Evaluating policy makers' performance

a reaction function test*

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

Preliminary draft: Comments very welcome

July 14, 2022

Abstract

How can we evaluate and compare the performance of policy makers after their term in office? An evaluation based on realized outcomes is flawed since different policy makers can face different initial conditions, shocks and economic environment. In this paper propose to evaluate policy makers based on the distance-to-optimality of their reaction function and show that this measure is comparable across policy makers. The distance-to-optimality can be computed from well-known statistics —the impulse responses to policy and non-policy shocks—, so that our method does not require specifying the underlying economic model nor the reaction function. We illustrate the methodology by comparing the performance of the Fed across different time periods.

JEL classification: C14, C32, E32, E52.

Keywords: Optimal policy, Policy evaluation, Hypothesis testing, Specification tests.

*We thank Christian Brownlees, Ryan Chahrour, Kirill Evdokimov, Sylvain Leduc, Barbara Rossi and Adam Shapiro for helpful comments. Mesters acknowledge support from the Spanish Ministry of Economy and Competitiveness through the Ramon y Cajal fellowship (RYC2019-028287-I), the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S), and the Netherlands Organization for Scientific Research (NWO) through the VENI research grant (016.Veni.195.036). 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 can we evaluate and compare the performance of policy makers after their term in office? A naive approach would be to evaluate performance based on realized economic or social outcome variables. For instance, we could compare central bank chairs based on average inflation and unemployment outcomes, or presidents of countries based on average growth, inequality or even CO_2 emissions. While such approach is commonly adopted in the popular press, there are a number of possible criticisms: (i) different policy makers face different initial conditions, e.g. a president can inherit a strong or weak economy from her predecessor, (ii) different policy makers face different economic shocks, e.g. oil price shocks can affect the ability of central banks to control inflation but their occurrence is not controlled by the central bank, and (iii) different policy makers face different structural economic environments, e.g. a Covid lockdown created new a environment under which offsetting adverse shocks was arguably more complicated.

This triplet of confounding factors coming from different initial conditions, different structural shocks and different economic environments severely complicates the evaluation of policy makers based on realized outcomes.[1]

To make progress it is instructive to consider an ideal, yet infeasible, approach for comparing policy makers: an experimental setting. Consider setting up a laboratory, in which different policy makers enter and are subjected to the same initial conditions and economic environment. Subsequently they are exposed to the same sequence of shocks and can make decisions that aim to control some outcome variables —the policy objectives—. Afterwards, we can compare the average outcomes and conclude which policy maker performed better. Since the only variation in outcomes comes from the decisions of the policy makers, such comparison would be on equal grounds.

The objective of this paper is to develop econometric methods that can bring us closer to mimicking this ideal experiment in an observational setting. We restrict our analysis to comparing policy makers who have at least one common policy instrument at their disposal. As such we may compare policy makers from the same institution who operated in different time periods, but also policy makers from similar institutions across different countries.

A first step towards mimicking the ideal experiment comes from realizing that in practice different macro policy makers never face identical sequences of economic shocks. This implies that any evaluation approach that aims to compare how different policy makers responded to different shocks needs to condition on the shocks. By comparing average outcomes conditional on shocks we can avoid imbalances stemming from initial conditions and different shock sequences.

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

That said, conditioning on shocks alone does not equalize the different economic environments that different policy makers typically face. Environments are defined as all relationships/coefficients in the economy that are not directly under the control of the policy maker, and we note that in some environments it may simply be harder or easier to offset shocks. In other words, what is missing after conditioning on the shocks is what the policy maker could have done to offset the shocks given the environment she faced.

These observations provide the starting point for our paper. First, we formalize the actions that the policy maker could take in terms of a generic reaction function which captures how the policy maker responds to all endogenous variables and exogenous shocks. Importantly, we will not assume that we know or can estimate this function, we merely pose its existence. Second, we aggregate the different policy objectives in a loss function. The choice for the variables in the loss function is determined by the researcher conducting the comparison.

Our evaluation of the policy maker's performance is based on measuring the distance between the policy maker's reaction function and the optimal reaction function —the systematic response to endogenous and exogenous variables that minimizes the loss function—. This approach has two important advantages. First the distance to the optimal reaction to economic shocks can be estimated without relying on a specific structural model or a specific reaction function. Second, the distance measure is invariant to different (i) initial conditions, (ii) shocks faced, and (iii) economic environments.

Specifically, we approach the evaluation of policy makers as a hypothesis testing problem, more specifically as a specification testing problem in the spirit of Hausman (1978) and Breusch and Pagan (1980). A specification test consists in verifying whether a necessary condition underlying a method, a model or here an optimal reaction function is indeed verified. A policy maker's reaction function is optimal if it minimizes the policy maker's loss function, i.e., if it provides the best response to the state of the economy in order to achieve the policy objectives. It follows that under optimality, the gradient of the loss function with respect to the response to all non-policy shocks should be zero, similar to a standard score test. The important benefit of a score test is that it does not require the estimation of the parameters that are fixed under the null (e.g. Breusch and Pagan, 1980). In our case, the reaction function is "fixed under the null", and there is no need to estimate or know that reaction function.

Instead, we show that the gradient with respect to the response to non-policy shocks depends only on familiar statistics: the impulse responses of the policy objectives to both policy shocks and non-policy shocks. Specifically, the gradient is found to be equal to the inner-product (over the impulse response horizons) between the impulse responses of the policy objectives to policy and non-policy shocks. When the gradient is zero the impulse responses to the policy and non-policy shocks should be orthogonal and there is no adjustment to the reaction function that can lower the loss function. By drawing on a large macro-econometric literature on the estimation of impulse responses, we can compute the gradient using standard methods, and evaluate whether the gradient is zero to assess whether the reaction function was optimal.

Since the gradient condition must hold for all policy and non-policy impulse responses, we can construct subset tests that allow to verify whether the gradient condition holds for any combination of policy instruments and non-policy shocks that affect the economy. As a concrete example for monetary policy one can think of testing whether the reaction function that characterizes the short term interest rate was optimal for off-setting oil price shock, or TFP shocks, or fiscal shocks, etc. The benefit of such subset tests is twofold: (i) they are easy to interpret economically and provide concrete evidence for which shocks the policy maker did not respond to optimally, and (ii) they are feasible, as in practice we will not be able to identify all shocks in the economy, but the macro econometric literature has been able to identify a subset of different shocks (e.g. Ramey, 2016; Stock and Watson, 2012).

Next, we rescale the subset gradient statistics (by the negative inverse of the Hessian) to measure the distance of the reaction function to nearest reaction function that responded on average optimal to the non-policy shocks considered. In other words, we measure how far from optimality was the policy maker's systematic response to non-policy shocks. This distance is conditional on the specific non-policy shocks considered and precisely measures what could have been done given the economic environment. Moreover, as with the gradient the "distance to optimality" statistic only depends on the impulse responses to policy and non-policy shocks.

We adopt the distance to optimality statistic to (i) evaluate the absolute performance of policy makers in responding to a particular subset of structural shocks, and (ii) to compare the performance of different policy makers across different environments.

The intuition for our approach comes from combining two insights. First, recall that policy makers can modify the effects of non-policy shocks on the policy objectives by changing their reaction function. In other words, the impulse response to the non-policy shocks encodes how the policy maker responded systematically to non-policy shocks. Second, to infer what could have been done to offset non-policy shocks we note that the impulse responses of changes in the policy makers response to non-policy shocks are proportional to the impulse responses to policy shocks. To see this, note that the effect of any exogenous shock to the policy instruments is characterized in equilibrium by the impulse responses to the policy shocks. Hence, the innovation induced by responding differently to any exogenous non-policy shock has a proportional equilibrium effect.

Combining these insights it follows that if the impulse responses to policy changes —the causal effects of responding differently to non-policy shocks— are orthogonal to the impulse responses to non-policy shocks, there is nothing more that can be done to offset the expected effects of the non-policy shocks. Moreover, the projection of the impulse responses to non-policy shocks on the impulse responses to policy shocks —a regression in impulse response space— provides the adjustment to the reaction function that makes the reaction function optimal in the direction of the systematic response to those non-policy shocks. This our distance to optimality statistic.

Informally put, we can evaluate reaction functions without estimating them as the effect of an (unknown) reaction function is *already* encoded in the impulse responses to non-policy shocks, capturing what the policy maker did, and policy shocks, capturing what the policy maker could have done. Thus estimating impulse responses to policy and non-policy shocks is enough to assess (i) the optimality of a policy maker's reaction function, (ii) compute the distance of the reaction function to optimality in the direction of the response to non-policy shocks and (iii) compare the reaction functions among different policy makers.

We show that the aforementioned results hold in a broad class of forward looking macro models similar to those considered in McKay and Wolf (2022) and Barnichon and Mesters (2022). Specifically, within this class we show that if the researcher is able to identify at least one policy and one non-policy shock, we can evaluate the performance of the policy makers based on these shocks. This implies that any existing identification and estimation method can be used to evaluate policy makers' systematic performance based on their reaction function. Clearly, being able to identify more shocks will improve our ability to evaluate policy makers. Prominent members in the general model class include New Keynesian models and structural vector autoregressive models (SVAR) which are often adopted for policy making. We explicitly illustrate our approach for these examples, but we note that also more modern macro models, such as HANK models, are included in the class. Furthermore, extensions for time-varying parameters and state dependence can be easily considered.

To illustrate our method, we systematically evaluate the Fed's reaction function over the past 60 years. Notably, we revisit and refine the influential findings of a large literature (see Judd and Rudebusch, 1998; Taylor, 1999; Clarida, Galí and Gertler, 2000) that concluded that the Fed conduct of monetary policy improved after 1980, because it satisfied the Taylor principle —a central bank should react more than one-to-one in the fact of inflation movements—.[2] However, beyond that Taylor principle, that literature can say little about the optimality of the reaction function, whether the central bank was reacting too much or (still) too little after 1980.

In addition to not relying on a simple interest rate rule, our approach can provide a

[2]Based on Taylor rule estimates, these papers found that the parameters of the Taylor rule shifted around 1980 and that the Fed responded more vigorously to inflation variations after 1979, though this conclusion has not gone unchallenged (e.g., Orphanides, 2003).

much more refined evaluation of the reaction function. Our evaluation can establish whether the Fed systematically reacted to much or too little to certain shocks. Consistent with conventional wisdom that Fed monetary policy was not appropriate in the pre-Volcker period, we can reject that the Fed's reaction to oil price shocks and TFP shocks was appropriate: the Fed did not raise the fed funds rate sufficiently in the face of policy oil price shocks, and the Fed did not raise the fed funds sufficiently in the face of negative TFP shocks. After 1985 however, and in line with previous model-based studies (e.g., Galı, López-Salido and Vallés, 2003), we find that the reaction to both oil shocks and technology shocks (notably during the tech boom of the late 90s) was appropriate, i.e., we cannot reject optimality.

The remainder of this paper is organized as follows. We continue the introduction by briefly relating our approach to the existing literature. The next section illustrates our method for a simple New Keynesian model and a structural VAR model. Section 3 presents the general environment. Sections 4 and 5 provide the key results for evaluating reaction functions and comparing them across policy makers. The results from the empirical study for monetary policy are discussed in Section 6. Extensions are presented in Section 7 and Section 8 concludes.

Relation to the literature

The difficulties associated with evaluating and comparing policy makers based on realized outcomes were outlined in Fair (1978). Blinder and Watson (2016) improve on such approach by projecting out specific non-policy shocks to assess the contribution of such shocks to explain the unconditional difference between policy makers. We instead project the observed variables on the non-policy shocks and thus explicitly study how the policy maker responded on average to these shocks and, importantly, incorporate how the policy maker could have responded, given the different environments that they faced.

A different approach for evaluating policy makers is based on structural models. An early example is Fair (1978) who uses optimal control methods, but more modern DGSE approaches (e.g. Gali and Gertler, 2007) can also be considered. An possible downside of such approach is that the structural model may be mis-specified. In our approach we can reduce this risk as we only require the estimation of impulse responses for which more robust reduced form econometric models can be adopted.

In addition, evaluation based on structural models necessarily requires specifying a reaction function. For instance in monetary policy models, this reaction function often takes the form of a Taylor-type rule, a rule specifying how the policy rate is set as a function of inflation and the output gap. In practice however, policy makers do not mechanically follow simple Taylor-type rules, and they respond to much more information than captured by a few variables.[3] In fact, in many practical policy settings, e.g. fiscal or climate, the reaction function is rarely explicitly modeled, discussed, let alone known, yet it is clear that decisions are based on some information set. As such, the reaction function can be viewed as a complex function that incorporates a large information set, but is unlikely that it can be written down explicitly.[4]

Our approach avoids needing to know the reaction function by effectively approaching the evaluation of policy makers from a hypothesis testing perspective (e.g. Hausman, 1978; Breusch and Pagan, 1980; Hansen, 1982). The adopted score testing principal fixes the reaction function under the null and avoids the need to know or estimate this function. Specification tests have been used extensively in the econometrics literature to evaluate models and methods, but not in the context of policy evaluation.

In general, comparing policy makers based on their reaction functions necessarily excludes exogenous, non-systematic, mistakes that policy makers may also make from the analysis. Comparisons based on the optimal policy perturbation statistic of Barnichon and Mesters (2022) across different policy makers can be used to also incorporate such shocks, but computing this statistic requires the conditional forecasts of the policy maker, which are not always available. Here we are solely interested in comparing the average, systematic behavior of policy makers, i.e. reaction function comparisons.

From a general macro perspective, this paper uncovers an important but so far overlooked link between the structural impulse response literature (e.g. Ramey, 2016) and the optimal policy literature. While impulse response estimates have primarily been used as a guide for model building (e.g., Ramey, 2016), our paper provides a novel and important role for impulse response estimates: as a testbed for the optimality of reaction functions.

Related, recent works have shown that impulse response estimates are more generally useful in the context of policy making. Specifically, Barnichon and Mesters (2022) highlight how impulse responses can form the basis of a sufficient statistic approach to improving policy decisions with minimal modeling assumptions, while McKay and Wolf (2022) show how impulse responses can be used to construct any policy rule counter-factual. Both works require the identification of policy news shocks to compute optimal policy rules or paths. Here our objectives are different as we are solely interested in evaluating policy makers, and

[3]As Svensson puts it, "An optimal policy responds to all relevant state variables (including all relevant information), and there are many more relevant state variables and much more relevant information than current inflation and output" (e.g. Svensson, 2003).

[4]Policy makers repeatedly voice their concern that simple algebraic instrument rules are too simple to capture the complexity of the underlying economy. For instance, in the context of monetary policy Svensson (2017) writes "Taylor-type rules are too restrictive and mechanical, not taking into account all relevant information, and the ability to handle the complex and changing situations faced by policy makers". Algebraic rules cannot capture ex-ante all relevant contingencies, and a lot of information may simply be "non-rulable" (Kocherlakota, 2016; Blinder, 2016). See Blanchard (2018); Blanchard, Leandro and Zettelmeyer (2020) for similar arguments in the context of fiscal policy.

not in optimal policy making. For this particular task we show that all existing evidence for the identification and estimation of impulse response functions can be used to evaluate the performance of policy makers.

2 Some examples

To provide the intuition for our approach we informally present two examples to illustrate how we can evaluate and compare policy makers based on their reaction function without having access to the specification of the underlying economic model nor access to their reaction function. The first example is the baseline New Keynesian model where the optimal policy is defined under discretion, see Galí (2015, Section 5.1.1). In this example optimal policy is well understood allowing us to highlight the main mechanisms linking back to the broad NK literature (e.g. Galí, 2015). The second example is based on the more econometric structural VAR model (e.g. Sims, 1982; Kilian and Lütkepohl, 2017), which allows us to highlight modifications that arise when multiple policy and non-policy shocks, as well as dynamics, are present.

2.1 Three equation New Keynesian model

We consider evaluating a monetary policy maker based on the standard loss function

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

with π_t the inflation gap and x_t the output gap. The log-linearized baseline New-Keynesian model (Galí, 2015) is defined by a Phillips curve and an intertemporal (IS) curve given by

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

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

with i_t the nominal interest rate set by the central bank and ξ_t an iid cost-push shock with variance σ_ξ^2.

To illustrate our approach we consider a simple policy rule augmented with a policy shock ϵ_t, i.e.,

$$i_t = \phi_\pi \pi_t + \phi_\xi \xi_t + \epsilon_t \ , \tag{4}$$

where $\phi = (\phi_\pi, \phi_\xi)$ is the reaction function which captures the systematic response of the central bank. We assume that the structural shocks ξ_t and ϵ_t have mean zero and are mutually uncorrelated.[5]

Equilibrium allocation

For any $\phi_\pi > 1$ we can write the model for $Y_t = (\pi_t, x_t)'$ and i_t as

$$Y_t = \Gamma\xi_t + \mathcal{R}\epsilon_t \qquad \text{and} \qquad i_t = \Theta_\xi \xi_t + \Theta_\epsilon \epsilon_t \ , \tag{5}$$

with

$$\Gamma = \begin{bmatrix} \frac{1-\kappa\phi_\xi/\sigma}{1+\kappa\phi_\pi/\sigma} \\ \frac{-\phi_\pi/\sigma-\phi_\xi/\sigma}{1+\kappa\phi_\pi/\sigma} \end{bmatrix} \ , \quad \mathcal{R} = \begin{bmatrix} \frac{-\kappa/\sigma}{1+\kappa\phi_\pi/\sigma} \\ \frac{-1/\sigma}{1+\kappa\phi_\pi/\sigma} \end{bmatrix} \ , \quad \Theta_\xi = \frac{\phi_\pi + \phi_\xi}{1+\kappa\phi/\sigma} \quad \text{and} \quad \Theta_\epsilon = \frac{1}{1+\kappa\phi_\pi/\sigma} \ ,$$

where Γ and $\mathcal{R}$ capture the responses of the structural shocks ξ_t and ϵ_t on the policy objectives. Similarly, Θ_ξ and Θ_ϵ capture the equilibrium relationship between the structural shocks and the interest rate. Note that each term depends on the reaction function ϕ.

For future reference it is useful to make explicit that representation (5) arises from premultiplying the display below by D^{-1}.

$$\underbrace{\begin{bmatrix} 1 & -\kappa & 0 \\ 0 & 1 & 1/\sigma \\ -\phi_\pi & 0 & 1 \end{bmatrix}}_{=D} \begin{bmatrix} \pi_t \\ x_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \phi_\xi \end{bmatrix} \xi_t + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \epsilon_t \ . \tag{6}$$

The point to note is that innovations to the reaction function, in this case ϵ_t, enter the equilibrium representation (5) via the expression $D^{-1}(0,0,1)'$.

Optimal reaction function

An optimal reaction function is defined as any $\phi = (\phi_\pi, \phi_\xi)$ that minimizes the expected loss function subject to the equations that describe the economy. Formally, let Φ denote the subset of reaction functions that lead to a unique equilibrium. The set of optimal reaction functions is given by

$$\Phi^{\text{opt}} = \left\{ \phi \in \Phi : \phi \in \underset{\phi\in\Phi}{\operatorname{argmin}} \, \mathbb{E}\mathcal{L}_t \quad \text{s.t.} \quad (2) - (4) \right\} \ .$$

We note that since the economy (2)-(3) is linear and the reaction function (4) allows the policy maker to respond to all available information (i.e. all endogenous variables and struc-

[5] As we will see below, the benefit of including ξ_t is that it ensures the existence of a unique equilibrium under the optimality, irrespective of the parameter values in (2)-(3) (e.g. Galí, 2015, page 133).

tural shocks) the set Φ^{opt} is not a priori constrained.[6] In the vast majority of applications Φ^{opt} will include infinitely many reaction functions as optimal responses to structural shocks can often be interchanged with optimal responses to the endogenous variables. Note that we purposely allow for this possibility as we do not wish to impose any knowledge of the elements to which the policy maker responds.

Testing the reaction function

Consider a policy maker who uses the reaction function ϕ^0. We are interested in testing whether $\phi^0 \in \Phi^{\text{opt}}$. To construct a test statistic, we consider a thought experiment where the proposed policy rule (4) is modified with some change τ in response to the cost-push shock

$$i_t = \phi^0_\pi \pi_t + (\phi^0_\xi + \tau)\xi_t + \epsilon_t \ . \tag{7}$$

If for any $\tau \neq 0$ we are able to lower the expected loss function we may conclude that ϕ^0 was not optimal.

To verify whether this is the case, suppose that ϕ^0 leads to a unique equilibrium, i.e. $\phi^0 \in \Phi$. Then, following the same steps that led to (5) we have that under (7)

$$Y_t = (\Gamma^0 + \mathcal{R}^0 \tau)\xi_t + \mathcal{R}^0 \epsilon_t \ , \tag{8}$$

where Γ^0 and $\mathcal{R}^0$ denote the responses to the structural shocks under the rule ϕ^0 and are defined in (5). To see how this expression arises, consider the modified version of (6) under the augmented reaction function (7):

$$\underbrace{\begin{bmatrix} 1 & -\kappa & 0 \\ 0 & 1 & 1/\sigma \\ -\phi^0_\pi & 0 & 1 \end{bmatrix}}_{=D} \begin{bmatrix} \pi_t \\ x_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \phi^0_\xi \end{bmatrix} \xi_t + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tau\xi_t + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \epsilon_t \ . \tag{9}$$

The key insight is that $\tau\xi_t$ enters the equilibrium allocation for $Y_t = (\pi_t, x_t)'$ in exactly the same way as ϵ_t, i.e. via $D^{-1}(0,0,1)'$. This is understandable when viewing $\tau\xi_t$ as an innovation to the policy rule, which in equilibrium has effect $\mathcal{R}^0$ on the policy objectives. Indeed any innovation to the instrument has an effect proportional to $\mathcal{R}^0$ in equilibrium, the monetary shock is only a special case, which is usually normalized to have a unit effect on i_t.

Now, as mentioned, if $\phi^0 \in \Phi^{\text{opt}}$ there should not be any $\tau \neq 0$ that is able to lower the

[6] For instance, if the output gap x_t was affected by an additional shock, say a productivity shock, we would include the output gap and the productivity shock in the policy rule as well.

loss function. A necessary condition for this is that the gradient of the loss function with respect to τ evaluated at $\tau = 0$ should be zero. This leads to the following equivalence

$$\phi^0 \in \Phi^{\text{opt}} \qquad \Longleftrightarrow \qquad \nabla_\tau \mathbb{E}\mathcal{L}_t\big|_{\tau=0} \propto \mathcal{R}^{0\prime}\Gamma^0 = 0 \ . \tag{10}$$

This is a key result of this paper: if and only if the reaction function is optimal the impulse responses of the supply shock on the policy objectives Γ^0 should be orthogonal to the impulse responses of the policy shock on the policy objectives $\mathcal{R}^0$. Intuitively, since Γ^0 and $\mathcal{R}^0$ depend on ϕ^0 they encode the information on the reaction function. To see that $\mathcal{R}^{0\prime}\Gamma^0 = 0$ implies $\phi^0 \in \Phi^{\text{opt}}$, we note that the second derivative is given by $\nabla^2_\tau \mathbb{E}\mathcal{L}_t = \sigma^2_\xi \mathcal{R}'\mathcal{R} > 0$ and hence a global minimum (non-unique) is obtained at ϕ^0 and thus $\phi^0 \in \Phi^{\text{opt}}$.

Intuitively, the inner product $\mathcal{R}^{0\prime}\Gamma^0$ captures exactly (i) what the policy maker did on average to offset the cost-push shock, i.e. Γ^0, and (ii) what the policy maker could have done to offset the cost-push shock, i.e. $\mathcal{R}^0$. Noting that the latter follows from our previous observation that changes in the response to the reaction function, i.e. changes in ϕ_ξ, have an equilibrium effect on Y_t that is proportional to $\mathcal{R}^0$.

In practice, we can use different macro econometric methods and identification approaches to estimate the impulse responses Γ^0 and $\mathcal{R}^0$ (e.g. Ramey, 2016). We emphasize that these methods typically only require the estimation of a reduced form econometric model in combination with some identification strategy, e.g. short-run, long-run, sign, heteroskedasticity, non-Gaussianity, etc. This implies that for testing whether $\Gamma^{0\prime}\mathcal{R}^0 = 0$ we do not need to know the specific underlying structural model nor the reaction function ϕ^0.

That said, in this example we can easily verify the optimality condition as we know that the optimal reaction function takes the form $\phi^{\text{opt}}_\xi = (\kappa\sigma - \phi^{\text{opt}}_\pi)/(1+\kappa^2)$ for any $\phi^{\text{opt}}_\pi > 1$, see Galí (2015, page 133). Writing out $\mathcal{R}^{0\prime}\Gamma^0$ under any $\phi^0 = \phi^{\text{opt}}$ gives

$$\begin{aligned}
\Gamma^{0\prime}\mathcal{R}^0 &= (1+\kappa\phi^0_\pi/\sigma)^{-2}[-\kappa/\sigma \ , \ -1/\sigma] \begin{bmatrix} 1-\kappa\phi^0_\xi/\sigma \\ -\phi^0_\pi/\sigma - \phi^0_\xi/\sigma \end{bmatrix} \\
&= \frac{-\kappa/\sigma + \kappa/\sigma + \phi^0_\pi/\sigma^2 - \phi^0_\pi/\sigma^2}{(1+\kappa\phi^0_\pi/\sigma)^2} = 0 \ .
\end{aligned}$$

This mechanically shows that under an optimal rule the gradient condition holds.

Measuring the distance to optimality

Next, we show how the gradient with respect to the response to the cost-push shock can be exploited to measure the distance to the set of optimal reaction functions. We have

$$\nabla_\tau \mathbb{E}\mathcal{L}_t = \sigma^2_\xi \mathcal{R}^{0\prime}(\Gamma^0 + \mathcal{R}^0\tau) \ .$$

Setting the gradient to zero and solving for τ gives the distance

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

The coefficient τ^* measures exactly how much more or less the policy maker should have responded to the cost-push shock in order to minimize the loss function. It is easy to see that $\phi^* = (\phi^0_\pi, \phi^0_\xi + \tau^*) \in \Phi^{\text{opt}}$. The distance τ^* is intuitively interpretable as the projection of Γ^0 (what the policy maker did) on $\mathcal{R}^0$ (what the policy maker could have done).

We note that if $\Gamma^0 = 0$ cost-push shocks do not affect Y_t and apart from policy mistakes ϵ_t the loss function is minimized. This shows that from the perspective of the policy maker *systematically* minimizing the loss function is equivalent to minimizing $\Gamma^{0\prime}\Gamma^0$. The policy maker can do this by responding differently to ξ_t which has an effect proportional to $\mathcal{R}^0$. In impulse response space the systematic policy problem becomes

$$\Gamma^0 = -\mathcal{R}^0\tau + \Delta \ ,$$

where the minus sign is because we aim to minimize the variation in Γ^0 and Δ is some remainder.[7] The distance to the optimal reaction function is measured by the least squares projection coefficient τ^*.

Comparing reaction functions

The results so far can be used to compare the reaction functions of different policy makers. To avoid excessive notation at this stage, consider comparing two policy makers that used reaction functions ϕ^0 and ϕ^1, respectively, and let the economic environment that they faced by captured by the parameters θ^0 and θ^1, respectively, which include all coefficients in the Phillips (2) and IS (3) curves.

For each policy maker we compute the distance to the set of optimal reaction functions, noting that since the economic environments are different these sets are not the same. We have

$$\tau_0^* = -(\mathcal{R}^{0\prime}\mathcal{R}^0)^{-1}\mathcal{R}^{0\prime}\Gamma^0 \qquad \text{and} \qquad \tau_1^* = -(\mathcal{R}^{1\prime}\mathcal{R}^1)^{-1}\mathcal{R}^{1\prime}\Gamma^1 \ .$$

The key insight is that while the environments are different, τ_0^* and τ_1^* provide measures for the same quantity: the distance to the optimal reaction function as measured in units of response to the structural shocks. It is important to emphasize that the impulse responses take care of possible changes in the volatility of structural shocks, i.e. each impulse response

[7] The regression $\Gamma^0 = -\mathcal{R}^0\tau + \Delta$ can be viewed as an artificial regression in the sense of Davidson and MacKinnon (2001). Testing $\phi^0 \in \Phi^{\text{opt}}$ by verifying $\tau = 0$ is analog to an artificial regression based specification test, see Davidson and MacKinnon (2004, Chapter 15) for more examples.

is defined as the response to a one-unit change in the policy or non-policy shock.

In sum, this example illustrates how we can evaluate and compare policy makers based on their reaction function without knowing the reaction function. The proposed evaluation metric τ is invariant to different economic environments.

2.2 Structural VAR model

The next example that we consider is the structural VAR model (e.g. Sims, 1982). While SVAR models are not uniformly robust to the Lucas (1976) critique, they remain a useful tool for macro policy making (e.g. Antolin-Diaz, Petrella and Rubio-Ramírez, 2021) and are commonly used for impulse response estimation under different identification strategies. As such we view it useful to illustrate how our reaction function testing approach applies in a general SVAR model. We will not repeat all intuition obtained from the New Keynesian example, but constrain ourselves to highlighting the implications of dynamic responses and the presence of multiple policy and non-policy shocks.

Let the M-dimensional vector of policy targets Y_t be aggregated in the loss function

$$\mathcal{L}_t = Y_t' Y_t \ . \tag{11}$$

Suppose the policy maker has $K < M$ policy instruments P_t available to minimize the loss function. The economy for $X_t = (Y_t', P_t')'$ is given by the SVAR model

$$A(L)X_t = e_t \ , \qquad \text{where} \qquad A(L) = A_0 - A_1 L - \ldots - A_q L^q \ , \tag{12}$$

where L denotes the lag operator and $e_t = (\xi_t', \varepsilon_t')'$ is the vector of structural shocks; $\xi_t \in \mathbb{R}^M$ capturing the non-policy shocks and $\varepsilon_t \in \mathbb{R}^K$ the policy shocks. We assume that these shocks are mean zero and mutually uncorrelated.

For exposition purposes we will assume that the researcher is only able to identify one non-policy shock $\xi_{1,t}$ and one policy shock $\varepsilon_{1,t}$. The equation for the first policy instrument is given by

$$a_{p_1 p_1} P_{1,t} + a_{p_1 p_{2:K}} P_{2:K,t} + a_{p_1 y} Y_t = a_{p_1 x,1} X_{t-1} + \ldots + a_{p_1 x,q} X_{t-q} + \varepsilon_{1,t} \ , \tag{13}$$

Let $\phi = \{a_{pp}, a_{py}, a_{px,1}, a_{px,q}\}$ denote the reaction function of the policy maker of which the coefficients in (13) are a subset. Since, not all shocks can be identified we cannot estimate ϕ nor the subset in (13) as, for instance, $a_{p_1 y}$ cannot be recovered.

Let Φ be the set of reaction functions for which (12) can be inverted. For any $\phi \in \Phi$ we get

$$X_t = C(L)e_t \qquad \text{with} \quad C(L) = \begin{bmatrix} \Gamma_1(L) & \Gamma_{2:M}(L) & \mathcal{R}_1(L) & \mathcal{R}_{2:K}(L) \\ \Theta_{\xi_1}(L) & \Theta_{\xi_{2:M}}(L) & \Theta_{\varepsilon_1}(L) & \Theta_{\varepsilon_{2:K}}(L) \end{bmatrix} , \tag{14}$$

where $C(L)$ is the moving-average polynomial. The set of optimal reaction function is defined similar as before $\Phi^{\text{opt}} = \{\phi \in \Phi : \phi \in \arg\min_{\phi \in \Phi} \mathbb{E}\mathcal{L}_t\}$.

Next, consider a policy maker who proposes $\phi^0 \in \Phi$, we aim to measure the distance of ϕ^0 to Φ^{opt} in the direction of the response to $\xi_{1,t}$.[8] To measure this distance we consider the augmented policy reaction function for $P_{1,t}$

$$a^0_{p_1p_1} P_{1,t} + a^0_{p_1p_{2:K}} P_{2:K,t} + a^0_{p_1y} Y_t = a^0_{p_1x,1} X_{t-1} + \ldots + a^0_{p_1x,q} X_{t-q} + \tau \xi_{1,t} + \varepsilon_{1,t} \ , \tag{15}$$

where τ is a constant that measures the additional response to $\xi_{1,t}$. The moving average representation for the policy objectives Y_t under (15) is given by

$$Y_t = \left(\Gamma^0_1(L) + \tau \mathcal{R}^0_1(L)\right) \xi_{1,t} + \Gamma^0_{2:M}(L)\xi_{2:M,t} + \mathcal{R}^0(L)\varepsilon_t \ , \tag{16}$$

where again we see that the equilibrium effect of adjusting by τ is proportional to the, now dynamic, effect of the policy shock $\mathcal{R}^0_1(L)$.

To know whether ϕ^0 is optimal we compute the gradient of the loss function with respect to τ. We have

$$\nabla_\tau \mathbb{E}\mathcal{L}_t = \sigma^2_{\xi_1} \mathcal{R}^{0'}_1 (\Gamma^0_1 + \mathcal{R}^0_1 \tau)$$

where $\mathcal{R}^0$ and Γ^0_1 are the impulse response coefficients in the polynomials $\mathcal{R}_1(L)$ and $\Gamma_1(L)$ that capture the effect of $\varepsilon_{1,t}$ and $\xi_{1,t}$ on $Y_t, Y_{t+1}, \ldots$ under ϕ^0. It follows that at if $\phi^0 \in \Phi^{\text{opt}}$, the gradient with respect to τ evaluated at $\tau = 0$ should be zero, and we have that

$$\mathcal{R}^{0'}_1 \Gamma^0_1 = 0 \ .$$

The impulse responses of policy and non-policy shocks are orthogonal to each other. Moreover, we may set the gradient to zero to compute the distance to Φ^{opt} in the direction of $\xi_{1,t}$. We have

$$\tau^* = -(\mathcal{R}^{0'}_1 \mathcal{R}^0_1)^{-1} \mathcal{R}^{0'}_1 \Gamma^0_1$$

which is the distance under ϕ^0. This measure can again be compared across policy makers who have different reaction functions and operate in different environments.

[8] Note that in the SVAR model there is no coefficient that explicitly capture the response to $\xi_{1,t}$, instead the policy maker's response to $\xi_{1,t}$ follows from her responses to the endogenous variables Y_t. Obviously this does not prevent us from looking in the direction $\tau\xi_{1,t}$ directly.

We conclude that if the underlying economy can be written as a structural vector autoregressive model, we can evaluate and compare policy makers based on their reaction function by simply evaluating τ^* — a simple function of the impulse responses to policy and non-policy shocks —. The benefit of this approach is that it can be adopted even in scenarios where A cannot be entirely identified and the reaction function cannot be estimated. Moreover, any type of identification and estimation strategy can be used to recover $\mathcal{R}^0$ and Γ^0. Finally, it is easy to verify that the same approach holds for all time-varying parameter SVAR models (e.g. Primiceri, 2005), regime switching SVAR models (e.g. Sims and Zha, 2006) and state dependent SVAR models (e.g. Barnichon, Debortoli and Matthes, 2021).

3 Environment and objectives

In this section we describe our general framework and objectives. We restrict ourselves to linear-quadratic environments, which can be justified for small fluctuations around the steady-state. Clearly, this restriction rules out potentially interesting models and in Section 7 we carefully explore which further generalizations are feasible.

3.1 Environment

We consider a policy maker who aims to stabilize an $M_y \times 1$ vector y_{t+h} of gaps —deviations of the policy objectives from their targets— over horizons $h = 0, 1, \ldots$. Denote by $\mathbf{Y}_t = (y_t', y_{t+1}', \ldots)'$ the path of the vector of policy objectives y_t for a given time period t. The policy maker's loss function writes

$$\mathcal{L}_t = \mathbb{E}_t \mathbf{Y}_t' \mathcal{W} \mathbf{Y}_t \ , \tag{17}$$

where $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$, with $\mathcal{F}_t$ the time t information set. The matrix $\mathcal{W} = \text{diag}(\beta \otimes \lambda)$ includes preferences $\lambda = (\lambda_1, \ldots, \lambda_{M_y})'$ capturing the weights on the different variables and discount factors $\beta = (\beta_0, \beta_1 \ldots)'$ for the different horizons.

The policy maker has M_p instruments to minimize the loss function. Denote by $p_t = (p_{1,t}, \ldots, p_{M_{p},t})'$ the vector of instruments at t. Stacking the policy instruments, we get the policy path $\mathbf{P}_t = (p_t', p_{t+1}', \ldots)'$. A generic model for $\mathbf{Y}_t$ is completed by the path for the additional endogenous variables $\mathbf{W}_t = (w_t', w_{t+1}', \ldots)'$ and the paths of the structural shocks $\mathbf{\Xi}_t = (\xi_t, \xi_{t+1}, \cdots)'$. At time t the economy can be written as

$$\begin{cases} \mathcal{A}_{yy}\mathbb{E}_t\mathbf{Y}_t - \mathcal{A}_{yp}\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{yw}\mathbb{E}_t\mathbf{W}_t & = & \mathcal{B}_{y\xi}\mathbb{E}_t\mathbf{\Xi}_t \\ \mathcal{A}_{ww}\mathbb{E}_t\mathbf{W}_t - \mathcal{A}_{wp}\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{wy}\mathbb{E}_t\mathbf{Y}_t & = & \mathcal{B}_{w\xi}\mathbb{E}_t\mathbf{\Xi}_t \end{cases} , \tag{18}$$

where $\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ denote conformable linear maps. This model is general and accommodates many models found in the literature; prominent examples include New Keynesian models and more modern heterogeneous agents models. Additional exogenous variables can be easily included but we exclude them to simplify the exposition.

Turning to the policy block; we postulate that the policy maker follows a generic policy rule of the form

$$\mathcal{A}_{pp}\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{py}\mathbb{E}_t\mathbf{Y}_t - \mathcal{A}_{pw}\mathbb{E}_t\mathbf{W}_t = \mathcal{B}_{p\xi}\mathbb{E}_t\mathbf{\Xi}_t + \mathbb{E}_t\boldsymbol{\epsilon}_t \ , \tag{19}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_t', \epsilon_{t+1}', \ldots)'$ a sequence of policy shocks and $\mathbb{E}_t\boldsymbol{\epsilon}_t$ transforms the policy shocks into policy news shocks, i.e. exogenous shocks to the future path of $\mathbf{P}_t$ that are released at time t. The policy rule (19) imposes no restrictions as it allows the policy maker to respond to all available variables and shocks in the economy. We collect all parameters of the non-policy block in $\theta = \{\mathcal{A}_{yy}, \mathcal{A}_{yp}, \mathcal{A}_{yw}, \mathcal{A}_{ww}, \mathcal{A}_{wp}, \mathcal{A}_{wy}, \mathcal{B}_{y\xi}, \mathcal{B}_{w\xi}\}$, which can be thought of as describing the economic environment.

We collect the coefficients of the policy rule in $\phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{A}_{pw}, \mathcal{B}_{p\xi}\}$ and refer to these as the reaction function. We denote by Φ the set of all policy rules ϕ for which the model (18)-(19) implies a unique equilibrium. In addition, we normalize all news shocks — $\mathbb{E}_t\mathbf{\Xi}_t$ and $\mathbb{E}_t\boldsymbol{\epsilon}_t$ — to be unconditionally mean zero and mutually uncorrelated. The variance map of $\mathbb{E}_t\mathbf{\Xi}_t$ is given by $\mathbf{\Sigma}_\Xi = \mathbb{E}\mathbf{\Xi}_t\mathbf{\Xi}_t'$.

We define the set of optimal reaction functions as follows

$$\Phi^{\text{opt}} = \left\{ \phi \in \Phi : \phi \in \underset{\phi\in\Phi}{\operatorname{argmin}}\, \mathbb{E}\mathcal{L}_t \quad \text{s.t} \quad (18) - (19) \right\} \ . \tag{20}$$

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

3.2 Objectives

We have two different yet related objectives. First, in Section 4 we consider a policy maker with reaction function ϕ^0 and we test whether ϕ^0 is optimal, i.e. we consider

$$H_0 : \phi^0 \in \Phi^{\text{opt}} \qquad \text{vs.} \qquad H_1 : \phi^0 \notin \Phi^{\text{opt}} \ . \tag{21}$$

We first derive relevant conditions for conducting this test under the assumption that we can identify all policy $\mathbb{E}_t\boldsymbol{\epsilon}_t$ and non-policy $\mathbb{E}_t\mathbf{\Xi}_t$ shocks. Next, we relax this assumption and propose conditions for the practical scenario where we can only identify some subsets of the shocks.

Second, with the testing framework in place we turn in Section 5 to quantifying the distance to optimality in the direction of responses to non-policy shocks. The distance can be used as a comparative metric for ranking the performance of policy makers.

4 Reaction function optimality tests

In this section we consider testing $H_0 : \phi^0 \in \Phi^{\text{opt}}$ given that the economy admits the generic representation (18)-(19). We first establish the main orthogonality conditions in population that will subsequently be used to test the optimality of the reaction function in finite samples.

4.1 Orthogonality conditions

We start by noting that under $H_0 : \phi^0 \in \Phi^{\text{opt}}$ we can write the policy objectives in terms of the contemporaneous shocks and the news shocks.

$$\mathbb{E}_t \mathbf{Y}_t = \Gamma^0 \mathbb{E}_t \mathbf{\Xi}_t + \mathcal{R}^0 \mathbb{E}_t \boldsymbol{\epsilon}_t \ , \tag{22}$$

where the maps Γ^0 and $\mathcal{R}^0$ capture the causal effects of the structural shocks $\mathbb{E}_t \mathbf{\Xi}_t$ and $\mathbb{E}_t \boldsymbol{\epsilon}_t$ under ϕ^0, respectively. The precise mapping from the model coefficients $\underset{..}{\mathcal{A}}^0$ and $\underset{..}{\mathcal{B}}^0$ to Γ^0 and $\mathcal{R}^0$ is provided in the appendix, but we will not require knowledge of this mapping.

Given the notation established in representation (22) the following proposition establishes the key result.

Proposition 1. *Given the generic model* (18)-(19) *we have that*

$$\phi^0 \in \Phi^{\text{opt}} \qquad \Longleftrightarrow \qquad \mathcal{R}^{0\prime} \mathcal{W} \Gamma^0 = 0 \ , \tag{23}$$

where Φ^{opt} *is defined in* (20).

The proof is shown in the appendix but follows exactly the same steps as in the simple examples above.

Expression (23) shows that under an optimal policy rule, the impulse response Γ^0 of the non-policy shocks on the policy objectives should be orthogonal to $\mathcal{R}^0$, the impulse responses of the policy shocks on the policy objectives. Intuitively, if a policy maker follows an optimal reaction function, the reaction function of the policy maker should have transformed the effects of the non-policy shocks such that there is no more the policy maker can do to lower the loss: the impulse responses to non-policy shocks should be orthogonal to the impulse response to changes in policy.

The key reason we do not need to know or estimate the functional form of the reaction function is that the effect of that reaction function is *already* encoded in the impulse responses Γ^0 and $\mathcal{R}^0$. Thus estimating Γ^0 (along with $\mathcal{R}^0$) is enough to assess the optimality of the policy reaction function ϕ^0.

More generally, condition (23) shows that it is not necessary to know the full structure of underlying model nor to estimate the underlying reaction function. Instead, two statistics are sufficient to assess the optimality of a rule: (i) the impulse responses to changes in policy, and (ii) the impulse responses to structural shocks.

Subset orthogonality condition

One obvious difficulty in practice is that there exists insufficient exogenous variation to estimate the causal effects $\mathcal{R}^0$ and Γ^0. For instance, estimating Γ^0 would require being able to identify all non-policy shocks in the economy.

Fortunately, this is not necessary as we can test the optimality of the reaction function by considering any subset of causal effects that can be estimated. As an example let us take one row of $\mathcal{R}^{0\prime}$, i.e. the impulse response of $\mathbf{Y}$ to a specific policy instrument, and one column of Γ^0, i.e. the impulse response of $\mathbf{Y}$ to a specific structural shock. Then we must have that $\mathcal{R}_i^{\prime}\mathcal{W}\Gamma_j = 0$. Which shows that any combination of policy instruments and structural shocks can be tested in isolation.

More generally, denote by $\mathcal{R}_a$ the effects of $\mathbb{E}_t\boldsymbol{\epsilon}_{a,t}$ on $\mathbb{E}_t\mathbf{Y}_t$ where $\mathbb{E}_t\boldsymbol{\epsilon}_{a,t}$ is a vector formed from any subset or linear combination of the policy news shocks in $\mathbb{E}_t\boldsymbol{\epsilon}_{a,t}$. Similarly, denote by Γ_b the impulse responses of $\mathbb{E}_t\mathbf{Y}_t$ to a subset $\mathbb{E}_t\boldsymbol{\Xi}_{t,b}$ (or linear combination) of the shocks $\mathbb{E}_t\boldsymbol{\Xi}$. The key requirement for the shocks in $\mathbb{E}_t\boldsymbol{\Xi}_b$ is that they are non-policy shocks, but it is not strictly necessary for them to have a specific interpretation.[9]

With this notation the following corollary summarizes the implication of a given subset orthogonality condition.

Corollary 1. *Given the generic model* (18)-(19) *we have that*

$$\mathcal{R}_a^{0\prime}\mathcal{W}\Gamma_b^0 \neq 0 \qquad \Longrightarrow \qquad \phi^0 \notin \Phi^{\text{opt}} \ . \tag{24}$$

The result is of great practically relevant as it shows that policy makers and researchers never have to estimate the entire causal maps $\mathcal{R}$ and Γ to test the reaction function. For instance, suppose a monetary policy maker is interested in testing the reaction of the short term interest rate to an oil price shock, then only the impulse responses of the policy objec-

[9]For instance, if a researcher is sure that a particular sequence of residuals does not depend on the policy shocks such sequence can be used to form $\mathbb{E}_t\boldsymbol{\Xi}_{t,b}$, see Gali and Gambetti (2020) for an approach along these lines.

tives, say inflation and unemployment, to interest rate shocks and oil price shocks are needed. Ideally, in this scenario we would include all interest rate news shocks, but in practice we can conduct a subset test after only identifying the contemporaneous policy shock.

4.2 Implementing reaction function tests

In this section we discuss how the necessary condition for the optimality of the reaction function can be implemented in practice. For practical relevance we focus on the testing the subset orthogonality condition $\mathcal{R}_a^{0\prime}\mathcal{W}\Gamma_b^0 = 0$ for the case where $\mathcal{R}_a^0$ and Γ_b^0 correspond to a finite subset or finite linear combination of policy and non-policy shocks, respectively. Corollary 1 implies that $\mathcal{R}_a^{0\prime}\mathcal{W}\Gamma_b^0 \neq 0$ implies that ϕ^0 is not optimal.

To implement the test we assume that are given a baseline sample $\{y_t, p_t\}_{t=1}^n$, and in some applications perhaps a subset of additional endogenous variables $\{w_t\}_{t=1}^n$ or instrumental variables $\{z_t\}_{t=1}^n$.

In general, we require that the economy was in a stable regime for the sampling period $t = 1, \ldots, n$, such that the causal effects $\mathcal{R}_a^0$ and Γ_b^0 are time-invariant. We relax this assumption in Section 7 below. Further, we note that any preferred identification and estimation method for the impulse responses can in principal be used. The exposition here outlines one possible approach based on using instrumental variables for identification.

Following empirical practice we truncate the relevant horizon at H and define $\mathbf{Y}_{t:t+H} = (y_t', \ldots, y_{t+H}')'$ as the truncated policy objectives at time t. The causal effects $\mathcal{R}_a$ and Γ_b can be estimated based on (22) by considering

$$\mathbf{Y}_{t:t+H} = \mathcal{R}_a^0 \mathbb{E}_t \boldsymbol{\epsilon}_{a,t} + \Gamma_b^0 \mathbb{E}_t \boldsymbol{\Xi}_{b,t} + \mathbf{U}_{t:t+H} \ , \tag{25}$$

where $\mathbf{U}_{t:t+H}$ includes all other structural shocks, both policy and non-policy shocks, as well as the forecast errors $\mathbf{Y}_{t:t+H} - \mathbb{E}_t \mathbf{Y}_{t:t+H}$. Note that we slightly abuse notation by not indexing the truncated matrices $\mathcal{R}_a^0$ and Γ_b^0 by H.

Clearly, if $\mathbb{E}_t \boldsymbol{\epsilon}_{a,t}$ and $\mathbb{E}_t \boldsymbol{\Xi}_{b,t}$ are observable we can simply use OLS to estimate the desired response, but in practice we typically only observe some proxies for these shocks, say $\boldsymbol{z}_{a,t}$ and $\boldsymbol{z}_{b,t}$ which can be used as instruments if they only correlate with $\mathbb{E}_t \boldsymbol{\epsilon}_{a,t}$ and $\mathbb{E}_t \boldsymbol{\Xi}_{b,t}$, respectively. Following Stock and Watson (2018), under the normalization that $\mathbb{E}_t \boldsymbol{\epsilon}_{a,t}$ has a unit effect on $\mathbb{E}_t \mathbf{P}_{a,t}$ and $\mathbb{E}_t \boldsymbol{\Xi}_{b,t}$ has unit effect on some other endogenous variables $\mathbb{E}_t \mathbf{W}_{b,t}$, we can replace the shocks by the endogenous variables and use the proxies as instruments to estimate $\mathcal{R}_a^0$ and Γ_b^0. Specifically, we may consider

$$\mathbf{Y}_{t:t+H} = \mathcal{R}_a^0 \mathbb{E}_t \mathbf{P}_{a,t} + \Gamma_b^0 \mathbb{E}_t \mathbf{W}_{b,t} + \mathbf{U}_{t:t+H} \ .$$

The resulting IV estimates $\widehat{\mathcal{R}}_a^0$ and $\widehat{\Gamma}_b^0$ can be used to evaluate the orthogonality conditions $\widehat{\mathcal{R}}_a^{0\prime}\mathcal{W}\widehat{\Gamma}_b^0$, and confidence bands can be constructed using either asymptotic theory based on the delta method to compute the variance of $\text{vec}(\widehat{\mathcal{R}}_a^{0\prime}\mathcal{W}\widehat{\Gamma}_b^0)$, or using bootstrap methods, see Montiel Olea and Plagborg-Møller (2021) for an efficient approach.

In addition, we may want to regularize such regression estimates using a structural VAR framework (e.g. Plagborg-Møller and Wolf, 2021) or by directly penalizing the impulse responses $\mathcal{R}_a^0$ and Γ_b^0 following (Barnichon and Brownlees, 2018). Li, Plagbørg-Møller and Wolf (2022) show simulation results that compare a variety of regularization approaches in their ability to improve finite sample inference.

Ignoring measurement error

In the setting where we only have access to two proxies, one for the policy shock $\mathbb{E}_t\epsilon_{a,t}$ and for one non-policy shock $\mathbb{E}_t\xi_{b,t}$ we can ignore the measurement error in the proxies when constructing the test. Here the shocks $\mathbb{E}_t\epsilon_{a,t}$ and $\mathbb{E}_t\xi_{b,t}$ may also be a linear combination of the shocks $\mathbb{E}_t\boldsymbol{\epsilon}_{t:t+H}$ and $\mathbb{E}_t\boldsymbol{\Xi}_{t:t+H}$, but we suppress this for ease of notation.[10]

To see this we note that the population IV estimators for $\mathcal{R}_a^0$ and Γ_b^0 are in this case given by

$$\mathcal{R}_a^{0,IV} = \frac{\text{Cov}(\mathbf{Y}_{t:t+H}, z_{a,t})}{\text{Cov}(\mathbb{E}_t p_{a,t}, z_{a,t})} \qquad \text{and} \qquad \Gamma_b^{0,IV} = \frac{\text{Cov}(\mathbf{Y}_{t:t+H}, z_{b,t})}{\text{Cov}(\mathbb{E}_t w_{b,t}, z_{b,t})} \ ,$$

Note that the denominators in both expressions are scalar and non-zero if we assume that the instruments are relevant. This implies that the subset orthogonality condition under the null $\mathcal{R}_a'\mathcal{W}\Gamma_b^0 = 0$, implies that we must have

$$\text{Cov}(z_{a,t}, \mathbf{Y}_{t+H})\mathcal{W}\text{Cov}(\mathbf{Y}_{t:t+H}, z_{b,t}) = 0 \ .$$

We can test the orthogonality condition by replacing the population covariances by their sample counterparts. This test statistic has the benefit that its limiting distribution does not rely on the strength of the correlation between $z_{a,t}$ and $\mathbb{E}_t p_{a,t}$, nor $z_{b,t}$ and $\mathbb{E}_t w_{b,t}$.[11] In a VAR framework with the proxy variables ordered first, or a VARX framework, this implies that we do not need to rescale the impulse responses to the structural shocks by the impulse responses to $p_{a,t}$ and $w_{b,t}$ (e.g. Paul, 2019; Plagborg-Møller and Wolf, 2021).

[10] This is important for empirical applications as many of the identified proxies, think for instance the high frequency identified monetary policy shocks of Kuttner (2001) are in fact a linear combination of horizon specific shocks.

[11] Such test is the natural analog to the reduced form test for instrumental variable estimation, see Chernozhukov and Hansen (2008).

5 Ranking policy makers

In this section we consider comparing the performance of among two or more policy makers based on the same loss function. We measure performance by computing the nearest distance to the set of optimal reaction function in the direction of the response to the shocks $\mathbb{E}_t\mathbf{\Xi}_{b,t}$. This can be done by computing the adjustment to the response to the shocks $\mathbb{E}_t\mathbf{\Xi}_{t,b}$ that ensures that the gradient is equal to zero.

To formalize the result let $\mathcal{B}_{p_a\xi_b}$ denote the subset of $\mathcal{B}_{p\xi}$ that captures the response of $\mathbb{E}_t\mathbf{\Xi}_{b,t}$ on $\mathbb{E}_t\mathbf{P}_{a,t}$ in (19), and $\mathcal{B}_{p_{-a}\xi_{-b}}$ captures the remaining responses. This allows to formalize the following result.

Corollary 2. *Given the generic model* (18)-(19) *we have that the subset gradient condition* (24) *can be satisfied by setting*

$$\mathcal{R}_a^{0'}\mathcal{W}(\Gamma_b^0 + \mathcal{R}_a^0\mathcal{T}_{a,b}^*) = 0 \qquad \textit{where} \qquad \mathcal{T}_{a,b}^* = -(\mathcal{R}_a^{0'}\mathcal{W}\mathcal{R}_a^0)^{-1}\mathcal{R}_a^{0'}\mathcal{W}\Gamma_b^0 \ , \tag{26}$$

and

$$\mathbb{E}\mathcal{L}_t(\phi^*) \leq \mathbb{E}\mathcal{L}_t(\phi^0) \ ,$$

with $\phi^* = \{\mathcal{A}_{pp}^0, \mathcal{A}_{py}^0, \mathcal{A}_{pw}^0, \mathcal{B}_{p_a\xi_b}^0 + \mathcal{T}_{a,b}^*, \mathcal{B}_{p_{-a}\xi_{-b}}^0\}$.

The rescaled subset gradient $\mathcal{T}_{a,b}^*$ measures the distance —in units of the policy instruments— to the set of optimal reaction functions in the direction of $B_{p_a\xi_b}$. It tells exactly how much more or less the policy maker should have responded to the non-policy shocks $\mathbb{E}_t\mathbf{\Xi}_{b,t}$ using the policy instruments $\mathbb{E}_t\mathbf{P}_{a,t}$. We typically report this measure together with the orthogonality condition as it allows to comment on the economic magnitude of the error in the reaction function ϕ^0.

Moreover, we use such distance measures to compare the performance of different policy makers based on their reaction function. Specifically, for each policy maker we may compute $\mathcal{T}_{a,b}^*$ and compare how far these matrices are from zero. This measures the difference in terms of units of the policy instrument, i.e. for one extra unit of the non-policy shocks how much additional change in the policy instrument is needed. In addition, we may rescale such distance measures to make them interpretable in terms of units of the outcome variables. Specifically, $\mathcal{R}_a^0\mathcal{T}_{a,b}^*$ measures the distance to local optimality in terms of the units of the observable variables, and we can compare such measures across different policy makers.

Formally, let $\mathcal{T}_{a,b}^*(j)$ be the distance to optimality under the reaction function ϕ^j and the model parameters θ^j as stemming from (18). We understand $\mathcal{T}_{a,b}^*(j)$, for $j = 1, \ldots, d$, as the distance statistics for d different policy makers.[12]

[12] Alternatively, $\mathcal{T}_{a,b}^*(j)$'s may correspond to the same policy maker in different time periods.

We rank policy makers based on the entries of $\mathcal{T}^*_{a,b}(j)$, for each combination of instrument and non-policy shock we then obtain a different ranking. Each ranking is informative about a specific dimension of policy. To obtain an overall ranking we can simply sum the entries of $\mathcal{T}^*_{a,b}(j)$. When we look at the difference between entries of say $\mathcal{T}^*_{a,b}(1)$ and $\mathcal{T}^*_{a,b}(2)$ we measure the relative distance to optimality under the reaction functions ϕ^1 and ϕ^2, given the environments θ^1 and θ^2.

Care must be taken when ranking policy makers based on $\mathcal{T}^*_{a,b}(j)$ as the test only pertains to shocks $\mathbb{E}_t\boldsymbol{\epsilon}_{a,t}$ and $\mathbb{E}_t\boldsymbol{\Xi}_{b,t}$. For instance, a reaction function ϕ^1 can be more optimal for off-setting shocks $\mathbb{E}_t\boldsymbol{\Xi}_{b,t}$, but less so for other shocks. However, and important to remember, by adjusting ϕ^j based on $\mathcal{T}^*_{a,b}(j)$ as in Corollary 2 we can always lower the loss function, for any j.

Implementation

With the estimates and the joint distribution for the impulse responses $\widehat{\mathcal{R}}^0_a$ and $\widehat{\Gamma}^0_b$ we can also compute $\mathcal{T}^*_{a,b}(0)$ using the expression in Corollary 2. Specifically, using the bootstrap draws we can compute the mean and confidence set for

$$\mathcal{T}_{a,b}(0) = -(\mathcal{R}^{0\prime}_a \mathcal{W} \mathcal{R}^0_a)^{-1} \mathcal{R}^{0\prime}_a \mathcal{W} \Gamma^0_b$$

We note that simply plugging in the estimates for $\mathcal{R}^{0\prime}_a$ and Γ^0_b is also asymptotically valid, but result in a finite sample bias as such plug-in estimate ignores the variance of $\mathcal{R}^{0\prime}_a \mathcal{W} \mathcal{R}^0_a$, see the discussion in Barnichon and Mesters (2022).

It of often attractive to compute the distance measures for multiple policy makers simultaneously, as formal comparison tests can be conducted to assess whether certain distances were significantly larger. To illustrate, suppose that ϕ^1 was implemented over the periods $t = 1, \ldots, n_0$ and ϕ^2 over $t = n_0, \ldots, n$. The regression model (25) can be extended as follows

$$\begin{aligned}\mathbf{Y}_{t:t+H} &= \left[\mathbf{1}(t \leq n_0)\mathcal{R}^1_a + \mathbf{1}(t > n_0)\mathcal{R}^2_a\right] \mathbb{E}_t\boldsymbol{\epsilon}_{a,t} \\ &\quad + \left[\mathbf{1}(t \leq n_0)\Gamma^1_b + \mathbf{1}(t > n_0)\Gamma^2_b\right] \mathbb{E}_t\boldsymbol{\Xi}_{b,t} + \mathbf{U}_{t:t+H} \ , \qquad t = 1, \ldots, n \ , \qquad (27)\end{aligned}$$

where $\mathbf{1}()$ is the indicator function.

With this representation in hand the comparison tests can be implemented following the approach discussed in the previous section. We can estimate the impulse responses using any desired identification and estimation method. We note that it easier to estimate them jointly as this incorporates the correlation among the estimates which is requires for constructing confidence sets and comparison tests.

6 Testing the Fed reaction function

To illustrate our reaction function optimality tests, we evaluate the Fed's reaction function over the past 60 years. Notably, we revisit and refine the influential findings of (Judd and Rudebusch, 1998; Taylor, 1999; Clarida, Galí and Gertler, 2000) that concluded that the Fed conduct of monetary policy improved after 1980, because it satisfied the Taylor principle —a central bank should react more than one-to-one in the fact of inflation movements—. Based on Taylor rule estimates, these papers found that the parameters of the Taylor rule shifted around 1980 and that the Fed responded more vigorously to inflation variations after 1979, though this conclusion has not gone unchallenged (e.g., Orphanides, 2003). However, beyond that Taylor principle, that literature can say little about the optimality of the reaction function, whether the central bank was reacting too much or (still) too little after 1980.

We evaluate the Fed reaction function by testing the optimal response of the Fed to two different shocks: (i) oil price shocks and (ii) technology shocks.

As loss function we posit the usual dual inflation-unemployment mandate

$$\mathcal{L}_t = \frac{1}{2}\mathbb{E}_t\left(\|\Pi_{t:t+H}\|^2 + \|U_{t:t+H}\|^2\right) \ , \tag{28}$$

with $\Pi_{t:t+H} = (\pi_t - \pi^*, \ldots, \pi_{t+H} - \pi^*)'$ the vector of inflation gaps and $U_{t:t+H} = (u_t - u_t^*, \ldots, u_{t+H} - u_{t+H}^*)'$ the vector of unemployment gaps. Note that (28) implies a weighting matrix $\mathcal{W} = I$, reflecting the Fed's "balanced approach" between its two mandates, and we set the discount factor $\beta = 1$.

Implementation details

To implement the optimal reaction function tests we require two sets of impulse responses: the (i) causal effects of policy changes on π_t and u_t ($\mathcal{R}_a$), and (ii) the impulse responses of π_t and u_t to non-policy shocks (Γ_b).

To estimate $\mathcal{R}_a$, we rely on the Romer and Romer (2004) narrative measure of exogenous monetary policy changes to identify the impulse responses to changes in the current policy rate. The Romer and Romer series has the advantage of covering the longest sample period thanks to Tenreyro and Thwaites (2016)'s extension of the Romer and Romer series (1969-2007).

To estimate Γ_b — the impulse responses of π_t and u_t to oil supply shocks and technology shocks, we rely on two series of narratively identified shock proxies. As oil price shocks proxy we use the Hamilton (2003) oil shocks series, and as technology shocks proxy we use the Fernald (2012) series.

Main results

Focusing first on the Fed's reaction to oil price shocks, Figures 1 and 2 plot the impulse responses estimated using quarterly data over two sub-samples: 1969-1985 and 1986-2007. In each figure, the top row plots the effects of oil price shocks on inflation and unemployment, while the bottom row plots the dynamic causal effects of changing the fed funds rate on inflation and unemployment.

Prior 1985, we can see that monetary policy has significant effects on both unemployment and inflation, though with substantial inertia. In addition, note how the response of inflation substantially lags that of unemployment, reaching its peak effect after 10 quarters (see also Barnichon and Mesters, 2021). Turning to the effect of an oil price shock, we can see that the response of inflation is rather rapid and short-lived, particularly in th epost-85 perdior. Thus, while can reject that the Fed's reaction to oil price shocks was appropriate at the 68% confidence level in the pre-1985 period —not raising the fed funds rate enough in response to the increase in oil prices— (Figure 1), we can no longer reject optimality in the post-1985 period (Figure 2).

Turning to the Fed's reaction to TFP shocks, Figures 3 and 4 plot the impulse responses estimated using quarterly data over two sub-samples: 1969-1985 and 1986-2007. In the pre-1985 period, a positive TFP shock leads to a decline in both unemployment and inflation. Moreover, notice how the the impulse response of unemployment to a change in policy overlaps with the impulse response of unemployment to a TFP shock. As a result, the impulse responses are not orthogonal and we can reject the optimality of the Fed's reaction function in the face of TFP shocks. In the face of the negative technology shocks of the 1970s, the Fed should have further raised interest rates ($\tau^* = -0.8$). Post-1985 however, the impulse responses are orthogonal ($\tau^* \approx -0.8$) and we cannot reject optimality of the reaction function.

7 Extensions

The previous sections showed that we can implement reaction function optimality and comparison tests as long as the underlying model admits a representation like (18)-(19). In this section we discuss some further extensions that allow the method to be used for several generalizations of this linear model. Specifically, we discuss extensions with time-varying parameters (e.g. Paul, 2019), state dependence (e.g. Auerbach and Gorodnichenko, 2013) and multiple policy regimes (e.g. Sims and Zha, 2006). We discuss the extensions in the context of reaction function optimality tests, noting that comparison tests follow naturally.

7.1 Time varying parameters

The baseline model (18)-(19) can be extended to allow for time-varying parameters.

$$\left\{\begin{array}{rcl} \mathcal{A}_{yy,t}\mathbb{E}_t\mathbf{Y}_t - \mathcal{A}_{yp,t}\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{yw,t}\mathbb{E}_t\mathbf{W}_t & = & \mathcal{B}_{y\xi,t}\mathbb{E}_t\boldsymbol{\Xi}_t \\ \mathcal{A}_{ww,t}\mathbb{E}_t\mathbf{W}_t - \mathcal{A}_{wp,t}\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{wy,t}\mathbb{E}_t\mathbf{Y}_t & = & \mathcal{B}_{w\xi,t}\mathbb{E}_t\boldsymbol{\Xi}_t \\ \mathcal{A}_{pp,t}\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{py,t}\mathbb{E}_t\mathbf{Y}_t - \mathcal{A}_{pw,t}\mathbb{E}_t\mathbf{W}_t & = & \mathcal{B}_{p\xi,t}\mathbb{E}_t\boldsymbol{\Xi}_t + \mathbb{E}_t\boldsymbol{\epsilon}_t \ , \end{array}\right. , \tag{29}$$

where now the maps $\mathcal{A}_{\cdot\cdot,t}$ and $\mathcal{B}_{\cdot\cdot,t}$ are allowed to vary over time. We generally restrict the source of time-variation to be exogenous and unrelated to the structural shocks (e.g. Cogley and Sargent, 2001, 2005; Primiceri, 2005). The time varying reaction function is now given by $\phi_t = \{\mathcal{A}_{pp,t}, \mathcal{A}_{py,t}, \mathcal{A}_{pw,t}, \mathcal{B}_{p\xi,t}\}$ and the objective is to test whether some given reaction function ϕ_t^0 was optimal in the sense that it minimized the loss function $\mathbb{E}\mathcal{L}_t$. If $\phi_t^0 \in \Phi$, i.e. the set of reaction functions that lead to a unique equilibrium, we can write the policy objectives as

$$\mathbb{E}_t\mathbf{Y}_t = \mathcal{R}_t^0\mathbb{E}_t\boldsymbol{\epsilon}_t + \Gamma_t^0\mathbb{E}_t\boldsymbol{\Xi}_t \ ,$$

where now the impulse responses to the policy and non-policy shocks are time-varying. The following corollary establishes that we can continue use the orthogonality condition to test whether the reaction function was optimally set.

Corollary 3. *Given the generic time-varying model* (29) *we have that*

$$\phi_t^0 \in \Phi_t^{\text{opt}} \qquad \Longleftrightarrow \qquad \mathcal{R}_t^{0\prime}\mathcal{W}\Gamma_t^0 = 0 \ , \tag{30}$$

where Φ_t^{opt} *is defined as*

$$\Phi_t^{\text{opt}} = \left\{\phi_t \in \Phi : \phi_t \in \underset{\phi_t\in\Phi}{\operatorname{argmin}}\, \mathbb{E}\mathcal{L}_t \quad s.t \quad (29)\right\} \ .$$

The proof is similar to the proof of Proposition 1. The key insight is that the same orthogonality condition $\mathcal{R}_t^{0\prime}\mathcal{W}\Gamma_t^0 = 0$ is necessary and sufficient to determine whether the reaction function was optimally set if the coefficients of the general model were given by the time t maps $\mathcal{A}_{\cdot\cdot,t}$ and $\mathcal{B}_{\cdot\cdot,t}$. It follow immediately that failure to satisfy any subset orthogonality condition $\mathcal{R}_{a,t}^{0\prime}\mathcal{W}\Gamma_{b,t}^0 \neq 0$ implies that $\phi_t^0 \notin \Phi_t^{\text{opt}}$.

Implementing subset tests for $\mathcal{R}_{a,t}^{0\prime}\mathcal{W}\Gamma_{b,t}^0 = 0$ can be done using any of the time-varying parameter SVAR methods that have been developed (e.g. Cogley and Sargent, 2001, 2005; Primiceri, 2005; Petrova, 2019; Paul, 2019). These methods usually adopt a random walk specification for the coefficients which fits in our main requirement that the variation the parameters is driven by exogenous forces. Alternatively, time-varying parameter local projection methods can equally well be adopted (e.g. Ruisi, 2019; Lusompa, 2021). Typically, such time-varying parameter approaches are Bayesian in nature such that credible sets for $\mathcal{R}^{0\prime}_{a,t}\mathcal{W}\Gamma^0_{b,t} = 0$, or subsets hereof, can be easily constructed.

7.2 State Dependence

Time-variation macro models is often believed to be caused by state dependence and numerous works have documented evidence for various forms of state dependence in the effects of fiscal and monetary policy (e.g. Auerbach and Gorodnichenko, 2012, 2013; Tenreyro and Thwaites, 2016; Ramey and Zubairy, 2018; Barnichon, Debortoli and Matthes, 2021; Ascari and Haber, 2021; Eichenbaum, Rebelo and Wong, 2022). The common thread in such works is that state dependence is governed by some time-t pre-determined variable that is independent of the reaction function.

Our main results continue to hold in this setting, and the two impulse response estimates $\mathcal{R}^0$ and Γ^0 are necessary and sufficient to test the reaction function. The only difference is that $\mathcal{R}^0$ and Γ^0 need to be conditioned on the state of the economy. As an illustration, consider the specification of Auerbach and Gorodnichenko (2013) where the economy can be in two states, depending on the value of some state variable z_t. In that case, the generic model (18)-(19) is modified in that the maps $\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ become functions of z_t, i.e., $\mathcal{A}_{..}(z_t)$ and $\mathcal{B}_{..}(z_t)$.

With two states, each map can be written as $\mathcal{A}_{..}(z_t) = F(z_t)\mathcal{A}_{..(1)} + (1-F(z_t))\mathcal{A}_{..(2)}$ where $F(z_t)$ can be interpreted as a measure of probability of being in state 1 at time t based on some time t predetermined variable z_t.[13] The causal effects of the policy and non-policy shocks will then depend on the state of the economy $\mathcal{R}^0(z_t)$ and $\Gamma^0(z_t)$. The orthogonality condition $\mathcal{R}^0(z_t)\mathcal{W}\Gamma^0(z_t) = 0$ continues to have the same properties as in proposition 1, and subset orthogonality conditions can be used to test whether the reaction function was optimal for any state z_t of the economy.

8 Conclusion

In this paper, we showed that it is possible to evaluate unspecified reaction functions, because the effect of an (unspecified) reaction function is already encoded in the impulse responses to non-policy shocks, which are estimable.

Under an optimal reaction function, the impulse responses to policy shocks should be orthogonal the impulse responses to non-policy shocks. By exploiting this property, we construct reaction function optimality tests that can be used to detect and improve non-optimal reaction functions.

[13] A popular functional form for $F(.)$ is $F(z_t) = \exp(-\gamma z_t)/[1 + \exp(-\gamma z_t)]$ with γ a tuning parameter.

Appendix

A: Proofs

Proof of Proposition 1. Let $\mathcal{T}$ be a linear map, sufficiently small such that $\phi = \{\mathcal{A}^0_{pp}, \mathcal{A}^0_{py}, \mathcal{A}^0_{pw}, \mathcal{B}^0_{p\xi} + \mathcal{T}\} \in \Phi$. If $\phi^0 \in \Phi^{\text{opt}}$, $\mathbb{E}\mathcal{L}_t$ cannot be lowered by any $\mathcal{T} \neq 0$. Similar as in (22) we obtain the equilibrium representation

$$\mathbf{Y}_t = (\Gamma^0 + \mathcal{R}^0\mathcal{T})\mathbb{E}_t\mathbf{\Xi}_t + \mathbf{V}_t \ ,$$

where $\mathbf{V}_t = \mathcal{R}^0\mathbb{E}_t\boldsymbol{\epsilon}_t + \mathbf{Y}_t - \mathbb{E}_t\mathbf{Y}_t$ and note that $\mathbb{E}[\mathbb{E}_t\mathbf{\Xi}_t\mathbf{V}_t'] = 0$. The expected loss $\mathbb{E}\mathcal{L}_t$ becomes

$$\begin{aligned}\mathbb{E}\mathcal{L}_t &= \frac{1}{2}\mathbb{E}\left([\Gamma^0 + \mathcal{R}^0\mathcal{T}]\mathbb{E}_t\mathbf{\Xi}_t + \mathbf{V}_t\right)' \mathcal{W}\left([\Gamma^0 + \mathcal{R}^0\mathcal{T}]\mathbb{E}_t\mathbf{\Xi}_t + \mathbf{V}_t\right) \\ &= \frac{1}{2}\text{Tr}\left\{[\Gamma^0 + \mathcal{R}^0\mathcal{T}]'\mathcal{W}[\Gamma^0 + \mathcal{R}^0\mathcal{T}]\mathbf{\Sigma}_\Xi\right\} + \frac{1}{2}\mathbb{E}\left(\mathbf{V}_t'\mathcal{W}\mathbf{V}_t\right)\end{aligned}$$

The derivative of the map $\mathcal{T} \to \mathbb{E}\mathcal{L}_t$ is given by

$$\mathcal{R}^{0'}\mathcal{W}(\Gamma^0 + \mathcal{R}^0\mathcal{T})\mathbf{\Sigma}_\Xi \ . \tag{31}$$

Evaluating at $\mathcal{T} = 0$ and setting the derivative to zero implies

$$\mathcal{R}^{0'}\mathcal{W}\Gamma^0 = 0 \ ,$$

is a necessary condition for optimality. Noting that $\mathcal{T} \to \mathbb{E}\mathcal{L}_t$ is a convex map, it follows that $\mathcal{R}^0\mathcal{W}\Gamma^0 = 0$ is also sufficient for $\mathcal{T} = 0$ being a global minimizer, and thus $\phi^0 \in \Phi^{\text{opt}}$. □

Proof of Corollary 1. The right implication of proposition 1 implies the claim. □

Proof of Corollary 2. The left hand side of (26) is equivalent to the subset of (31) corresponding to a, b and after post-multiplying (26) by $\mathbf{\Sigma}_\Xi^{-1}$. The right hand side follows by direct calculation. The second part follows directly by noting $\mathcal{T}^*_{a,b}$ sets the local gradient to zero, hence it lowers the loss convex loss function. □

References

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

Ascari, Guido, and Timo Haber. 2021. "Non-Linearities, State-Dependent Prices and the Transmission Mechanism of Monetary Policy." *The Economic Journal*, 132(641): 37–57.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy*, 4: 1–27.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2013. "Output Spillovers from Fiscal Policy." *American Economic Review*, 103(3): 141–46.

Barnichon, Regis, and Christian Brownlees. 2018. "Impulse Response Estimation By Smooth Local Projections." *Review of Economics and Statistics.*

Barnichon, Regis, and Geert Mesters. 2021. "The Phillips Multiplier." *Journal of Monetary Economics*, 117: 689–705.

Barnichon, Regis, and Geert Mesters. 2022. "A Sufficient Statistics Approach to Macro Policy Evaluation." working paper.

Barnichon, Regis, Davide Debortoli, and Christian Matthes. 2021. "Understanding the Size of the Government Spending Multiplier: It's in the Sign." *The Review of Economic Studies*, 89(1): 87–117.

Blanchard, Olivier. 2018. "On the future of macroeconomic models." *Oxford Review of Economic Policy*, 34(1-2): 43–54.

Blanchard, Olivier, Alvaro Leandro, and Jeromin Zettelmeyer. 2020. "Redesigning EU fiscal rules: From rules to standards." *Economic Policy.*

Blinder, Alan. 2016. "Comment on: Rules versus discretion: A reconsideration." *Brookings Papers on Economic Activity*, 2016(2): 1–55.

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

Breusch, Trevor S., and Adrian R. Pagan. 1980. "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics." *The Review of Economic Studies*, 47(1): 239–253.

Chernozhukov, Victor, and Christian Hansen. 2008. "The reduced form: A simple approach to inference with weak instruments." *Economics Letters*, 100(1): 68–71.

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

Cogley, Timothy, and Thomas J. Sargent. 2001. "Evolving Post-World War II U.S. Inflation Dynamics." *NBER Macroeconomics Annual*, 16: 331–373.

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

Davidson, Russell, and James G. MacKinnon. 2001. "Chapter 1 - Artificial regressions." In . *Companion to Theoretical Econometrics*, , ed. B. Baltagi, 16 – 37. Blackwell.

Davidson, Russell, and James G. MacKinnon. 2004. *Econometric Theory and Methods.* New York:Oxford University Press.

Eichenbaum, Martin, Sergio Rebelo, and Arlene Wong. 2022. "State Dependent Effects of Monetary Policy: The Refinancing Channel." *American Economic Review*. forthcoming.

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

Fernald, John G. 2012. "A quarterly, utilization-adjusted series on total factor productivity." Federal Reserve Bank of San Francisco Working Paper Series 2012-19.

Galí, Jordi. 2015. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications.* Princeton University Press.

Gali, Jordi, and Luca Gambetti. 2020. "as the U.S. Wage Phillips Curve Flattened? A Semi-Structural Exploration." In . *Changing Inflation Dynamics, Evolving Monetary Policy*, , ed. G. Castex, J. Gali and D. Saravia, 149–172. Central Bank of Chile.

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.

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

Hansen, Lars Peter. 1982. "Large sample properties of generalized method of moments estimators." *Econometrica*, 50: 1029–1054.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica*, 46(6): 1251–1271.

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.

Kocherlakota, Narayana. 2016. “Rules versus discretion: A reconsideration.” *Brookings Papers on Economic Activity*, 2016(2): 1–55.

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.

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

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

Lusompa, Amaze. 2021. “Local Projections, Autocorrelation, and Efficiency.” Working paper.

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

Montiel Olea, José Luis, and Mikkel Plagborg-Møller. 2021. “Local Projection Inference Is Simpler and More Robust Than You Think.” *Econometrica*, 89(4): 1789–1823.

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

Paul, Pascal. 2019. “The Time-Varying Effect of Monetary Policy on Asset Prices.” *The Review of Economics and Statistics*. Forthcoming.

Petrova, Katerina. 2019. “A quasi-Bayesian local likelihood approach to time varying parameter VAR models.” *Journal of Econometrics*, 212(1): 286–306. Big Data in Dynamic Predictive Econometric Modeling.

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

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

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

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

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

Ruisi, G. 2019. "Time-varying local projections." Working paper.

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

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

Stock, James H., and Mark W. Watson. 2012. "Disentangling the Channels of the 2007-09 Recession." *Brookings Papers on Economic Activity*, 81.

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.

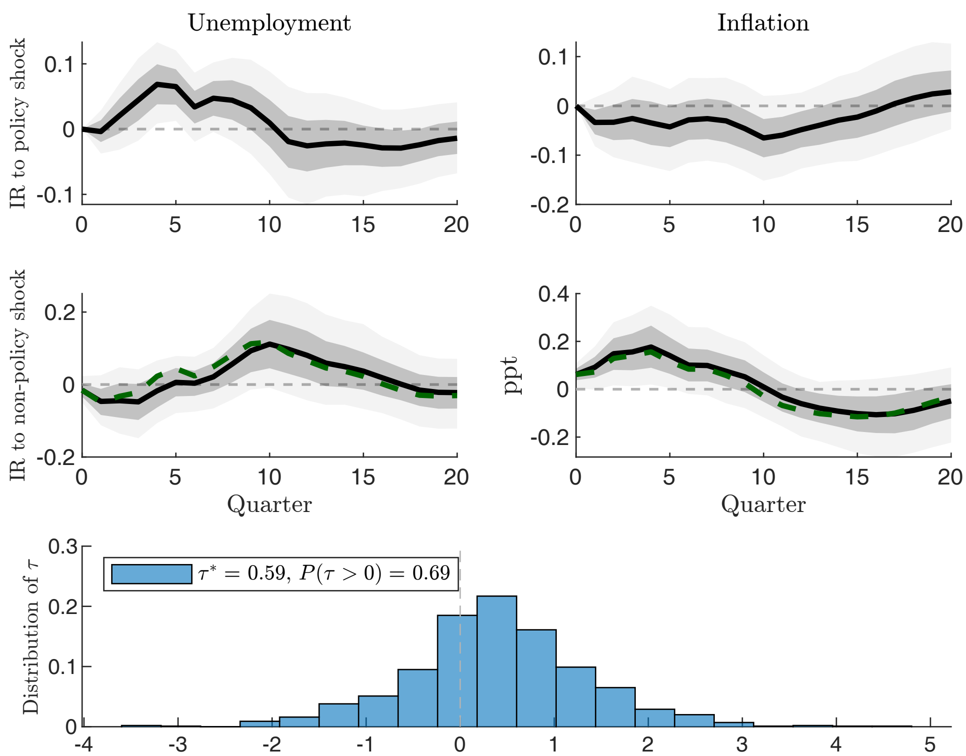
Svensson, Lars E.O. 2003. "What Is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules." *Journal of Economic Literature*, 41(2): 426–477.

Svensson, Lars E.O. 2017. "What Rule for the Federal Reserve? Forecast Targeting." National Bureau of Economic Research.

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

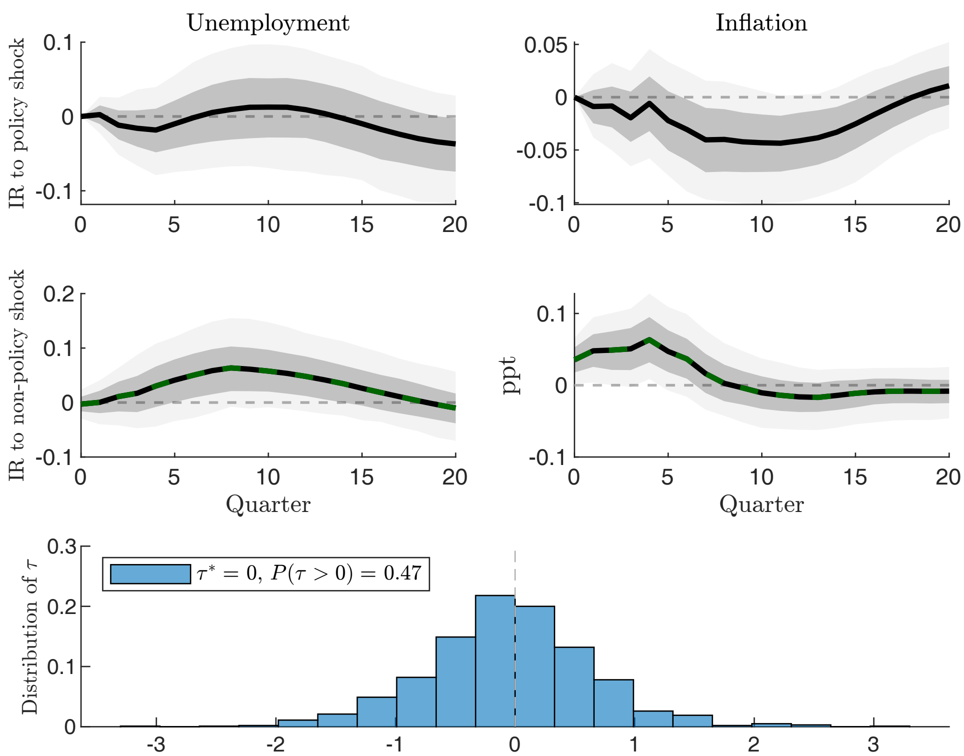
Tenreyro, Silvana, and Gregory Thwaites. 2016. "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions." *American Economic Journal: Macroeconomics*, 8(4): 43–74.

Figure 1: Impulse responses to an oil shock, pre-1985 data



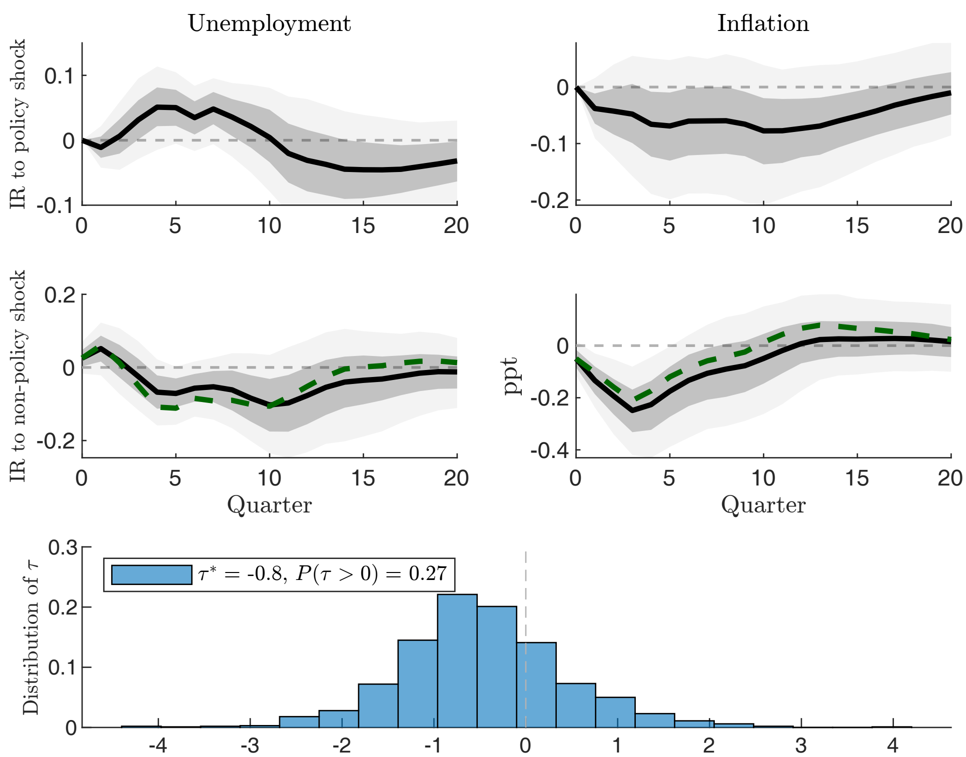
Notes: Top panel: impulse responses of π and u to a monetary policy shock ($\mathcal{R}$) identified using using Romer and Romer monetary shocks. Mid panel: impulse responses of π and u to an oil price shock (Γ) identified using the narrative series of Hamilton (2003). All impulse responses are estimated using pre-1985 data. The green lines denote the impulse responses after adjustment of the reaction function by τ^*, the mean of the statistic $\tau = -(\mathcal{R}'\mathcal{R})^{-1}\mathcal{R}'\Gamma$. Bottom panel: distribution of the τ statistic.

Figure 2: Impulse responses to an oil shock, post-1985 data



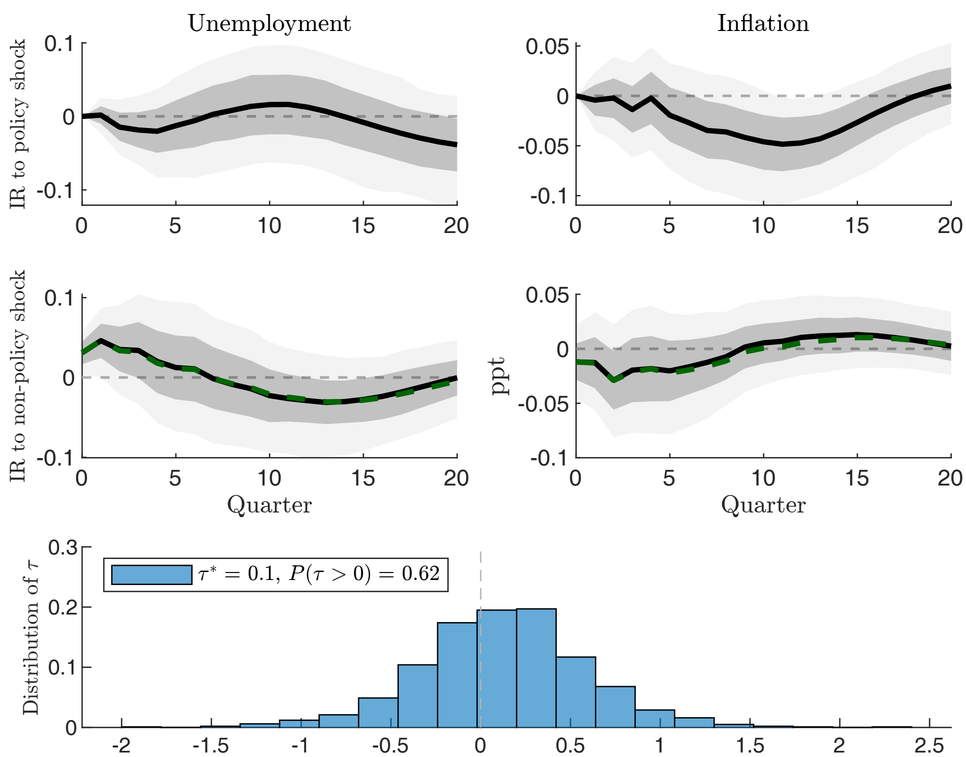
Notes: Top panel: impulse responses of π and u to a monetary policy shock ($\mathcal{R}$) identified using using Romer and Romer monetary shocks. Mid panel: impulse responses of π and u to an oil price shock (Γ) identified using the narrative series of Hamilton (2003). All impulse responses are estimated using post-1985 data. The green lines denote the impulse responses after adjustment of the reaction function by τ^*, the mean of the statistic $\tau = -(\mathcal{R}'\mathcal{R})^{-1}\mathcal{R}'\Gamma$. Bottom panel: distribution of the τ statistic.

Figure 3: Impulse responses to a TFP shock, pre-1985 data



Notes: Top panel: impulse responses of π and u to a monetary policy shock ($\mathcal{R}$) identified using using Romer and Romer monetary shocks. Mid panel: impulse responses of π and u to a TFP shock (Γ) from Fernald (2012). All impulse responses are estimated using pre-1985 data. The green lines denote the impulse responses after adjustment of the reaction function by τ^*, the mean of the statistic $\tau = -(\mathcal{R}'\mathcal{R})^{-1}\mathcal{R}'\Gamma$. Bottom panel: distribution of the τ statistic.

Figure 4: Impulse responses to a TFP shock, post-1985 data



Notes: Top panel: impulse responses of π and u to a monetary policy shock ($\mathcal{R}$) identified using using Romer and Romer monetary shocks. Mid panel: impulse responses of π and u to a TFP shock (Γ) from Fernald (2012). All impulse responses are estimated using post-1985 data. The green lines denote the impulse responses after adjustment of the reaction function by τ^*, the mean of the statistic $\tau = -(\mathcal{R}'\mathcal{R})^{-1}\mathcal{R}'\Gamma$. Bottom panel: distribution of the τ statistic.