing part of the *future* shocks in $\boldsymbol{\Xi}$ and $\boldsymbol{\epsilon}$. Our perfect foresight notation masks this requirement, but it is useful to clarify that in practice this requires the identification of news shocks. To see this, note that we can decompose ξ_t and ϵ_t as[10]

$$\xi_t = \sum_{j=0}^{t} \underbrace{\mathbb{E}_j \xi_t - \mathbb{E}_{j-1}\xi_t}_{\xi_{t,j}} \qquad \text{and} \qquad \epsilon_t = \sum_{j=0}^{t} \underbrace{\mathbb{E}_j \epsilon_t - \mathbb{E}_{j-1}\epsilon_t}_{\epsilon_{t,j}} \ , \tag{17}$$

where $\mathbb{E}_j(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_j)$, with $\mathcal{F}_j$ the information set available at time j. The increment $\xi_{t,j} \equiv \mathbb{E}_j\xi_t - \mathbb{E}_{j-1}\xi_t$ is the component of ξ_t that is released at time $j \leq t$. In other words $\xi_{t,j}$ is a news shock released at $j \leq t$, and (17) decomposes the shock ξ_t —a shock realized at time t— as a sum of news shocks $\xi_{t,j}$ revealed all the way until time t with $\xi_t = \sum_{j=0}^{t} \xi_{t,j}$. Similarly for $\epsilon_{t,j}$. By construction the news shocks are serially uncorrelated.

Thus, to identify the impulse responses in (16), we require observing proxies for the news shocks in $\boldsymbol{\xi}_0 = (\xi_{0,0}, \xi_{1,0}, \xi_{2,0}, \ldots)'$ and $\boldsymbol{\epsilon}_0 = (\epsilon_{0,0}, \epsilon_{1,0}, \epsilon_{2,0}, \ldots)'$.[11] For notational convenience we drop the zero subscript and work under perfect foresight.

Evaluation criteria

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

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

[11] Note that in practice our approach will not require the identification of all news shocks.

We measure performance using the loss function

$$\mathcal{L}(\phi;\theta) = \mathbb{E}\mathbf{Y}'\mathcal{W}\mathbf{Y} \ , \tag{18}$$

where $\mathcal{W}$ is a diagonal matrix, with non-negative entries, which selects and weights the specific variables and horizons that are part of the researcher's evaluation criteria. The loss (18) is the researcher's evaluation criterion for scoring policy maker performance —an input into our framework—.

In terms of the impulse responses of Lemma 1 the loss function can be written as

$$\mathcal{L}(\phi;\theta) = \mathrm{Tr}(\Gamma(\phi,\theta)'\mathcal{W}\Gamma(\phi,\theta)) + \mathrm{Tr}(\mathcal{R}(\phi,\theta)'\mathcal{W}\mathcal{R}(\phi,\theta)) \ , \tag{19}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace operator that takes the sum of the diagonal elements of the maps.

The actions of the policy maker are summarized by the reaction function ϕ. We define a reaction function to be optimal if it minimizes the loss function (18). Formally, the set of optimal reaction functions is given by

$$\Phi^{\mathrm{opt}} = \left\{ \phi : \phi \in \operatorname*{argmin}_{\phi\in\Phi} \mathcal{L}(\phi;\theta) \quad \text{s.t} \quad (15) \right\} \ . \tag{20}$$

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

For a given $\phi^{\mathrm{opt}} \in \Phi^{\mathrm{opt}}$ the minimum loss given the environment θ is given by

$$\mathcal{L}^{\mathrm{opt}} = \mathcal{L}(\phi^{\mathrm{opt}};\theta) = \mathrm{Tr}(\Gamma(\phi^{\mathrm{opt}},\theta)'\mathcal{W}\Gamma(\phi^{\mathrm{opt}},\theta)) \ , \tag{21}$$

as $\mathcal{B}^{\mathrm{opt}}_{p\epsilon} = 0$ canceling the second term in (19).

4 Policy evaluation with sufficient statistics

In this section, we show how we can evaluate a policy maker after her term by measuring the distance between the loss under her reaction function, denoted by ϕ^0, and the loss under the optimal reaction function. To that effect, we define the *Distance to Minimum Loss* (DML) as

$$\Delta = \mathcal{L}^0 - \mathcal{L}^{\mathrm{opt}} \ , \tag{22}$$

where $\mathcal{L}^0 = \mathcal{L}(\phi^0;\theta)$ is the policy maker's loss.

4.1 Computing the distance to minimum loss

Following the same steps as the simple example of Section 2, we propose to measure the DML by considering a thought experiment where we adjust the policy maker's reaction coefficients to the structural shocks. Specifically, consider the auxiliary reaction function

$$\mathcal{A}^0_{pp}\mathbf{P} - \mathcal{A}^0_{py}\mathbf{Y} = \mathcal{B}^0_{p\xi}\boldsymbol{\Xi} + \mathcal{B}^0_{p\epsilon}\boldsymbol{\epsilon} + \mathcal{B}^0_{p\epsilon}(\mathcal{T}_\xi\boldsymbol{\Xi} + \mathcal{T}_\epsilon\boldsymbol{\epsilon}) \ , \tag{23}$$

where $\mathcal{T} = (\mathcal{T}_\xi, \mathcal{T}_\epsilon)$ adjusts the response to the structural shocks $\boldsymbol{\Xi}$ and $\boldsymbol{\epsilon}$.[12] Each element of $\mathcal{T}$ corresponds to a different rule counterfactual, in which we modify how one element of the policy path responds to one of the shocks. Note that similar as in the simple example (i.e. equation (8)) all responses are scaled by the variance of policy shocks, here $\mathcal{B}^0_{p\epsilon}$, to ensure that the scaling of the adjustments is similar as for the original irfs.

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

Lemma 2. *Consider the generic model* (15) *with* $\phi^0 \in \Phi$ *and the modified policy rule* (23). *We have*

$$\mathbf{Y} = (\Gamma^0 + \mathcal{R}^0\mathcal{T}_\xi)\boldsymbol{\Xi} + (\mathcal{R}^0 + \mathcal{R}^0\mathcal{T}_\epsilon)\boldsymbol{\epsilon} \ , \tag{24}$$

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

The reaction adjustment $\mathcal{T}$ affects the equilibrium by changing the impulse responses to non-policy shocks from Γ^0 to $\Gamma^0 + \mathcal{R}^0\mathcal{T}_\xi$ and the impulse responses to policy shock from $\mathcal{R}^0$ to $\mathcal{R}^0 + \mathcal{R}^0\mathcal{T}_\epsilon$, so that knowledge of the impulse response matrix $\mathcal{R}^0$ is sufficient to compute the policy rule counterfactuals embedded in the $\mathcal{T}$ adjustments. This property echoes the general result of McKay and Wolf (2023), who show that it is possible to reproduce any policy rule counterfactual with an appropriate combination of policy news shocks at different horizons.[13]

We now define the auxiliary loss function

$$\mathsf{L}(\mathcal{T}; \phi^0, \theta) = \mathbb{E}\mathbf{Y}'\mathcal{W}\mathbf{Y} \qquad \text{with} \quad \mathbf{Y} = (\Gamma^0 + \mathcal{R}^0\mathcal{T}_\xi)\boldsymbol{\Xi} + (\mathcal{R}^0 + \mathcal{R}^0\mathcal{T}_\epsilon)\boldsymbol{\epsilon} \ , \tag{25}$$

which allows to trace out how changing the reaction to any individual shock affects the loss.

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

[13]The derived counterfactual is robust to the Lucas critique provided that the coefficients of the macro block (here θ) are invariant to changes in the coefficients of the policy rule (here, ϕ). This property holds in most modern macro models as in our generic model (15).

The *Optimal Reaction Adjustment* (ORA) is the adjustment that minimizes the auxiliary loss function, i.e. $\mathcal{T}^* = \text{argmin}_{\mathcal{T}} \mathsf{L}(\mathcal{T};\phi^0,\theta)$, and we have

$$\mathcal{T}^* = [\mathcal{T}_\xi^*, \mathcal{T}_\epsilon^*] \qquad \text{with} \qquad \mathcal{T}_\xi^* = -(\mathcal{R}^{0\prime}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0\prime}\mathcal{W}\Gamma^0 \quad \text{and} \quad \mathcal{T}_\epsilon^* = -\mathbf{I} \, , \tag{26}$$

where $\mathbf{I}$ is the identity map. Note that the ORA has the same geometric interpretation as discussed in the simple example, and $\mathcal{T}_\xi^*$ is equal to the (weighted) least-square regression of the selected non-policy impulse responses Γ^0 on the selected policy impulse responses $\mathcal{R}^0$. The only difference is the weighting matrix $\mathcal{W}$, which is merely a selection tool used to select the variables that comprise the researcher's evaluation criteria.

Using the ORA we can establish the following key result for the distance to minimum loss defined in (22).

Proposition 1. *Given the generic model* (15)*, with* $\phi^0 \in \Phi$*, we have that*

1. $\mathsf{L}(\mathcal{T}^*;\phi^0,\theta) = \mathcal{L}^{\text{opt}}$
2. *The DML statistic* $\Delta = \mathcal{L}^0 - \mathcal{L}^{\text{opt}}$ *is given by*

$$\Delta = \Delta_\xi + \Delta_\epsilon \ ,$$

where

$$\Delta_\xi = \text{Tr}(\Gamma^{0\prime}\mathcal{W}\mathcal{R}^0(\mathcal{R}^{0\prime}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0\prime}\mathcal{W}\Gamma^0) \qquad \textit{and} \qquad \Delta_\epsilon = \text{Tr}(\mathcal{R}^{0\prime}\mathcal{W}\mathcal{R}^0) \ .$$

The first part of the proposition states that the auxiliary loss function, when evaluated at the ORA statistic, attains the minimum loss as defined in (21). This is our identification result: knowledge of the optimal reaction to the different structural shocks is sufficient to fully characterize the optimal policy rule and thus to compute the minimum attainable loss $\mathcal{L}^{\text{opt}}$. The second part of the proposition states that the DML —the distance between $\mathcal{L}^0$ and $\mathcal{L}^{\text{opt}}$— can be computed from sufficient statistics alone:—the impulse responses $\mathcal{R}^0$ and Γ^0 to policy and non-policy shocks—.

4.2 Policy evaluation with subset shock identification

Identifying the DML requires the identification of all the elements of $\mathcal{R}^0$ and Γ^0 which in turn requires to identify all the different types of non-policy shocks that affected the economy (including news shocks) as well as all the different policy shocks (including news shocks). In practice, this may not be possible, and a researcher may only be able to estimate a subset of all policy and non-policy shocks. Fortunately, it is possible to evaluate policy makers even with subset shock identification.

To formalize this, let $\boldsymbol{\epsilon}_a$ denote any subset of $\boldsymbol{\epsilon}$ and $\boldsymbol{\Xi}_b$ denote any subset of $\boldsymbol{\Xi}$. To measure the corresponding subset distance to minimum loss, we proceed similarly to Section 4.1 and modify a subset of the reaction coefficients to the structural shocks. That is we consider the augmented policy rule

$$\mathcal{A}^0_{pp}\mathbf{P} - \mathcal{A}^0_{py}\mathbf{Y} = \mathcal{B}^0_{p\xi}\boldsymbol{\Xi} + \mathcal{B}^0_{p\epsilon}\boldsymbol{\epsilon} + \mathcal{B}^0_{p\epsilon_a}(\mathcal{T}_{\xi,ab}\boldsymbol{\Xi}_b + \mathcal{T}_{\epsilon,aa}\boldsymbol{\epsilon}_a) \ , \tag{27}$$

where now only the responses to the identifiable shocks are adjusted by $\mathcal{T}_{ab} = (\mathcal{T}_{\xi,ab}, \mathcal{T}_{\epsilon,as})$. Proceeding exactly as with Proposition 1, we can derive a subset ORA statistic $\mathcal{T}^*_{ab} = \text{argmin}_{\mathcal{T}_{ab}} \mathsf{L}(\mathcal{T}_{ab}; \phi^0, \theta)$ and derive a closed-form expression for the sub-distance to minimum loss $\Delta_{ab} = \mathcal{L}^0 - \mathsf{L}(\mathcal{T}^*_{ab}; \phi^0, \theta)$ with

$$\Delta_{ab} = \Delta_{\xi,ab} + \Delta_{\epsilon,aa} \tag{28}$$

where $\Delta_{\xi,ab} = \text{Tr}(\Gamma^{0\prime}_b \mathcal{W}\mathcal{R}^0_a(\mathcal{R}^{0\prime}_a\mathcal{W}\mathcal{R}^0_a)^{-1}\mathcal{R}^{0\prime}_a\mathcal{W}\Gamma^0_b)$ and $\Delta_{\epsilon,aa} = \text{Tr}(\mathcal{R}^{0\prime}_a\mathcal{W}\mathcal{R}^0_a)$, where $\mathcal{R}_a$ denotes the subset of $\mathcal{R}$ corresponding to the policy shocks $\boldsymbol{\epsilon}_a$ and Γ_b denotes the subset of Γ corresponding to the non-policy shocks $\boldsymbol{\Xi}_b$. Trivially, we have $\Delta_{ab} \leq \Delta$, with $\Delta_{ab} = \Delta$ in the limit where all shocks are identified.

The sub-distance $\Delta_{\xi,ab}$ captures the distance between the loss under the baseline policy rule and the minimum loss for a policy maker optimizing the reactions of her a policy instruments to the b non-policy shocks. A subset policy evaluation is then based on how well a policy maker used a specific set of instruments (captured by a) to handle of specific set of non-policy shock (captured by b). The sub-distance Δ_{ab} provides a summary measure of overall performance for the subset of shocks we could identify, and each individual element of $\mathcal{T}_{\xi,ab}$ has an economic interpretation; corresponding to a specific action behind a suboptimal policy, i.e., a specific coefficient in the policy rule.[14]

There is one caveat to subset evaluation however: with subset identification, a policy maker is evaluated in specific "directions" that may or may not correspond to the most relevant margins of improvement. In other words, a subset policy evaluation could be limited in scope.

To guard against this possibility and to get a sense of the "exhaustivity" of a subset-based assessment, we can place bounds on Δ_a, the *total* distance to minimum loss for the policy

[14]To give a concrete example, say we identified contemporaneous shocks to the central bank's policy rate and contemporaneous shocks to oil prices. Then the corresponding ORA entry will assess how the policy maker used the contemporaneous policy rate to minimize the effects of contemporaneous oil shocks: a shock-specific policy evaluation.

instruments captured by subset a, i.e.

$$\Delta_a = \Delta_{\xi,a} + \Delta_{\epsilon,aa} \ , \qquad \text{with} \qquad \Delta_{\xi,a} = \text{Tr}(\Gamma^{0\prime}\mathcal{W}\mathcal{R}_a^0(\mathcal{R}_a^{0\prime}\mathcal{W}\mathcal{R}_a^0)^{-1}\mathcal{R}_a^{0\prime}\mathcal{W}\Gamma^0) \ . \tag{29}$$

For instance, if the set a only includes contemporaneous policy shocks, the distance Δ_a measures how well the policy maker used its contemporaneous policy instrument in response to *all* the non-policy shocks that could affect the economy.

We have the following proposition

Proposition 2 (DML bounds)**.** *Given the generic model* (15)*, with* Φ *non-empty, we have*

$$\Delta_{ab} \leq \Delta_a \leq \Delta_{ab} + \mathcal{E}_{ab}^0 \ , \tag{30}$$

where

$$\mathcal{E}_{ab}^0 \equiv \mathcal{L}^0 - \mathcal{L}_{ab}^0 \qquad \textit{with} \qquad \mathcal{L}_{ab}^0 = \Gamma_b^{0\prime}\Gamma_b^0 + \mathcal{R}_a^{0\prime}\mathcal{R}_a^0 \ . \tag{31}$$

The term $\mathcal{E}_{ab}^0$ is the unexplained loss; the loss that cannot be accounted by our subsets a and b of identified shocks.[15] Intuitively, the lower bound corresponds to the case where the unexplained loss is already minimal, that is could not have been lowered with another reaction function, while the upper bound corresponds to the case where the unexplained loss could have been entirely set to zero with a different reaction function.

The larger the fraction of the total loss that can be identified with the subset b, the tighter the confidence bounds, and thus the more exhaustive the overall policy evaluation: in the limit where we can identify all structural shocks, the unexplained loss is zero ($\mathcal{E}_{ab}^0 = 0$), the upper and lower bounds coincide, and we have an exhaustive policy evaluation for the policy instruments captured by the subset a.

In the other direction — not being able to identify all policy shocks — can also lead to a non-exhaustive evaluation. Across different policy instruments there is little that can be done as knowing the causal effects of a certain policy instruments implies nothing about the loss that could have been avoided using other policy instruments. However, when $\boldsymbol{\epsilon}_a$ captures a subset of the path of single specific policy instrument,[16] considering a higher level of time aggregation allows to probe the sensitivity of the results to "missing horizons" in the

[15] Importantly, it is possible to estimate $\mathcal{E}_{ab}^0 = \mathcal{L}^0 - \mathcal{L}_{ab}^0$. First, because the realized loss gives an estimate of $\mathcal{L}^0 = \mathbb{E}(\mathbf{Y}'\mathcal{W}\mathbf{Y})$ computed under ϕ^0. To see that, let $\mathbf{Y}_t^w$ denote the vector of selected elements of $\mathcal{W}^{1/2}\mathbf{Y}_t$, where $\mathbf{Y}_t$ is the time t realization of $\mathbf{Y}$. Suppose that the evaluation period is from $t = 1, \ldots, n$, then $\frac{1}{n}\sum_{t=1}^n \mathbf{Y}_t^{w\prime}\mathbf{Y}_t^w$ provides an estimate for $\mathcal{L}^0$. Further, $\mathcal{L}_{ab}^0$ can be measured from sufficient statistics.

[16] To give a concrete example; if the subset ϵ_a only features short-horizon policy shocks, then the subset-based evaluation will only informative about how well policy makers used the policy path at short horizons.

subset $\boldsymbol{\epsilon}_a$. In the limit where the time unit step becomes the entire period of evaluation,[17] the distance Δ_a captures to the total distance to minimum loss Δ as the policy problem becomes static (as in Section 2), and the subset policy evaluation becomes exhaustive.[18]

4.3 Policy comparison with subset shock identification

Consider now a researcher aiming to compare p policy makers. Each policy maker operates in an economy that can be described by the general model (15), but the parameters θ and ϕ that govern the model may vary across policy makers, say θ^j and ϕ^j, for $j = 0, \ldots, p$, where ϕ^j denote the reaction function chosen by policy maker j.

Since the DML measures the distance to an optimal loss *given* the environment, we can use the DML to compare policy makers evolving in different environments. This comparison would rank policy makers based on how well they used their different policy instruments in response to all the different types of non-policy shocks that affected their economy.

Again, in practice it may not be possible to identify all the shocks required to compute Δ. In that case, we can compare policy makers based on the subset statistics described above, keeping in mind the limitations associated with the sizes of the $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\Xi}_b$ subsets. And should one worry about the limited scope of a subset-based comparison, one can combine time aggregation and bounding to assess robustness to subset identification of all policy and non-policy shocks.

In addition, we can exploit the ORA/DML complementary to get a sense of the reasons for a particular ranking. Specifically, since each element of the subset ORA $\mathcal{T}^*_{ab}$ captures the distance to the optimal reaction to a specific shock, we can compare the ORAs across policy makers, provided we can identify the *same* policy and non-policy shocks across policy makers. This requires identifying the same type of non-policy shock (e.g., an oil shock vs a financial shock), as well as the same *dynamic* type of shock, for instance identifying the same *contemporaneous* oil shocks. Depending on the nature of the underlying shock and on the time period considered, this requirement may or may not be easy to fulfill.[19] We will discuss such cases in the empirical applications.

[17]This is the route followed by Blinder and Watson (2016), who evaluate US presidents from average realizations over the entire policy makers' tenure.

[18]Another promising avenue could exploit the recent "VAR-plus" approach proposed by Caravello, McKay and Wolf (2024) and use structural assumptions on the transmission of shocks in order to complement the subset shock evidence.

[19]If a researcher suspects a problem of "dynamic shock heterogeneity" across policy makers, a solution is again to consider a higher level of time aggregation, as this will mechanically mute dynamic heterogeneity across shocks.

4.4 Computing ORA and DML statistics

An attractive feature of the DML statistics and its variants is that they can be readily computed from standard econometric methods. The sufficient statistics underlying the DML —impulse responses to structural shocks— are well studied statistics, and we can draw on a large macro-econometric literature precisely devoted to the estimation of these statistics, from the identification of structural shocks (e.g., Ramey, 2016; Stock and Watson, 2018) to the estimation of impulse responses from VARs or local projections (e.g., Kilian and Lütkepohl, 2017; Plagborg-Møller and Wolf, 2021; Li, Plagborg-Møller and Wolf, 2024).

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

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

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

Using the approximating distribution $\widehat{F}$, we can simulate draws for $\mathcal{R}_a$ and Γ_b, and compute $\Delta_{\Xi,ab}$ and $\Delta_{\epsilon,aa}$ for each draw. Given the sequence of draws we can construct confidence sets for the corresponding ORAs and DMLs.

Last, to compute bounds on the overall DML Δ, we need to compute estimates for $\mathcal{L}^0$ and $\mathcal{E}^0$. To estimate $\mathcal{L}^0$, we use the realized losses as discussed in footnote 15 and we can construct an estimate for $\mathcal{E}^0$ from $\mathcal{E}^0 = \mathcal{L}^0 - \mathcal{L}^0_{ab}$, where we estimate $\mathcal{L}^0_{ab}$ from (31).

5 Evaluating US monetary policy, 1879-2019

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

During the Gold Standard period, there was no active monetary policy (the Federal Reserve did not exist yet), and we use this period as a benchmark to see what a fictional policy institution could have done in this period. The Gold Standard monetary regime is now generally considered a sub-optimal regime with excessive fluctuations in inflation and unemployment (e.g. Friedman and Schwartz, 1963). In that context, this passive monetary policy period is instructive as a benchmark against which we can compare later Fed performances.

The early Fed period starts with the founding of the Fed in 1913 and ends with the US entering the second world war. The post-war period starts in 1951 with the Fed regaining some independence after the Treasury-Fed accord (e.g. Romer and Romer, 2004*a*).[20] The post Volcker period covers the Great Moderation period and ends right before the pandemic.

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

$$\mathcal{L} = \frac{1}{2}\mathbb{E}\sum_{h=0}^{H} \beta^h(\pi_{t+h}^2 + \lambda u_{t+h}^2) \ , \tag{32}$$

where π_t denotes the inflation gap, u_t the unemployment rate gap, β the discount factor and λ the preference parameter. While the targets π^* and u^* are irrelevant to rank/assess reaction functions,[21] we posit that $\pi^* = 2$ and $u^* = 5$ in order to compute realized losses in the naive approach that we describe next.

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

5.1 Naive approach

To provide a benchmark for our results, we first take a naive approach where we evaluate the Fed based on $\mathcal{L}^0$, which we estimate from realized outcomes for inflation and unemployment (Figure 2). Table 1 report realized losses for inflation and unemployment ($\mathcal{L}_x = \sum_{j=t_s}^{t_e} x_j^2$ for $x = \pi, u$) as well as the total realized loss ($\mathcal{L}_\pi + \mathcal{L}_u$).

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

[20] 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.

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

throughout.

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

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

5.2 Econometric implementation

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

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

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

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

We have

$$A_0 y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + e_t \ , \tag{33}$$

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

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

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

With the draws of the parameters from the posterior density we can compute the impulse responses to a policy shock ϵ_t (denoted by $\mathcal{R}_0^0$) and the impulse responses to non-policy shock ξ_t (denoted by Γ_0^0). We will report the corresponding subset ORA and DML statistics, as well as the corresponding bounds on Δ_a. In addition, to understand the mechanism that drive the DML $\tau_{\xi,0}^*$ we report the ORA adjusted impulse responses $\Gamma_0^* \equiv \Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$.

5.3 Shock identification

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

5.3.1 Monetary policy shocks

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

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

Post World War II regime For the Post World War II period we use the Romer and Romer (2004*b*) identified monetary policy shocks as instruments. Since there was no use of forward guidance before 1990 —Fed policymakers' views on the future policy path was "closely guarded" before 1990 (Rudebusch and Williams, 2008)—, we can consider that the Romer and Romer (2004*b*) monetary shocks capture solely contemporaneous policy shocks ($\epsilon_{t,t}$) and not news shocks to policy.[22]

Early Fed regime During the Early Fed period we use the Friedman and Schwartz (1963) dates extended by Romer and Romer (1989) as instruments to identify monetary policy shocks. We include five episodes —1920Q1, 1931Q3, 1933Q1, 1937Q1 and 1941Q3— where movements in money were "unusual given economic developments" (Romer and Romer, 1989). In the words of Romer and Romer (1989), these "unusual movements arose, in Friedman and Schwartz's view, from a conjunction of economic events, monetary institutions and the doctrines and beliefs of the time and of particular individuals determining policy". Since the concept of forward guidance in policy did not exist, we consider that the Friedman and Schwartz's dates capture solely contemporaneous policy shocks ($\epsilon_{t,t}$) and not news shocks to policy.[23].

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

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

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

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.[24] To guard ourselves against this possibility, we will also use an identification of monetary shock that is consistent across regimes, which will ensure that the monetary shocks are identified in the exact same way across regimes. Specifically, we use sign restrictions, another popular method to identify monetary shocks (e.g., Uhlig, 2005). This approach has the benefit that the same identification scheme can be implemented over the entire sampling period. With the VAR including inflation, unemployment, the policy rate and the growth rate of the monetary base, we impose the following sign restrictions: a positive monetary shock raises the short-term rate in impact, lowers money growth on impact, and lowers inflation and raises unemployment after a year. Other than that, the responses are unconstrained.[25]

5.3.2 Non-policy shocks

We now describe the identification of our five types of non-policy shocks. We again rely on standard identification methods in the literature, and we assume that these identification schemes identify the same shocks across periods.[26]

Financial shocks As financial shocks we use narratively identified bank panics. Each included panic was triggered by either a run on a particular trust fund or by foreign developments. The dates for the banking panics are taken from Reinhart and Rogoff (2009), Schularick and Taylor (2012) and Romer and Romer (2017). To capture the severity of the bank run, each non-zero entry is rescaled by the change in the BAA-AAA spread at the time of the run, similar to the re-scaling of Bernanke et al. (1997) and Romer and Romer (2017).[27]

[24] 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.

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

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

[27] 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

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

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

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

Inflation expectation shocks An important feature of a successful central bank is the anchoring of inflation expectations. In this context, we aim to measure how well the Fed has been responding to innovations to inflation expectations —a clear example being the de-anchoring of inflation expectations in the 1970s (Reis, 2021). To do so, we aim to identify inflation expectation shocks .

As measure of inflation expectations, we rely on the Livingston survey that has been continuously run over 1946-2019,[28] and includes a question about 8-months ahead inflation expectations. Prior to World War II, there are no systematic inflation expectation survey, so we instead rely on Cecchetti (1992)'s measure of 6-months ahead inflation expectations for the Early Fed period.[29]

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

1919 with yields on 10-year maturity government bonds from the Macro History database (Jordà et al., 2019).

[28]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).

[29]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.

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

Preliminaries

Before presenting our estimates for the ORA and DML statistics over the different periods, Figure 3 presents a decomposition of $\mathcal{L}^0$ based on (31).

A few comments are in order: (i) our five non-policy shocks and one monetary shock explain about 60 percent of our estimate for $\mathcal{L}^0$. The period with the lowest share is the Pre-Fed period with about 40 percent and the highest share is the Early Fed with about 70 percent. Overall, these are large shares, providing reassurance that our overall policy assessment will be based on the most important disturbances that affected the economy over each period. In terms of shock composition, the main contributors in each period lines up well with common wisdom: The Early Fed is characterized by large contributions of financial shocks, monetary shocks and inflation expectation shocks (consistent with e.g., Romer, 1992). The Post WWII period is characterized by large contributions of energy and inflation expectation shocks (e.g., Blinder, 2022). The Post Volcker period is characterized by a very small loss overall (the Great Moderation).

DML and ORA

Looking back at figure 1, recall that there are two ways to evaluate a policy maker: the distance to minimum loss (the DML Δ) and the distance to the optimal reaction coefficient (the ORA $\mathcal{T}^*$): both depict the same sub-optimal policy but from a different angle. The ORA measures the sub-optimality of the reaction function, while the DML measures the consequence of that sub-optimal reaction on the loss function. Depending on the variance of the underlying shocks and the environment (i.e., the shape of the loss function in Figure 1), the same ORA may imply small or large distances to minimum loss. For these reasons, we will comment on both the DML and the ORA as we present our results.

Table 2 reports our estimated subset DMLs for our five different shocks over four periods. Each subset DML captures how much loss could have been avoided with a better reaction to a specific type of shock. The rightmost column reports our estimated bounds for the total DML, and that column can be seen as our overall policy evaluation for each period. It reports bounds on the total loss that could have been avoided with a better reaction function, see Proposition 2.

To understand the determinants of these losses, Table 3 reports the ORA statistics computed over the four periods for our five non-policy shocks. Recall that the ORA is an adjustment to the coefficient $\mathcal{B}_{p\xi}$ in the policy rule,[30] so that a negative ORA indicates that the policy rate was set too high after a specific type of shock, either because the policy rate was increased too much or because it did not lowered enough.

In the main text, we focus on the main lessons of our exercise, leaving a more in-depth presentation of our results for the appendix. Our main results are as follows: (i) we estimate large and uniform improvements in the conduct of monetary policy, but *only* in the last 30 years, (ii) we cannot reject that the Fed's reaction to recent financial shocks (notably the 2007-2008 financial crisis) was appropriate, in contrast to the "highly" sub-optimal reaction of the Fed during the Great Depression, (iii) the Fed reaction function during the 1960s-1970s is almost as sub-optimal as the reaction function of the 1920s-1930s Fed, though the consequences of these sub-optimal reaction —the distances to minimum loss— were much more limited in the post WWII period, and (iv) despite much larger DMLs in the 1920s-1930s, the reaction function of the early Fed is no worse than the passive Gold Standard.

Improved policy in the Post Volcker period

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

Looking first at our DML bounds (Table 2, right column), the Early Fed stands out with a much larger distance to minimum loss than in any other period: the DML range is comfortably outside all our other estimated ranges for the other periods. A mirror image of that Early Fed is the Post Volcker period with much superior performances: the DML range is much smaller than at any other time and lies outside all other estimated ranges. Interestingly, the Post WWII period and Pre Fed periods are relatively similar in terms of overall performance with similar ranges Δ.

To better understand the determinants of these losses, we can turn to the average ORAs over each period (Table 3, right column). For the first 100 years of our study, we find *no* material improvement in the reaction function (as captured by the ORA), with similar deviations from optimality over the first three periods. Comparing the rows of Table 3 for the three periods before Volcker, we can see ORAs of similar magnitudes with the average absolute ORA hovering around 0.6 for 100 years.

It is only in the last 30 years that we can detect improvements in the reaction function. In the post Volcker period, the ORAs are substantially smaller (and non-significant) than in the other periods, with an average absolute ORA of 0.2. In fact, the Post Volcker ORA

[30] 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.

statistics are smaller across *all* non-policy shocks, meaning that policy performance improves in all dimensions, from the responses to supply-type shocks like energy price shocks and TFP shocks to the responses to demand-type shocks like government spending shocks and financial shocks.[31] That said, the consequences of these sub-optimal reaction functions are very different across periods.

In terms of policy shocks however, i.e., deviations from a stable function, the conduct of policy did improve substantially before Volcker. While the early Fed stands out with much larger monetary shocks and large DML Δ_ϵ with the Fed being directly responsible for large disturbances affecting the economy, the post World War II sees a much smaller Δ_ϵ. In other words, even though the reaction function is not substantially superior in the post WWII period, erratic behavior in the conduct of policy was much improved after WWII, in contrast to the stop-and-go policies of the 30s or the over-reaction of the early 20s (e.g., Friedman and Schwartz, 1963; Romer, 1992).

Responding to financial shocks

We will now focus in more details on the reaction to financial shocks, contrasting the Post Volcker Fed with the Early Fed of the 1920s-1930s. In a 2002 speech in honor of Milton Friedman 90th birthday, (then) Fed governor Bernanke famously said: "Regarding the Great Depression. You're right, we did it. We're very sorry. But thanks to you, we won't do it again." (Bernanke, 2002). In an irony of history, the speech was made a full five years before the 2007-2008 financial crisis; a crisis that saw an unprecedented Fed response (see e.g., Bernanke, 2013) *with* Bernanke as Fed chairman.

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

To see this, we can first contrast the financial ORAs —the ORAs for financial shocks— estimated for the Early Fed period and for the Post Volcker period. With $\tau^* = -1.2$ (statistically significant), the Fed reaction to banking panics was too tight —a result echoing previous findings in the literature (e.g., Friedman and Schwartz, 1963; Hamilton, 1987)—. In contrast, the estimated ORA for the post Volcker Fed is four times smaller with $\tau^* = -0.3$ and not statistically significant, indicating that the post Volcker Fed period reacted much more appropriately and pointing to large improvements in the Fed's reaction to financial

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

shocks.[32] As a result, while the 2007-2008 financial disruptions were substantial, the corresponding estimated DML over the post-Volcker period is tiny (0.1); two orders magnitude smaller than the DML for financial shocks for the Early Fed (27.7).

To better appreciate this reaction function improvement, Figures 4 and 5 display the impulse responses underlying the financial ORAs estimated for 1913-1941 and 1990-2019. The top rows show the impulse responses of inflation, unemployment and the interest rate to a monetary policy shock, while the bottom rows show the responses of the same variables to a financial shock.[33]

For the Early Fed period, notice how the Fed *raised* the discount rate in response to financial shocks. Combined with the decline in inflation caused by the financial shock, this means that the real policy rate increased substantially and monetary policy was contractionary, confirming earlier work on the monetary factors behind the Great Depression (e.g., Friedman and Schwartz, 1963; Hamilton, 1987). The ORA corrects this sub-optimal reaction function and turns the table on monetary policy by running an expansionary policy.[34] To see that, Figure 4 (dashed green line) reports the ORA adjusted impulse responses —$\Gamma_0^* = \Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$—, which depict how the ORA adjustment translates into different policy path responses to non-policy shocks and "improved" (i.e., more stable) impulse responses of inflation and unemployment. The ORA leads to a major adjustment to the policy path —the policy rate now goes down substantially on impact—, and the paths of inflation and unemployment are consequently much more stable. In contrast, for the Post Volcker period (Figure 5) the policy rate goes down following a financial shock (black line, lower-right panel), and the ORA only slightly adjusts the response of the policy rate (green line), leading to modest adjustments to the responses of inflation and unemployment.

The Great Inflation

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

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

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

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

had been.

Overall, we find that the Fed's reaction function during the 60s-70s is on a par with the reaction function of the early Fed, with ORAs of similar magnitudes, though the nature and the sizes of the underlying shocks is different. Post World War II, the Fed reaction was too weak following all the different supply-type shocks that we identified: energy price shocks, TFP shocks as well as inflation expectation shocks. In fact, the reaction to inflation expectation shocks over the 60s-70s displays the largest deviation from optimality over the entire 150 year of monetary history with $\tau^* = 1.2$, even slightly larger (in absolute value) than the Fed's poor reaction to bank runs during the Great Depression. The consequences of these sub-optimal reactions were much smaller however, with DMLs an order of magnitude smaller for the Post WWII period than for the Early Fed period.[35]

To better appreciate these sub-optimal reaction functions, Figure 6 plots the impulse responses underlying the ORAs for inflation expectation shocks (similar results hold for energy or TFP shocks, see the appendix). In response to an inflation expectation shock, inflation rises progressively, but the policy rate does not respond, leading to negative real interest rates and further increasing inflation. The (large) ORA adjustment restores the Taylor principle: after the ORA, the policy rate rises strongly following an inflation expectation shock (lower-right panel, Figure 6) and stems the rise in inflation (at the cost of higher unemployment).

The early Fed vs the passive Gold Standard

In contrast to the suggestive evidence of the naive approach (Table 1), the passive Gold Standard is *not* markedly superior to the early Fed. In other words, the founding of the Fed did not deteriorate performance relative to the passive monetary regime of the Gold Standard. Instead, the reaction functions were just as "bad" before and after the founding of the Fed, though the consequences (in terms of DMLs) of these sub-optimal reactions were much larger for the Early Fed period.

Comparing the ORA before and after the founding of the Fed (Table 3, first two rows), we observe similar deviations from optimality. During the passive Gold Standard, monetary policy is (unsurprisingly) too passive in the face of adverse shocks: be it bank runs or military buildups.[36] Comparing ORAs across two the periods, we can see that (i) the excessive

[35]In general, there are two broad possible reasons why a given ORA translates into larger losses and hence into a larger DML: (i) the magnitudes of the disturbances themselves were larger (in this case, in the Early Fed period relative to the post WWII period), and/or (ii) the Early Fed economy was less resilient in the face of adverse disturbances than the post WWII economy. Exploring the relative merits of these two alternative is left for future research.

[36]To give a few noteworthy "misses" of the passive Gold Standard, the ORAs call for lower interest rates (about 3/4 ppt) in the aftermaths of the 1893 and 1907 bank runs, as well as higher interest rates in response to higher military spending following the war against Spain in 1898, and the navy build-up of 1902-1904. See

passivity simply continued after the founding of the Fed —the ORAs are similar across the two periods—, and (ii) the excessive passivity of the early Fed is not limited to financial distress, and it also extends to other shocks, here government spending shocks.[37]

Robustness and caveats

In the appendix, we show robustness to (i) our identification of monetary shocks, (ii) possible dynamic heterogeneity in shocks across periods, and (iii) alternative periods. Overall, our results are consistent with our baseline estimates, with ORAs of similar magnitudes and levels of statistical significance.

As final comments, we note two important caveats. First, our analysis takes as starting point a loss function with equal weights on inflation and unemployment and studies whether different reaction functions could have achieved lower losses, thereby attributing some of the variation in inflation and unemployment to sub-optimal policies. Different loss functions *might* justify past reaction functions, and it is not our objective to argue in favor of one loss function vs another.[38] Instead, we provide a framework to evaluate and compare policy makers *given* a loss function that the researcher deems appropriate to evaluate performance.

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

6 Conclusion

In this paper, we propose to evaluate makers based on their Distance to Minimum Loss (DML): the distance to the minimum attainable loss given their environment. The approach can be implemented with minimal assumptions on the underlying structural economic model, because the DML can be computed from two sets of sufficient statistics: (i) the impulse responses of the policy objectives to non-policy shocks, and (ii) the same impulse responses to policy shocks. Importantly, explicit knowledge of the policy maker's reaction function is

Figure S1 and the Appendix for more details, notably the impulse responses underlying the ORAs estimated for 1879-1913.

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

[38]That said, varying the loss function weight on unemployment over [0.2.5] (instead of 1 as in this study) gave similar results.

not necessary, because the effects of an (unspecified) reaction function are already encoded in the impulse responses to shocks, which are estimable.

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

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

References

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

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

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

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

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

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

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

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

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

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

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

Blinder, Alan S. 2022. *A monetary and fiscal history of the United States, 1961–2021.* Princeton University Press.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kuttner, Kenneth N. 2001. "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics*, 47(3): 523–544.

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

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

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf. 2024. "Local Projections vs. VARs: Lessons From Thousands of DGPs." forthcoming.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix A: Details and Proofs

Proof of Lemma 1. Define

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{yy} & \mathcal{A}_{yp} \\ \mathcal{A}_{py} & \mathcal{A}_{pp} \end{bmatrix} \, , \; \mathcal{B}_\xi = \begin{bmatrix} \mathcal{B}_{y\xi} \\ \mathcal{B}_{p\xi} \end{bmatrix} \, , \; \mathbf{J} = \begin{bmatrix} \mathbf{0} \\ \mathcal{B}_{p\epsilon} \end{bmatrix} \text{ and } \mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{P} \end{bmatrix} \, . \tag{34}$$

The model (15) 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

$$\mathcal{D}_1 = \begin{bmatrix} \Gamma(\phi,\theta) \\ \Gamma_p(\phi,\theta) \end{bmatrix} \quad \text{and} \quad \mathcal{D}_2 = \begin{bmatrix} \mathcal{R}(\phi,\theta) \\ \mathcal{R}_p(\phi,\theta) \end{bmatrix} \; ,$$

where the maps $\Gamma(\phi,\theta)$ and $\mathcal{R}(\phi,\theta)$ 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

$$\Gamma(\phi,\theta) = \mathcal{D}(\mathcal{B}_{y\xi} + \mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}) \qquad \text{and} \qquad \mathcal{R}(\phi,\theta) = \mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\epsilon} \; , \tag{35}$$

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

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

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

and we obtain the equilibrium representation

$$\mathbf{Y} = (\Gamma(\phi,\theta) + \mathcal{R}(\phi,\theta)\mathcal{T})\boldsymbol{\Xi} + (\mathcal{R}(\phi,\theta) + \mathcal{R}(\phi,\theta)\mathcal{T}_\epsilon)\boldsymbol{\epsilon} \; , \tag{36}$$

where

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

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

Proof of Proposition 1. Part 1 requires showing $\mathcal{L}^{\text{opt}} = \mathsf{L}(\mathcal{T}^*; \phi^0, \theta)$. To do so we proceed mechanically and show $\{\min_\phi \mathcal{L} \text{ s.t. } (15)\} = \{\min_\mathcal{T} \mathsf{L}(\mathcal{T}; \phi^0, \theta) \text{ s.t. } (15) \text{ with } \mathcal{A}_{pp} = \mathcal{A}_{pp}^0, \; \mathcal{A}_{py} = \mathcal{A}_{py}^0, \mathcal{B}_{p\xi} = \mathcal{B}_{p\xi}^0, \; \mathcal{B}_{p\epsilon} = \mathcal{B}_{p\epsilon}^0\}$.

Note that $\mathbf{Y}$ can be written as

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

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

$$\begin{aligned}\mathcal{L} =&\mathbb{E}(\mathbf{Y}'\mathcal{W}\mathbf{Y})\\ =&\text{Tr}((\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi})'\mathcal{D}'\mathcal{W}\mathcal{D}(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}))\\ &+\text{Tr}((\mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\epsilon})'\mathcal{W}\mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\epsilon})\ .\end{aligned}$$

Regardless of the values of $\{\mathcal{A}_{pp},\mathcal{A}_{py},\mathcal{B}_{p\xi}\}$ the optimal solution for $\mathcal{B}_{p\epsilon}$ satisfies $\mathcal{B}_{p\epsilon}^{\text{opt}}=\mathbf{0}$. After setting $\mathcal{B}_{p\epsilon}=\mathcal{B}_{p\epsilon}^{\text{opt}}$ the derivative maps of $\mathcal{L}$ with respect to $\{\mathcal{A}_{pp},\mathcal{A}_{py},\mathcal{B}_{p\xi}\}$ are given by

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

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

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

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. It remains to show that $\mathcal{B}_{p\xi}^{\text{opt}}=\mathcal{B}_{p\xi}^0+\mathcal{B}_{p\epsilon}^0\mathcal{T}_\xi^*$ and $\mathcal{B}_{p\epsilon}^{\text{opt}}=\mathcal{B}_{p\epsilon}^0+\mathcal{B}_{p\epsilon}^0\mathcal{T}_\epsilon^*$. The latter is straightforward as $\mathcal{T}_\epsilon^*=-\mathbf{I}$. For the former we have

$$\begin{aligned}\mathcal{B}_{p\xi}^0+\mathcal{B}_{p\epsilon}^0\mathcal{T}_\xi^* &=\mathcal{B}_{p\xi}^0-\mathcal{B}_{p\epsilon}^0(\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma^0\\ &=\mathcal{B}_{p\xi}^0-\mathcal{B}_{p\epsilon}^0((\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1}\mathcal{B}_{p\epsilon}^0)'\mathcal{W}\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1}\mathcal{B}_{p\epsilon}^0)^{-1}(\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1}\mathcal{B}_{p\epsilon}^0)'\mathcal{W}\Gamma^0\\ &=\mathcal{B}_{p\xi}^0-((\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})^{-1}(\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\Gamma^0\\ &=\mathcal{B}_{p\xi}^0-((\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})^{-1}(\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\mathcal{D}\mathcal{B}_{y\xi}\\ &+((\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})^{-1}(\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}\\ &=-((\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})^{-1}(\mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1})'\mathcal{W}\mathcal{D}\mathcal{B}_{y\xi}=\mathcal{B}_{p\xi}^{\text{opt}}\ ,\end{aligned}$$

where the third equality uses that $\mathcal{B}_{p\epsilon}^0$ is invertible.

Part 2 uses the optimal characterization from part 1. We have

$$\begin{aligned}\Delta&=\mathcal{L}^0-\mathcal{L}^{\text{opt}}\\ &=\text{Tr}(\Gamma^{0'}\mathcal{W}\Gamma^0)+\text{Tr}(\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)-\text{Tr}((\Gamma^0+\mathcal{R}^0\mathcal{T}_\xi^*)'\mathcal{W}(\Gamma^0+\mathcal{R}^0\mathcal{T}_\xi^*))\\ &=\text{Tr}(\Gamma^{0'}\mathcal{W}\mathcal{R}^0(\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma^0)+\text{Tr}(\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)\end{aligned}$$

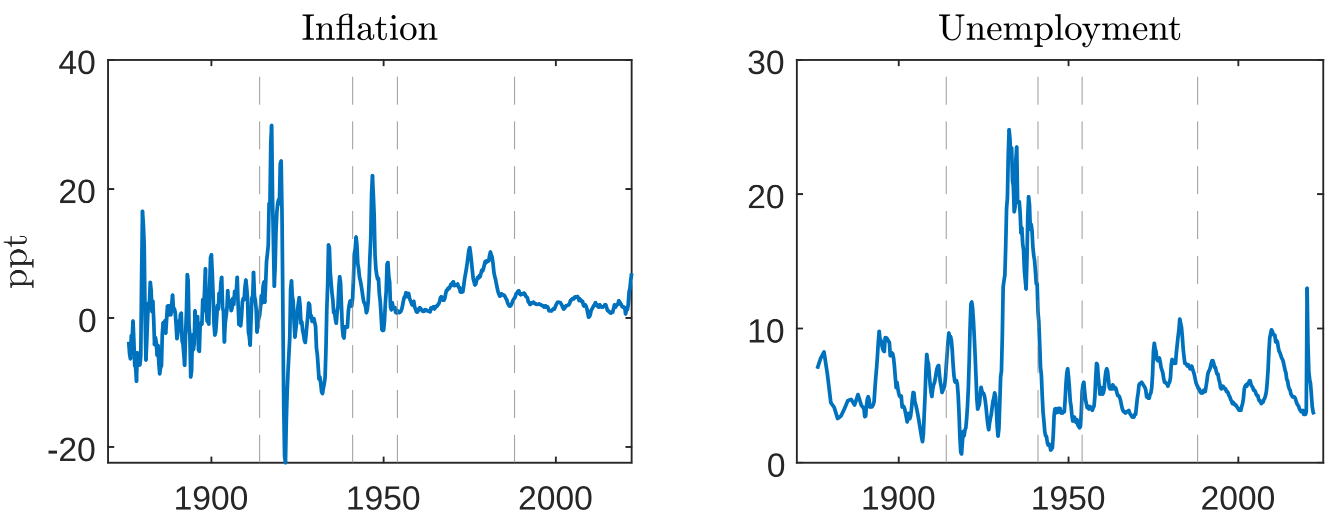
which completes the proof. □

Proof of Proposition 2. From $\Delta_{\Xi,a}=\text{Tr}(\Gamma^{0'}\mathcal{W}\mathcal{R}_a^0(\mathcal{R}_a^{0'}\mathcal{W}\mathcal{R}_a^0)^{-1}\mathcal{R}_a^{0'}\mathcal{W}\Gamma^0)$ and We can write the impulse response matrix Γ as $\Gamma=[\Gamma_b,\Gamma_{-b}]$ where Γ_{-b} contains the impulse responses to the non-policy shocks that were not identified.

$$
\begin{aligned}
\Delta_{\Xi,a} &= \mathrm{Tr}(\Gamma^{0'}\mathcal{W}\mathcal{R}_a^0(\mathcal{R}_a^{0'}\mathcal{W}\mathcal{R}_a^0)^{-1}\mathcal{R}_a^{0'}\mathcal{W}\Gamma^0) \\
&= \mathrm{Tr}(\Gamma_b^{0'}\mathcal{W}\mathcal{R}_a^0(\mathcal{R}_a^{0'}\mathcal{W}\mathcal{R}_a^0)^{-1}\mathcal{R}_a^{0'}\mathcal{W}\Gamma_b^0) + \mathrm{Tr}(\Gamma_{-b}^{0'}\mathcal{W}\mathcal{R}_a^0(\mathcal{R}_a^{0'}\mathcal{W}\mathcal{R}_a^0)^{-1}\mathcal{R}_a^{0'}\mathcal{W}\Gamma_{-b}^0) \\
&\leq \Delta_{ab} + \mathcal{E}_{ab}^0
\end{aligned}
$$

The lower bound is obtained by imposing that the unexplained loss is already minimal, that is could not have been lowered with another reaction function. The upper bound is obtained by imposing that the unexplained loss could have been entirely set to zero with an optimal reaction to the unidentified subsets of shocks. □

Figure 2: Inflation and unemployment, 1879–2019



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

Table 1: Realized losses

	Pre Fed 1879-1912	Early Fed 1913-1941	Post WWII 1951-1984	Post Volcker 1990-2019
$\mathcal{L}_\pi^0$	24.3	83.0	11.9	0.7
$\mathcal{L}_u^0$	5.7	81.0	6.5	5.8
$\mathcal{L}^0$	30.0	164.0	18.4	6.5

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

Table 2: Distances to Minimum Loss (DML)

	Δ_{ab}						Δ_a
	Bank panics	G	Energy	π^e	TFP	MP	[lower bound, upper bound]
Pre Fed 1879−1912	**1.5** (0.3,3.5)	**0.6** (0.1,2.1)	**0.2** (0,0.6)	—	**0.6** (0.1,1.7)	**1.5** (0.7,3.2)	**[2.9, 17.3]**
Early Fed 1913−1941	**27.7** (11.5,67.4)	**6.6** (0.8,24.3)	**1.6** (0.1,8.6)	**27.9** (9.4,70.4)	**2.3** (0.2,11.8)	**18.5** (8.8,37.9)	**[66.1, 101]**
Post WWII 1951−1984	—	**0.1** (0,0.8)	**0.9** (0.1,3.5)	**1.7** (0.3,5.5)	**0.4** (0,2.3)	**1.2** (0.4,3.2)	**[3.1, 11.0]**
Post Volcker 1990−2019	**0.1** (0,0.6)	**0.1** (0,0.4)	**0.2** (0,0.9)	**0** (0,0.2)	**0.1** (0,0.4)	**0.7** (0.3,1.9)	**[0.5, 2.6]**

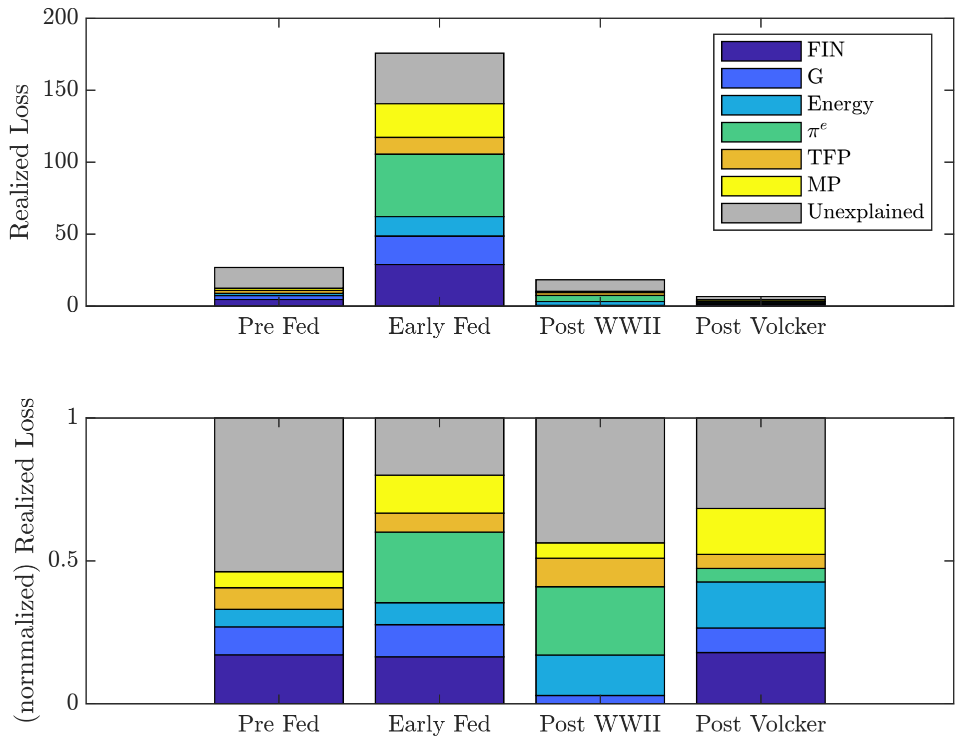
Notes: Median subset distance to minimum loss (DML, Δ_{ab}) together with 68% credible sets with each row reporting estimates for a different period. Each column reports the subset DML for one type of shock: the financial shocks are bank panics from Reinhart and Rogoff (2009), the government spending shocks (G) are from Ramey and Zubairy (2018), TFP shocks from Gali (1999), energy shocks are computed using the peak-over-threshold approach of Hamilton (1996), and inflation expectation shocks (π^e) are innovations to inflation expectations as measured from Cecchetti (1992) for Early Fed period and from the Livingston survey after 1946. For the Pre Fed period the TFP, G and Energy ORAs are computed over the 1890-1912 period. The monetary policy shocks (MP) are identified as described in the main text: using gold rush discoveries in the pre-Fed period, Romer and Romer (1989)'s Friedman-Schwartz dates in the early Fed period, Romer and Romer (2004) monetary shocks for the post WWII period and high-frequency surprises in the post Volcker period. The rightmost column reports median estimates for upper and lower bounds for the total distance to minimum loss (Δ_a).

Table 3: Optimal Reaction Adjustments (ORA)

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

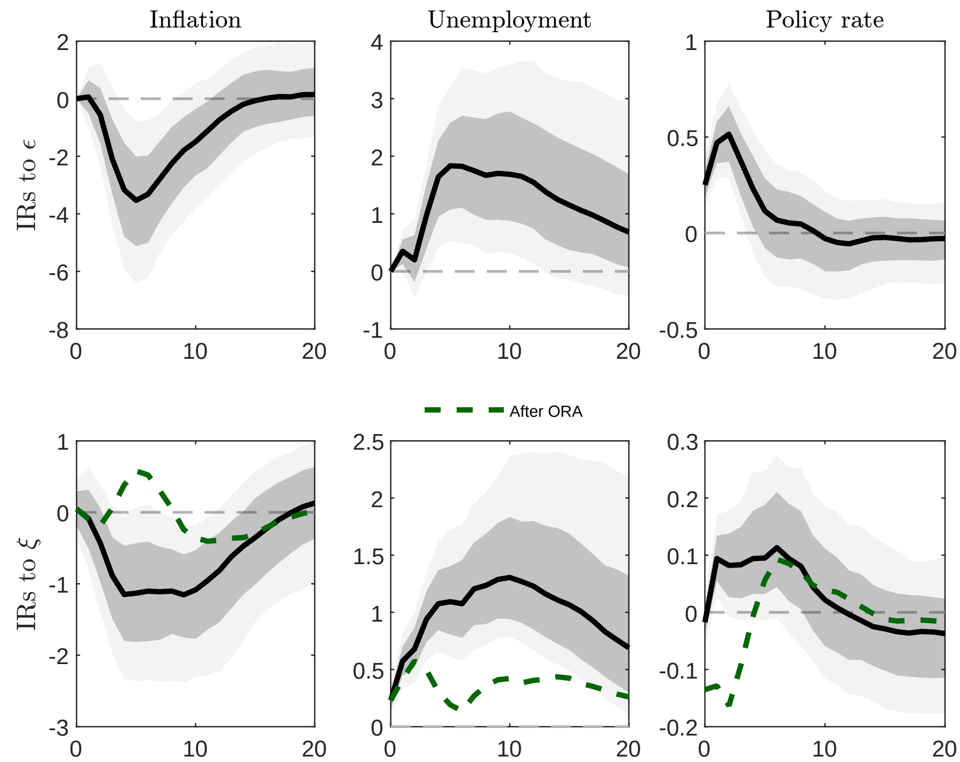
Notes: Median ORA statistics together with 68% credible sets. The monetary policy shocks are identified as described in the main text: using gold rush discoveries in the pre-Fed period, Romer and Romer (1989)'s Friedman-Schwartz dates in the early Fed period, Romer and Romer (2004) monetary shocks for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

Figure 3: Loss ($\mathcal{L}^0$) decomposition, 1879-2019



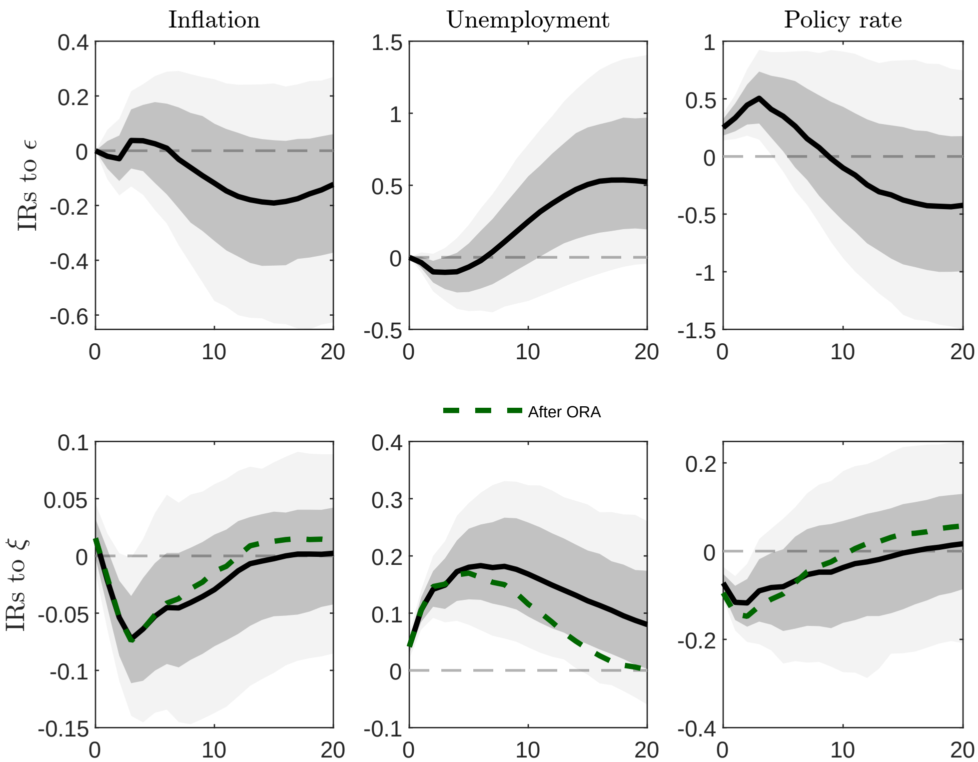
Notes: Decomposition of $\mathcal{L}^0$ into the contribution of each identified shock (financial, government spending, energy price, inflation expectation, TFP, monetary policy). The grey shaded area depicts the unexplained part of $\mathcal{L}^0$. The top panel reports decomposition for the level of $\mathcal{L}^0$, while the bottom panel reports the same decomposition but for the level of $\mathcal{L}^0$ normalized on 1 in each period.

Figure 4: Early Fed, 1913-1941, Reaction to Financial shocks



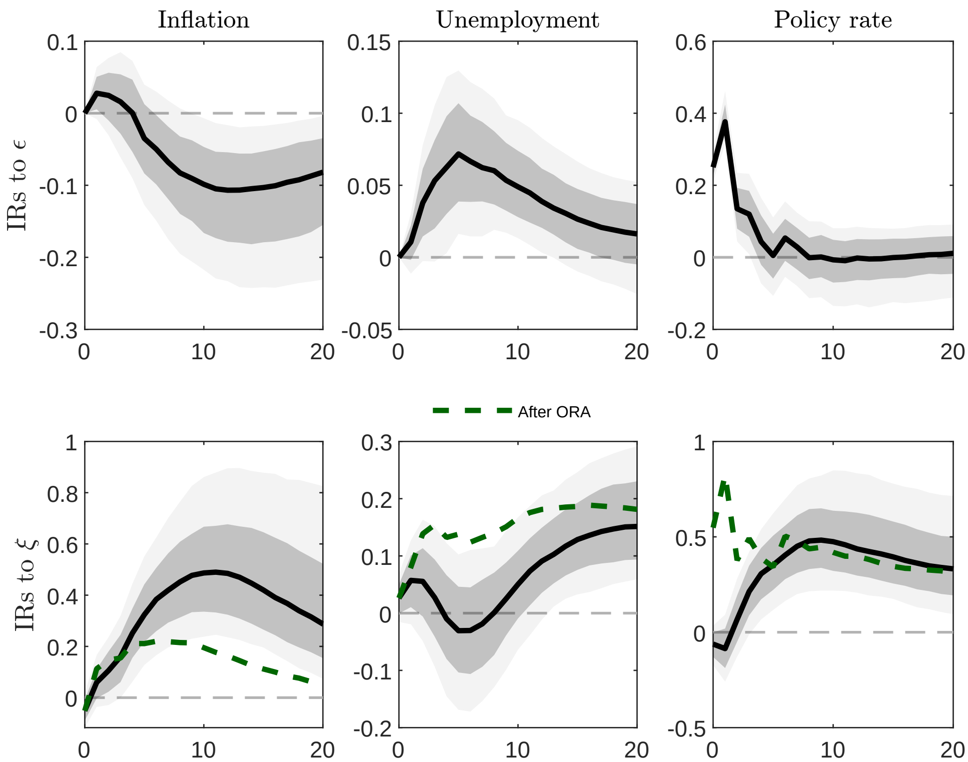
Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the Fed's discount rate to a monetary policy shock ϵ (resp. financial shock ξ). The dotted green lines show the ORA adjusted impulse responses: $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95% and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

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



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

Figure 6: Post WWII Fed, 1951-1984, Reaction to π^e shocks



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