What Can Time-Series Regressions Tell Us About Policy Counterfactuals?[†]

Alisdair McKay
FRB Minneapolis

Christian K. Wolf
MIT & NBER

This version: April 2022
First version: July 2021

Abstract: We show that, in a general family of linearized structural macroeconomic models, knowledge of the empirically estimable causal effects of contemporaneous and news shocks to the prevailing policy rule is sufficient to construct counterfactuals under alternative policy rules. If the researcher is willing to postulate a loss function, our results furthermore allow her to recover an optimal policy rule for that loss. Under our assumptions, the derived counterfactuals and optimal policies are fully robust to the Lucas critique. We then discuss strategies for applying these insights when only a limited amount of empirical causal evidence on policy shock transmission is available.

[†]Email: `alisdair.mckay@mpls.frb.org` and `ckwolf@mit.edu`. We received helpful comments from Isaiah Andrews, Marios Angeletos, Regis Barnichon, Marco Bassetto, Martin Beraja, Anmol Bhandari, Francesco Bianchi, Chris Carroll, Gabriel Chodorow-Reich, Hal Cole, Peter Ganong, Yuriy Gorodnichenko, Kilian Huber, Chen Lian, Geert Mesters, Ben Moll, Chris Moser, Emi Nakamura, Mikkel Plagborg-Møller, Pablo Ottonello, Richard Rogerson, Karthik Sastry, Jón Steinsson, Robert Ulbricht, Mike Waugh, Iván Werning, Tom Winberry, and seminar participants at various venues. We thank Christiane Baumeister and Bent Sørensen for valuable discussions. Jackson Mejia provided superb research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

An important function of macroeconomics is to predict the consequences of changes in policy. In this paper we revisit the role that evidence on policy *shocks*—that is, surprise deviations from a prevailing rule—can play in helping macroeconomists learn about policy *rule* counterfactuals. Existing work mainly uses such policy shocks in two ways. First, in what Christiano et al. (1999) call the "Lucas program", researchers first estimate the causal effects of a policy shock in the data, and then construct a structural macro model that matches these effects. This model is then trusted as a laboratory for predicting the effects of changes in policy rules. An alternative approach, proposed by Sims & Zha (2006), instead relies only on the estimated policy shock: the economy is subjected to a new policy shock at each date t, with the shocks chosen so that, t-by-t, the counterfactual rule holds.[1] Appealingly, this strategy does not require the researcher to commit to a model; on the other hand, it is subject to the Lucas critique: the private sector is surprised by the new policy shock at each t, rather than knowing at the initial date 0 that the policy rule has changed forever.

The contribution of this paper is to propose a method that constructs policy counterfactuals using empirical evidence on *multiple distinct* identified policy shocks, rather than just a single one. Like Sims & Zha (2006), the method does not rely on a particular parametric model structure; at the same time, for a family of models that nests many of those popular in the Lucas program, it yields counterfactuals that are robust to the Lucas critique. At the heart of our methodology lies an identification result. We prove that, for a relatively general family of macro models, the causal effects of contemporaneous as well as news shocks to a given policy rule are sufficient to construct Lucas critique-robust counterfactuals for alternative policy rules. The core intuition is that, by subjecting the economy to multiple distinct policy shocks at date 0 (rather than a new value of a single shock at each $t \geq 0$), we are able to enforce the contemplated counterfactual policy rule not just *ex post* along the equilibrium path (as done in Sims & Zha), but also *ex ante* in private-sector expectations. Under our assumptions, this suffices to sidestep the Lucas critique. While our exact identification result requires knowledge of the causal effects of a very large number of policy shocks, our proposed empirical method can be applied in the empirically relevant case of a researcher with access to only a couple of distinct shocks. We demonstrate the usefulness of the proposed approach with several applications to monetary policy rule counterfactuals.

[1] See Bernanke et al. (1997), Leeper & Zha (2003), Hamilton & Herrera (2004), Eberly et al. (2020), and Brunnermeier et al. (2021) for important extensions and recent applications of this method.

Identification result. The first part of the paper establishes the identification result. Our analysis builds on a general linear data-generating process, with one key added restriction: policy is allowed to affect private-sector behavior *only* through the current and future expected path of the policy instrument.[2] For example, for monetary policy, the private sector only cares about the expected future path of the nominal rate, and not whether this path is the result of the systematic component of policy—i.e., the policy *rule*—or due to *shocks* to a given rule. We consider an econometrician that lives in this economy and observes data generated under some baseline policy rule, where that baseline rule is subject to shocks. The econometrician then wishes to predict the effects of a switch to some alternative policy rule. Using standard time-series methods, she can estimate the causal effects of shocks to the prevailing policy rule (e.g. Ramey, 2016; Stock & Watson, 2018). Our identification result states that, if the econometrician has successfully estimated the effects of contemporaneous shocks to the policy rule as well as the effects of news about deviations from the rule *at all future horizons*, then those estimates contain all the information she needs to construct the counterfactual. Key to the proof is our assumption on how policy rules are allowed to shape private-sector behavior. Since only the expected future path of the policy instrument matters, any given *rule*—characterized by the instrument path that it implies—can equivalently be synthesized by adding well-chosen *shocks* to the baseline rule—all that is required is that those date-0 policy shocks imply the same instrument path from date-0 onwards as the counterfactual rule. Finally we show that, if additionally given a loss function, our econometrician can leverage the same logic to also characterize *optimal* policy.[3]

How general is the setting of this identification result? Our two key model restrictions are (i) linearity and (ii) the way that the policy instrument is allowed to shape private-sector behavior. We show that the key property (ii) is a feature shared by many standard linearized business-cycle models, including those with many frictions (Christiano et al., 2005), shocks (Smets & Wouters, 2007), and even rich micro heterogeneity (Kaplan et al., 2018; Ottonello & Winberry, 2020). Perhaps the most popular class of models violating our restriction is those with an asymmetry of information between policymaker and private sector, as in Lucas

[2]More precisely, the policy rule is allowed to matter only through (a) the expected path of the instrument and (b) equilibrium selection. Our assumptions on equilibrium existence and uniqueness for the different rules that we consider address equilibrium selection.

[3]To be clear, our identification results are entirely silent on the mapping from equilibrium outcomes to welfare, and so on the shape of loss functions. In particular, fully specified structural models are one way to arrive at such objective functions. However, given that objective functions in practice are often derived from a legislated mandate rather than economic theory, we believe it is useful to have a method of calculating optimal policy for an objective function that is taken as given.

(1972). In such models, private-sector agents solve a filtering problem, and the policy rule affects both the dynamics of the policy instrument as well as the information contained in that policy choice. In addition to this restriction (ii) on models, our linearity assumption (i) also limits the set of policy rule counterfactuals to which our method can be applied: our approach can be used to compare different cyclical stabilization policies (e.g., monetary or fiscal feedback rules), but is less well-suited to study policies that alter the steady state (e.g., changes in the inflation target or in the long-run fiscal system).

Counterfactuals with finitely many shocks. The main challenge to operationalizing our identification result is that empirical evidence on the effects of policy shocks is limited. Our theory says that we need to select a linear combination of time-0 policy shocks that perturbs the current and expected future path of the policy instrument exactly like the contemplated counterfactual rule. This is a daunting informational requirement: in general, to synthesize the effects of any possible policy instrument sequence of some length T, we would need access to T distinct identified policy shocks. While existing empirical evidence falls short of this ideal, recent research has however made progress on identifying the effects of at least *some* distinct policy shocks with rather different implications for future expected policy paths.[4] How much can be done with this available evidence?

The idea of our empirical method is to use the available evidence on policy shock transmission to provide a best Lucas critique-robust approximation to the desired policy counterfactual. Given estimates of the causal effects of some finite number n_s of distinct policy shocks, we face the challenge that our population identification result cannot be applied immediately: the counterfactual policy rule needs to hold in ex post equilibrium and ex ante expectation for a large number T of periods, but we only have access to $n_s \ll T$ shocks—more equations than unknowns. Our proposal is simply to choose the linear combination of date-0 shocks that enforces the desired counterfactual rule as well as possible, in a standard least-squares sense. Crucially, since this approach involves no ex post surprises dated $t = 1, 2, \ldots$, it is—under our assumptions—fully robust to Lucas critique concerns. Whether or not this best approximation is then in fact a sufficiently accurate representation of the desired counterfactual rule is invariably an application-dependent question.

[4] For monetary policy, different canonical shocks (e.g. Romer & Romer, 2004; Gertler & Karadi, 2015) lead to rather different responses of short-term rates. Other identification strategies explicitly aim to identify shocks at different parts of the yield curve (e.g. Gürkaynak et al., 2005; Inoue & Rossi, 2021). For fiscal policy, Ramey (2011) and Ramey & Zubairy (2018) estimate the effects of short-lived as well as more persistent shocks. Mertens & Ravn (2010) and Leeper et al. (2013) are similarly focussed on spending dynamics.

We demonstrate the uses and limitations of our method through several examples. Our object of interest is the propagation of a contractionary investment technology shock under different monetary rules. As the inputs to our method, we take the causal effects associated with two popular monetary policy shock series: those of Romer & Romer (2004) and Gertler & Karadi (2015). Armed with these causal effects, we apply our method to construct counterfactuals for alternative policy rules that: target the output gap; enforce a conventional Taylor-type rule; peg the nominal rate of interest; target nominal GDP; and minimize a simple dual-mandate loss function.[5] We find that, with the exception of the nominal rate peg, the counterfactual rules can be enforced to a quite high degree of accuracy. Our conclusion is that, at least for our particular investment shock, several rather different monetary policy rule counterfactuals can already be characterized quite sharply simply by combining existing pieces of empirical evidence on monetary policy shock transmission.

Finally, we note that our use of multiple distinct policy shocks also suggests a refinement of the original Sims & Zha procedure. In their approach, a researcher sets the value of a single policy shock at each t to enforce the counterfactual rule ex post, along the equilibrium path. Access to multiple shocks increases her degrees of freedom, allowing the rule to be enforced more with date-0 and less with date-$t > 0$ shocks, thus at least somewhat reducing the bite of expectations-related Lucas critique concerns. We use our monetary policy applications to also illustrate this hybrid of our method and the original proposal of Sims & Zha.

Counterfactuals with (partial) model structure. In some applications, it will not be possible to closely approximate the contemplated counterfactual rule through existing policy shock evidence. In that case a natural solution is to use a structural model to match the existing shock evidence, and then use the model to extrapolate to the effects of all other policy (news) shocks—standard impulse response matching as in e.g. Christiano et al. (2005), just now re-interpreted through the lens of our identification result. Our final contribution is to shed some light on how this extrapolation is achieved in "typical" models. We provide theoretical and quantitative results revealing that, in models that are popular in business-cycle analysis, the causal effects of contemporaneous and news policy shocks are often tightly linked. Our starkest example here is to show that—for a particular but important class of monetary policy counterfactuals—the required extrapolation of policy shock causal effects depends only on one model block: the Phillips curve. Researchers can use identified policy

[5]The fixed-rate counterfactual may also be interpreted as controlling for the endogenous response of monetary policy—a very popular counterfactual in Sims & Zha-type analyses.

shocks to estimate a parametric Phillips curve relationship, use the restrictions embedded in this Phillips curve to extrapolate from the identified shock to the effects of all other policy (news) shocks, and then construct rule counterfactuals using our identification results. We illustrate this observation with a return to some of our monetary policy rule applications.

Literature. Our identification result provides a bridge between the "Lucas program" as discussed in Christiano et al. (1999) and the empirical strategy of Sims & Zha (2006). By using multiple policy shocks at date 0 (rather than a single one at each $t \geq 0$) we are able to construct Lucas critique-robust counterfactuals without explicit model structure. Policy shock causal effects are thus model-robust "sufficient statistics" in the sense of Chetty (2009) and Nakamura & Steinsson (2018). Our results on causal effect extrapolation in "typical" macro models furthermore imply that individual policy shocks—while not sufficient statistics for all possible rule counterfactuals—can nevertheless serve as powerful "identified moments" (Nakamura & Steinsson, 2018) for model-based analysis of policy counterfactuals.

Our work also relates to other more recent contributions to counterfactual policy analysis. Beraja (2020) similarly forms policy counterfactuals without relying on particular parametric models. His approach relies on stronger exclusion restrictions in the non-policy block of the economy, but given those restrictions requires less empirical evidence on policy news shocks. Barnichon & Mesters (2021) use policy shock impulse responses to *test* the optimality of some given observed policy rule. We show that, under relatively mild additional structural assumptions, such policy shock impulse responses can in fact be used to fully characterize optimal policy rules for a given policymaker loss function.[6]

Finally, our identification result builds on recent advances in solution methods for structural macro models. At the heart of our analysis lies the fact that equilibria in such models can be characterized by matrices of impulse response functions (Auclert et al., 2021). As in Guren et al. (2021) and Wolf (2020), we connect this sequence-space representation to empirically estimable objects. In contemporaneous and independent work, De Groot et al. (2021) and Hebden & Winker (2021) show how to use similar arguments to efficiently *compute* policy counterfactuals by generating impulse responses to policy shocks from a structural model. Our focus is not computational—we aim to calculate policy counterfactuals directly from empirical evidence, forcing us to confront the fact that such evidence is limited.

[6]Kocherlakota (2019) presents a dynamic policy game in which the policymaker can select the optimal action via regression analysis. In his setting, the policy action does not cause the private sector to update its beliefs about the future strategy of the policymaker. Therefore policymaker payoffs only depend on the current policy choice and not on the future expected instrument paths that we emphasize in our analysis.

Outline. The remainder of the paper proceeds as follows. Section 2 presents the core identification result, mapping causal effects of policy *shocks* to counterfactuals for policy *rules*. Section 3 then introduces our empirical methodology and illustrates using applications to several monetary policy rule counterfactuals. Section 4 discusses how, in "typical" structural models, the causal effects of policy shocks at different horizons are linked. Section 5 concludes, and supplementary results are relegated to several appendices.

2 From policy shocks to policy rule counterfactuals

We begin in Section 2.1 by presenting a stylized version of our identification argument in a particular, familiar environment: the canonical three-equation New Keynesian model. We then in Sections 2.2 to 2.5 extend the argument to a general class of infinite-horizon linearized dynamic models and discuss its scope and limitations.

The main identification result is presented for a linearized perfect-foresight economy. Due to certainty equivalence, the equilibrium dynamics of a linear model with uncertainty will coincide with the solution to such a linearized perfect-foresight environment. We thus emphasize that all results presented below extend without any change to models with aggregate risk solved using conventional first-order perturbation techniques.[7]

2.1 A simple example

We begin with a discussion of our identification argument in the context of a simple and familiar model environment: the canonical three-equation New Keynesian model (Galí, 2015; Woodford, 2011). We also use this model to explain the relationship between our approach to constructing policy counterfactuals and that of Sims & Zha (2006).

Model. The variables of the economy are two private-sector aggregates—output y_t and inflation π_t—and a policy instrument—the nominal rate i_t. They are related through three equations: an Euler equation and a Phillips curve as the private-sector block,

$$y_t = y_{t+1} - \frac{1}{\gamma}(i_t - \pi_{t+1}), \tag{1}$$

$$\pi_t = \kappa y_t + \beta \pi_{t+1} + (\varepsilon_t + \theta \varepsilon_{t-1}), \tag{2}$$

[7] For example see Fernández-Villaverde et al. (2016), Boppart et al. (2018) or Auclert et al. (2021) for a detailed discussion of this point.

and a simple Taylor rule as the policy rule,

$$i_t \;=\; \phi\pi_t \;+\; \underbrace{\nu_{0,t}}_{\text{contemp. shock}} \;+\; \underbrace{\sum_{\ell=1}^{\infty}\nu_{\ell,t-\ell}}_{\text{news shocks}}. \tag{3}$$

In our perfect-foresight set-up, the two private-sector equations as well as the policy rule hold for $t = 0, 1, 2, \ldots$. These equations feature two kinds of disturbances. First, ε_t is a cost-push shock; for the illustrative analysis in this section, we will find it useful to assume that it induces a first-order moving average wedge in the Phillips curve (2). Second, there are the policy shocks $\nu_{\ell,t-\ell}$; here, $\nu_{0,t}$ is a conventional contemporaneous policy shock, while $\nu_{\ell,t-\ell}$ for $\ell > 0$ denotes a deviation from the policy rule at time t announced at $t - \ell$—an ℓ-period "news" shock. These policy shocks will turn out to be crucial for our identification result. As usual, given a vector of time-0 cost-push as well as policy (news) shocks $\{\varepsilon_0, \nu_{0,0}, \nu_{1,0}, \ldots\}$, a perfect-foresight transition path—or impulse response function—are paths of $\{y_t, \pi_t, i_t\}$ such that (1) - (3) all hold at all t.

For the subsequent analysis, the crucial property of this simple model economy is that the coefficients in the private-sector equations (1) - (2) are independent of the policy rule—i.e., γ, κ and β are unaffected by changes in ϕ. Equivalently, private-sector behavior is affected by policy only through the current and future values of the policy instrument i_t. Our general identification analysis in Sections 2.2 to 2.5 will discuss the generality and limitations of this crucial assumption.

Object of interest. Under the baseline policy rule, the impulse response of the economy to a cost-push shock is given as the solution of (1) - (3) for some cost-push shock ε_0 together with $\nu_{\ell,0} = 0$ for all ℓ. We wish to instead characterize the behavior of this economy in response to ε_0 not under the baseline policy rule (3), but instead under some counterfactual policy rule of the form

$$i_t = \tilde{\phi}\pi_t \tag{4}$$

where $\tilde{\phi} \neq \phi$. Note that this thought experiment supposes that the private sector perfectly understands the change in rule: the new relation between i and π holds at $t = 0, 1, 2, \ldots$. Our identification result characterizes the information required to construct this counterfactual.

The identification argument. We consider an econometrician living in our simple three-equation economy (1) - (3). Using conventional semi-structural time series methods

(Ramey, 2016), and with access to suitable identifying assumptions or instruments, that econometrician can in principle estimate the dynamic causal effects of the cost-push shock ε_t as well as the policy shocks $\{\nu_{\ell,t-\ell}\}_{\ell=0}^{\infty}$ under the baseline rule (3). Our main identification result states that this knowledge is sufficient to predict the counterfactual propagation of the shock ε_t under the alternative rule (4). While our formal result is stated and proved for a more general class of models in Sections 2.2 and 2.3, we here provide the core intuition using our simple three-equation model structure.

The key idea is to choose time-0 policy shocks $\nu_{\ell,0}$ to the baseline rule in order to mimic the desired counterfactual policy rule. To develop the argument, note first that, because our model has no endogenous state variables, the impulse response to a time-0 cost-push shock will die out after $t = 1$, by our assumption on shock persistence. We collect the 2×1 transition paths of $\{y_t, \pi_t, i_t\}$ in response to a cost-push shock ε_0 under the baseline rule as the vectors $\{\boldsymbol{y}_\phi(\varepsilon_0), \boldsymbol{\pi}_\phi(\varepsilon_0), \boldsymbol{i}_\phi(\varepsilon_0)\}$. Similarly, contemporaneous and one-period-ahead policy shocks also have no effects after $t = 1$. For $\ell \in \{0, 1\}$, we collect the corresponding 2×1 impulse responses under the baseline rule to a policy shock $\nu_{\ell,0}$ as the vectors $\{\Theta_{y,\nu_\ell,\phi}, \Theta_{\pi,\nu_\ell,\phi}, \Theta_{i,\nu_\ell,\phi}\} \times \nu_{\ell,0}$; e.g., $\Theta_{y,\nu_\ell,\phi}$ is the 2×1 impulse response path of y to an ℓ-period ahead shock to the baseline ϕ-rule (3). Now consider setting the two policy shocks to values $\{\tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ so that, under the baseline rule (3) and in response to the shocks $\{\varepsilon_0, \tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$, the counterfactual rule (4) holds at both $t = 0$ and $t = 1$ along the perfect foresight transition path; that is, we solve the following two equations in the two unknowns $\{\tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$

$$\boldsymbol{i}_\phi(\varepsilon_0) + \Theta_{i,\nu_0,\phi}\tilde{\nu}_{0,0} + \Theta_{i,\nu_1,\phi}\tilde{\nu}_{1,0} = \tilde{\phi} \times \left[\boldsymbol{\pi}_\phi(\varepsilon_0) + \Theta_{\pi,\nu_0,\phi}\tilde{\nu}_{0,0} + \Theta_{\pi,\nu_1,\phi}\tilde{\nu}_{1,0}\right]. \tag{5}$$

The left-hand side of this equation gives us the impulse response of the interest rate following our combination of three shocks $\{\varepsilon_0, \tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ under the baseline rule (3), while the right-hand side does the same for inflation, just scaled by $\tilde{\phi}$. By our informational assumptions, the econometrician can evaluate the system of equations (5) given ε_0 and for any candidate set of the two policy shocks $\{\tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$. Now suppose a solution $\{\tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ to (5) exists and compute the responses of $\{y_t, \pi_t, i_t\}$ to $\{\varepsilon_0, \tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ under the baseline policy rule. The content of our identification result is that those impulse responses are in fact identical to the impulse responses to ε_0 alone under the counterfactual rule (4). Crucially, this alternative computation uses *only* impulse responses under the baseline rule, and so in particular does not require direct knowledge of the structural equations (1)-(3).

The intuition underlying the identification result is straightforward. Since the private sector's decisions only depend on the expected path of the policy instrument $\{i_0, i_1, \dots\}$, it follows that it does not matter whether this path comes about due to the systematic conduct of policy or due to policy shocks. Equation (5) leverages this logic, looking for a combination of date-0 policy shocks that results in the counterfactual policy rule (4) holding both at $t = 0$ *and* in expectation at $t = 1$.

We emphasize that this argument inherently relies on knowledge of the causal effects of both the contemporaneous policy shock $\tilde{\nu}_{0,0}$ *as well as the policy news shock* $\tilde{\nu}_{1,0}$: it is only with those two that we can enforce the counterfactual rule along the entire transition path (which here consists of two time periods). With access only to the contemporaneous policy shock $\tilde{\nu}_{0,0}$, on the other hand, the researcher could only impose the counterfactual rule at $t = 0$, but not at $t = 1$. The method proposed by Sims & Zha (2006) is instead to subject the economy to another new surprise contemporaneous policy shock $\tilde{\nu}_{0,1}$ at $t = 1$; while this ex post enforces the counterfactual rule both at $t = 0$ and $t = 1$, the key difference is that the private-sector block did not at $t = 0$ *expect* the counterfactual rule to hold at $t = 1$; rather, the rule only holds at $t = 1$ because of yet another surprise. As a result, as long as policy at $t = 1$ matters for $t = 0$ decisions, the constructed counterfactual will differ from the true counterfactual $\{\boldsymbol{y}_{\tilde{\phi}}(\varepsilon_0), \boldsymbol{\pi}_{\tilde{\phi}}(\varepsilon_0), \boldsymbol{i}_{\tilde{\phi}}(\varepsilon_0)\}$. We will further elaborate on this connection between our identification result and the empirical methodology of Sims & Zha in Section 2.4.

Discussion & outlook. The identification result sketched in this section is special in two respects: first, it is presented within the context of a particular explicit structural model; and second, since impulse responses to ε_0 are non-zero only for two periods, the result required knowledge of the effects of two policy shocks. The remainder of this section will state and prove our main identification result in the context of a general class of infinite-horizon models. In terms of our informational requirements, the key change will be that the econometrician now needs to know the causal effects of *all* policy shocks $\{\nu_{\ell,0}\}_{\ell=0}^{\infty}$, rather than just the first two. The economic intuition on the other hand will be exactly the same: the argument will work as the long as the private-sector block depends on the policy rule only through the path of the policy instrument, as was the case here.

2.2 General model & objects of interest

We consider a linearized perfect-foresight, infinite-horizon model economy. Throughout, boldface denotes time paths for $t = 0, 1, 2, \ldots$, and all variables are expressed in deviations from the model's deterministic steady state.

The economy is summarized by the system

$$\mathcal{H}_w \boldsymbol{w} + \mathcal{H}_x \boldsymbol{x} + \mathcal{H}_z \boldsymbol{z} + \mathcal{H}_\varepsilon \boldsymbol{\varepsilon} = \mathbf{0} \tag{6}$$

$$\mathcal{A}_x \boldsymbol{x} + \mathcal{A}_z \boldsymbol{z} + \boldsymbol{\nu} = \mathbf{0} \tag{7}$$

w_t and x_t are n_w- and n_x-dimensional vectors of endogenous variables, z_t is a n_z-dimensional vector of policy instruments, ε_t is a n_ε-dimensional vector of exogenous structural shocks, and ν_t is an n_z-dimensional vector of policy shocks.[8] The distinction between w and x is that the variables in x are observable while the variables in w are not; specifically, x contains the outcomes of interest to the econometrician and the arguments of the counterfactual policy rule. The infinite-dimensional linear maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ summarize the non-policy block of the economy, yielding $n_w + n_x$ restrictions for each t. Our key assumption—echoing the simple model of Section 2.1—is that the maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ do not depend on the coefficients of the policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$; instead, policy only matters through the path of the instrument z, with the rule (7) giving n_z restrictions on the policy instruments for each t. As in our simple example, entries of the shock vectors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}$ for $t > 0$ should again be interpreted as news shocks. In particular, the policy shock vector $\boldsymbol{\nu}$ collects the full menu of contemporaneous and news shocks to the prevailing policy rule at all horizons, generalizing the two-shock set-up that was our focus in the simple three-equation model.

Given $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}\}$, an equilibrium is a set $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}\}$ that solves (6) - (7). We assume that the baseline rule $\{\mathcal{A}_x, \mathcal{A}_z\}$ is such that an equilibrium exists and is unique for any $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}\}$.

Assumption 1. *The policy rule in* (7) *induces a unique equilibrium. That is, the infinite-dimensional linear map*

$$\mathcal{B} \equiv \begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \mathcal{A}_x & \mathcal{A}_z \end{pmatrix}$$

is invertible.

Given $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}\}$, we write that unique solution as $\{\boldsymbol{w}_\mathcal{A}(\boldsymbol{\varepsilon}, \boldsymbol{\nu}), \boldsymbol{x}_\mathcal{A}(\boldsymbol{\varepsilon}, \boldsymbol{\nu}), \boldsymbol{z}_\mathcal{A}(\boldsymbol{\varepsilon}, \boldsymbol{\nu})\}$. As in the simple example, we often focus on impulse responses to exogenous shocks $\boldsymbol{\varepsilon}$ when the policy rule is followed perfectly ($\boldsymbol{\nu} = \mathbf{0}$); with some slight abuse of notation we will simply write those impulse responses as $\{\boldsymbol{w}_\mathcal{A}(\boldsymbol{\varepsilon}), \boldsymbol{x}_\mathcal{A}(\boldsymbol{\varepsilon}), \boldsymbol{z}_\mathcal{A}(\boldsymbol{\varepsilon})\}$.

[8]The boldface vectors $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\varepsilon}, \boldsymbol{\nu}\}$ stack the time paths for all variables (e.g., $\boldsymbol{x} = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_{n_x}')'$), and the linear maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ are conformable.

Discussion & scope. Our identification results in Section 2.3 and the empirical analysis in Section 3 will apply to *any* structural model that can be written in the general form (6) - (7). As emphasized before, in addition to linearity, the key property of the model for our purposes is that policy enters the non-policy block of the economy *only* through the realized path of the policy variables $\boldsymbol{z}$; equivalently, in the linearized economy with aggregate risk, policy matters only through its effects on the expected future path of the instrument z. How restrictive are those assumptions?

Our first observation is that many of the explicit, parametric structural models used for counterfactual and optimal policy analysis in the classical "Lucas program" literature (see Christiano et al., 1999) fit into our framework. Such models are routinely linearized, and their linear representation features the separation between policy rule and non-policy block that our results require. We illustrate this point by giving several examples of well-known models that are consistent with our assumptions. Our simple model in Section 2.1 has already illustrated that one particular canonical environment—the textbook three-equation New Keynesian model—fits into our framework.[9] By the same line of reasoning, even workhorse estimated business-cycle models (e.g. Christiano et al., 2005; Smets & Wouters, 2007) as well as recent quantitative HANK models (e.g. Auclert et al., 2020; McKay & Wieland, 2021) fit into our structure. For example, in standard HANK-type models, the standard Euler equation of the representative household is simply replaced by a more general "aggregate consumption function" (e.g. Auclert et al., 2018; Wolf, 2021):

$$\boldsymbol{c} = \mathcal{C}(\boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{i}, \boldsymbol{\varepsilon}^d) = \mathcal{C}_y \boldsymbol{y} + \mathcal{C}_\pi \boldsymbol{\pi} + \mathcal{C}_i \boldsymbol{i} + \boldsymbol{\varepsilon}^d$$

Such models continue to fit into our framework precisely because the derivative matrices $\mathcal{C}_\bullet$ depend only on the model's deterministic steady state, and not on policy rules that influence the economy's fluctuations *around* that steady state (e.g., a Taylor rule for interest rates). We will give a concrete numerical illustration of our identification result in the context of a quantitative HANK-type model in Section 2.4. Finally, as we discuss further in Appendix A.1, several interesting behavioral models (such as those of Gabaix (2020) or Carroll et al. (2018)) are also consistent with our assumptions.

While thus clearly quite general, our framework also has important limitations. First, since we leverage certainty equivalence of the linearized model economy, our identification

[9] For reference, we in Appendix A.1 explicitly write down the model (1) - (3) in the form of our general matrix system (6) - (7).

results will generally not yield *globally* valid policy counterfactuals. Second, the policy invariance assumption embedded in the equilibrium system (6) - (7) is not plausible for all kinds of policy rules: it generally holds for rules that only respond to aggregate perturbations of the macro-economy (such as Taylor rules), but will be violated by policies that change the model's steady state. For example, in the aggregate consumption function sketched above, changes in the long-run tax-and-transfer system will invariably affect the stationary distribution of households and thus the coefficient matrices $\mathcal{C}_\bullet$, so such policies are necessarily outside the purview of our analysis. Third, some models—even after linearization—do not feature a separation of policy and non-policy blocks as in (6) - (7). An important example are models featuring an asymmetry of information between the policymaker and the private sector (like Lucas, 1972). Here, private-sector agents solve a filtering problem, and in general the coefficients of the policy rule matter for this filtering problem both through the induced movements of the policy instrument *and* through the information contained in those movements. In particular, as we show in Appendix A.2, the standard Lucas island model induces an aggregate supply relation of the form

$$y_t = \theta \left[p_t - \mathbb{E}_{t-1}(p_t)\right]$$

where y_t denotes output and p_t is the price level. The coefficient θ depends on the policy rule for nominal demand growth simply because the rule affects the private sector's interpretation of changes in the island-level price, thus breaking the separation between the two blocks.

Objects of interest. As in our simple model, we wish to learn about systematic policy rule counterfactuals. Specifically, we consider an alternative policy rule

$$\tilde{\mathcal{A}}_x \boldsymbol{x} + \tilde{\mathcal{A}}_z \boldsymbol{z} = \mathbf{0} \tag{8}$$

Just like the baseline rule, this alternative policy rule is also assumed to induce a unique equilibrium.

Assumption 2. *The policy rule in (8) induces a unique equilibrium. That is, the infinite-dimensional linear map*

$$\tilde{\mathcal{B}} \equiv \begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \tilde{\mathcal{A}}_x & \tilde{\mathcal{A}}_z \end{pmatrix}$$

is invertible.

Given this alternative rule $\tilde{\mathcal{A}}$, we ask: what are the dynamic response paths $\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ to some given exogenous non-policy shock path $\boldsymbol{\varepsilon}$?

As a special case of the general counterfactual rule (8), we can study *optimal* policy rules corresponding to a given loss function. Specifically, we consider a policymaker with a quadratic loss function of the form

$$\mathcal{L} = \sum_{i=1}^{n_x} \lambda_i \boldsymbol{x}_i' W \boldsymbol{x}_i \tag{9}$$

where i indexes the n_x distinct (observable) macroeconomic aggregates collected in x, λ_i denotes policy weights, and $W = \mathrm{diag}(1, \beta, \beta^2, \cdots)$ allows for discounting.[10] As for our general counterfactual rule, we assume that the optimal policy problem has a unique solution.

Assumption 3. *Given any vector of exogenous shocks $\boldsymbol{\varepsilon}$, the problem of choosing the policy variable $\boldsymbol{z}$ to minimize the loss function* (9) *subject to the non-policy constraint* (6) *has a unique solution.*

We are then interested in two questions. First, what policy rule is optimal for such a policymaker? Second, given that optimal rule $(\mathcal{A}_x^*, \mathcal{A}_z^*)$, what are the corresponding dynamic response paths $\boldsymbol{x}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$ for a non-policy shock path $\boldsymbol{\varepsilon}$?

Finally, for both general as well as optimal counterfactual policy rules, we would like to go beyond counterfactuals for particular non-policy shock paths $\boldsymbol{\varepsilon}$, and instead also predict the effects of a rule change on *unconditional* macroeconomic dynamics. In particular, we would like to predict how the change in policy rule would affect the unconditional second-moment properties of the observed macroeconomic aggregates x.

The objective of the remainder of this section is to characterize the information required to recover these desired policy counterfactuals. The key insight is that, exactly as in our simple model, all of the required information can in principle be recovered from data generated under the baseline policy rule.

[10] We emphasize that our results are completely silent on the *shape* of the loss function, with structural modeling still the most natural way of obtaining a mapping from observables to welfare. We instead take as given the loss function and ask what kind of empirical evidence would be most useful to figure out how to minimize the loss. We furthermore note that our focus on a separable quadratic loss functions is in line with standard optimal policy analysis, but not essential. As shown in Appendix A.3, our results extend to the non-separable quadratic case, where the loss is now given by $\boldsymbol{x}'Q\boldsymbol{x}$ for a weighting matrix Q. While our approach in principle also applies to even richer loss functions, the resulting optimal policy rule will generally not fit into the form (8).

2.3 Identification results

We begin by defining the dynamic causal effects that lie at the heart of our identification results. By Assumption 1, we can write the solution to the system (6) - (7) as

$$\begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = \underbrace{-\mathcal{B}^{-1} \times \begin{pmatrix} \mathcal{H}_\varepsilon & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}}_{\equiv \Theta_{\mathcal{A}}} \times \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\nu} \end{pmatrix}$$

The linear map $\Theta_{\mathcal{A}}$ collects the impulse responses of $\boldsymbol{w}$, $\boldsymbol{x}$ and $\boldsymbol{z}$ to the non-policy and policy shocks $(\boldsymbol{\varepsilon}, \boldsymbol{\nu})$ under the prevailing baseline policy rule (7) with parameters $\mathcal{A}$. We will partition it as

$$\Theta_{\mathcal{A}} \equiv \begin{pmatrix} \Theta_{w,\varepsilon,\mathcal{A}} & \Theta_{w,\nu,\mathcal{A}} \\ \Theta_{x,\varepsilon,\mathcal{A}} & \Theta_{x,\nu,\mathcal{A}} \\ \Theta_{z,\varepsilon,\mathcal{A}} & \Theta_{z,\nu,\mathcal{A}} \end{pmatrix}. \tag{10}$$

All of our identification results will require knowledge of $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$—the impulse responses of the policy instruments z and macroeconomic observables x to contemporaneous as well as all possible future shocks $\boldsymbol{\nu}$ to the prevailing policy rule. Furthermore, to construct counterfactual paths that correspond to a given non-policy shock sequence $\boldsymbol{\varepsilon}$, we also require knowledge of the dynamic causal effects of that particular shock sequence under the baseline policy rule, $\{\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$. We emphasize that, in principle, all of these objects are estimable using data generated under the baseline policy rule: for example, given valid instrumental variables for all the distinct policy shocks $\boldsymbol{\nu}$ as well as a single instrument for the non-policy shock path $\boldsymbol{\varepsilon}$, the required entries of the Θ's can be estimated using semi-structural time-series methods (e.g. see Ramey, 2016, for a review).

These informational requirements are the natural generalization of those for the simple model in Section 2.1. First, since we are now considering an infinite-horizon economy, any given shock generates entire *paths* of impulse responses, corresponding to the rows of the Θ's. Second, rather than two policy shocks, we now need to know causal effects corresponding to the full menu of possible contemporaneous and news shocks $\boldsymbol{\nu}$—all columns of the Θ_ν's.

General counterfactual rule. We begin with the main object of interest—policy counterfactuals after a non-policy shock sequence $\boldsymbol{\varepsilon}$ under an alternative policy rule.

Proposition 1. *For any alternative policy rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$ that induces a unique equilibrium,* *we can recover the policy counterfactuals* $\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ *and* $\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ *as*

$$\begin{aligned}\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) &= \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) \equiv \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{x,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}} &(11)\\ \boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) &= \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) \equiv \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{z,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}} &(12)\end{aligned}$$

where $\tilde{\boldsymbol{\nu}}$ *is the unique solution of the system*

$$\tilde{\mathcal{A}}_x \left[\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{x,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}}\right] + \tilde{\mathcal{A}}_z \left[\boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{z,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}}\right] = \mathbf{0}. \tag{13}$$

Proof. The equilibrium system under the new policy rule can be written as

$$\begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \tilde{\mathcal{A}}_x & \tilde{\mathcal{A}}_z \end{pmatrix} \begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = \begin{pmatrix} -\mathcal{H}_\varepsilon \\ \mathbf{0} \end{pmatrix} \boldsymbol{\varepsilon} \tag{14}$$

By Assumption 2 we know that (14) has a unique solution $\{\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})\}$. To characterize this solution as a function of observables, suppose instead that we could find a $\tilde{\boldsymbol{\nu}}$ that solves (13). Since (6) also holds under the baseline policy rule, and since (13) imposes the new policy rule, it follows that any $(\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}))$ with $\tilde{\boldsymbol{\nu}}$ solving (13) are also part of a solution of (14). Since by assumption (14) has a unique solution, it follows that the system (13) is solved by at most one $\tilde{\boldsymbol{\nu}}$.

It remains to establish that the system (13) has a solution. For this consider the candidate $\tilde{\boldsymbol{\nu}} = (\tilde{\mathcal{A}}_x - \mathcal{A}_x)\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) + (\tilde{\mathcal{A}}_z - \mathcal{A}_z)\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$. Since the paths $\{\boldsymbol{w}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}), \boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})\}$ solve (14), it follows that they are also a solution to the system

$$\begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \mathcal{A}_x & \mathcal{A}_z \end{pmatrix} \begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = -\begin{pmatrix} \mathcal{H}_\varepsilon \boldsymbol{\varepsilon} \\ (\tilde{\mathcal{A}}_x - \mathcal{A}_x)\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) + (\tilde{\mathcal{A}}_z - \mathcal{A}_z)\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) \end{pmatrix} \tag{15}$$

But by Assumption 1 we know that the system (15) has a unique solution, so indeed the paths $\{\boldsymbol{w}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}), \boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})\}$ are that solution. It then follows from the definition of $\Theta_{\mathcal{A}}$ in (10) that the candidate $\tilde{\boldsymbol{\nu}}$ also solves (13), completing the argument. □

It follows from Proposition 1 that we can recover the desired counterfactual as a function of $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ and $\{\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$ alone. The key building block equation (13) is the infinite-horizon analogue of the bivariate system (5) from our simple two-period example in Section 2.1. The intuition is exactly the same: since we know the effects of all possible perturbations $\boldsymbol{\nu}$ of the baseline rule, we can always construct a date-0 shock vector $\tilde{\boldsymbol{\nu}}$ that mimics the equilibrium instrument path under the new instrument rule. But since the first model block (6) depends on the policy rule *only* via the expected instrument path, the equilibrium allocations under the new counterfactual rule and the perturbed prevailing rule are the same. The only difference relative to the simple model is that, because we now consider an infinite-horizon setting, we in general require evidence on contemporaneous and all possible future news shocks to the baseline rule in order to be able to mimic an arbitrary alternative policy rule.

Optimal policy. A very similar argument allows us to recover optimal policy rules corresponding to a given loss function.

Proposition 2. *Consider a policymaker with loss function (9). For any $\boldsymbol{\varepsilon}$, the solution to the optimal policy problem is uniquely implemented by the rule $\{\mathcal{A}_x^*, \mathcal{A}_z^*\}$ with*

$$\begin{aligned} \mathcal{A}_x^* &= \left(\lambda_1 \Theta'_{x_1,\nu,\mathcal{A}} W, \lambda_2 \Theta'_{x_2,\nu,\mathcal{A}} W, \ldots, \lambda_{n_x} \Theta'_{x_{n_x},\nu,\mathcal{A}} W\right), \qquad (16) \\ \mathcal{A}_z^* &= \mathbf{0}. \qquad (17) \end{aligned}$$

Given $\{\mathcal{A}_x^*, \mathcal{A}_z^*\}$, the corresponding counterfactual paths under the optimal policy rule, $\boldsymbol{x}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$, are characterized as in Proposition 1.

Proof. The solution to the optimal policy problem is characterized by the following first-order conditions:

$$\begin{aligned} \mathcal{H}_w'(I \otimes W)\boldsymbol{\varphi} &= \mathbf{0} \qquad (18) \\ (\Lambda \otimes W)\boldsymbol{x} + \mathcal{H}_x'(I \otimes W)\boldsymbol{\varphi} &= \mathbf{0} \qquad (19) \\ \mathcal{H}_z' W \boldsymbol{\varphi} &= \mathbf{0} \qquad (20) \end{aligned}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots)$ and φ is the multiplier on (6). By Assumption 3 we know that the system (18) - (20) together with (6) has a unique solution $\{\boldsymbol{x}^*(\boldsymbol{\varepsilon}), \boldsymbol{z}^*(\boldsymbol{\varepsilon}), \boldsymbol{\varphi}^*(\boldsymbol{\varepsilon})\}$.

Now consider the alternative problem of choosing deviations $\boldsymbol{\nu}^*$ from the prevailing rule to minimize (9) subject to (6) - (7). This second problem gives the first-order conditions

$$\begin{aligned} \mathcal{H}_w'(I \otimes W)\boldsymbol{\varphi} &= \mathbf{0} \qquad (21) \\ (\Lambda \otimes W)\boldsymbol{x} + \mathcal{H}_x'(I \otimes W)\boldsymbol{\varphi} + \mathcal{A}_x' W \boldsymbol{\varphi}_z &= \mathbf{0} \qquad (22) \\ \mathcal{H}_z'(I \otimes W)\boldsymbol{\varphi} + \mathcal{A}_z' W \boldsymbol{\varphi}_z &= \mathbf{0} \qquad (23) \\ W \boldsymbol{\varphi}_z &= \mathbf{0} \qquad (24) \end{aligned}$$

where φ_z is the multiplier on (7). It follows from (24) that $\boldsymbol{\varphi}_z = \mathbf{0}$. But then (21) - (23) together with (6) determine the same unique solution as before, and $\boldsymbol{\nu}^*$ adjusts residually to satisfy (7). The original problem and the alternative problem are thus equivalent.

Next note that, by Assumption 1, we can re-write the alternative problem's constraint set as

$$\begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = \Theta_{\mathcal{A}} \times \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\nu}^* \end{pmatrix} \tag{25}$$

The problem of minimizing (9) subject to (25) gives the optimality condition

$$\sum_{i=1}^{n_x} \lambda_i \Theta'_{x_i,\nu,\mathcal{A}} W \boldsymbol{x}_i = 0 \tag{26}$$

By the equivalence of the policy problems, it follows that (26) is an optimal policy *rule*, taking the form (16) - (17). Finally, the second part of the result follows from Proposition 1 since (26) is just a special example of a policy rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$.

□

Proposition 2 reveals that, in conjunction with a given policymaker loss function, the information required to construct valid counterfactuals for arbitrary policy rules also suffices to characterize *optimal* policy rules.[11] The intuition is exactly as before: since we know the causal effects of every possible policy perturbation $\boldsymbol{\nu}$ on the policymaker targets $\boldsymbol{x}$, we in particular know the space of those targets that is implementable through policy actions. At an optimum, we must be at the point of this space that minimizes the policymaker loss. As before, it does not matter whether this optimum is attained through some systematic policy rule or through shocks to an alternative rule.

[11]Note that, by mapping our perfect foresight economy to a linearized economy with aggregate risk, we can re-write that optimal policy rule as a forecasting targeting rule (Svensson, 1997):

$$\sum_{i=1}^{n_x} \lambda_i \Theta'_{x_i,\nu,\mathcal{A}} W \mathbb{E}_t \left[\boldsymbol{x}_i\right] = 0 \tag{27}$$

where now $\boldsymbol{x}_i = (x_{it}, x_{it+1}, \dots)'$. In words, expectations of future targets must always minimize the policymaker loss within the space of (expected) allocations that are implementable via changes in the policy stance. For a timeless perspective, (27) must apply to *revisions* of policymaker expectations at each t.

Unconditional second-moment properties. While Propositions 1 and 2 predict counterfactual dynamics for *particular* non-policy shock paths $\boldsymbol{\varepsilon}$, researchers may also be interested in *unconditional* second-moment properties following a change in policy rule. Of course, if researchers have estimated the effects of all distinct non-policy shocks hitting the economy, then such unconditional analysis is simple: apply Propositions 1 and 2 for each such shock and then collect the results in the form of a vector moving average representation.

In practice, however, researchers may not be able to isolate all distinct aggregate non-policy shocks. Our third identification result states that, in some cases, it is nevertheless possible to recover the desired counterfactual second-moment properties. Since the result requires some investment in additional notation, we only state the main idea here and relegate all details to Appendix A.4. The key assumption allowing us to make progress is that of "invertibility": we need to assume that the vector moving average representation of the observable data x and z under the baseline policy rule is invertible with respect to the structural shocks driving the economy. This assumption, while restrictive (Plagborg-Møller & Wolf, 2021a), is routinely imposed in conventional Structural Vector Autoregression analysis (Fernández-Villaverde et al., 2007). Under this assumption, researchers need not be able to separately observe all of the individual aggregate shocks; instead, it suffices to simply apply our counterfactual prediction results in Propositions 1 and 2 to Wold innovations and then again collect the results in the form of a counterfactual vector moving average. Appendix A.4 also discusses why this argument fails in the non-invertible case.

Discussion. The identification results in Propositions 1 and 2 offer a bridge between the "Lucas program" as presented in Christiano et al. (1999) and purely empirical approaches to policy counterfactual analysis (as in Sims & Zha, 2006). The propositions reveal that, under our assumptions, impulse responses to contemporaneous and news policy *shocks*—objects that are estimable using semi-structural empirical techniques—are sufficient statistics for predicting the effects of changes in systematic policy *rules*. Key to our argument is the use of *multiple distinct* policy shocks. By using many such shocks (all at date 0), counterfactual rules can be imposed not just ex post but also in ex ante expectation, which is enough to fully sidestep the Lucas critique. We further elaborate on the connection to the approach of Sims & Zha—which uses one policy shock, set to a new level at each date t—in Section 2.4.

Our results also resonate with recent attempts to bring insights from the "sufficient statistics" approach popular in public finance to macroeconomics (Chetty, 2009; Nakamura & Steinsson, 2018). For a large family of structural models and policy rule counterfactuals, policy shock impulse responses turn out to be precisely such sufficient statistics for systematic policy rule counterfactuals.

2.4 Illustration

We now provide a visual illustration of our theoretical identification results. As our laboratory we use the structural HANK model of Wolf (2021), with details of the model parameterization relegated to Appendix A.1. In this environment we compute policy counterfactuals in two ways: first by using the structural equations of the model and solving the model with a counterfactual policy rule, and second by using our identification results.

We begin by solving the model with a baseline policy rule of

$$i_t = \phi_\pi \pi_t \tag{28}$$

for $\phi_\pi = 1.5$. In particular, we recover a) the impulse responses $\{\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$ to a contractionary cost-push shock ε_0 and b) the causal effects of *all* policy shocks $\boldsymbol{\nu}$, $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$. We emphasize that those causal effects would be estimable for an econometrician living in our model laboratory and with access to valid instrumental variables for the cost-push shock ε as well as the policy shocks $\{\nu_{0,t}, \nu_{1,t}, \dots\}$.

We now entertain the following counterfactual policy rule:

$$i_t = \phi_i i_{t-1} + (1 - \phi_i)(\phi_\pi \pi_t + \phi_y y_t) \tag{29}$$

for $\phi_i = 0.9$, $\phi_\pi = 2$, $\phi_y = 0.5$. Figure 1 shows model-implied impulse responses to a cost-push shock ε_0 under the baseline rule (28) (grey) and the counterfactual rule (29) (orange), where both of these lines are computed from the structural equations of the model. Next, following Proposition 1, we use the estimable effects of policy shocks to the baseline policy rule to construct the counterfactual, with results shown as the navy blue line. As expected, the outcome is identical to the one from the true structural solution of the model (i.e., orange and blue lines coincide). Finally, the right panel shows the sequence of shocks $\tilde{\boldsymbol{\nu}}$ that maps the baseline rule into the counterfactual rule. Since the new rule is more accommodating, the sequence of shocks is persistently negative (i.e., the shocks are expansionary). Appendix A.5 illustrates Proposition 2 with a similar application to optimal policy counterfactuals.

Link to Sims & Zha (2006). Our identification result enforces the counterfactual rule using date-0 contemporaneous and news policy shocks $\boldsymbol{\nu}$. To shed further light on this logic

Alternative Policy Rule, HANK Model

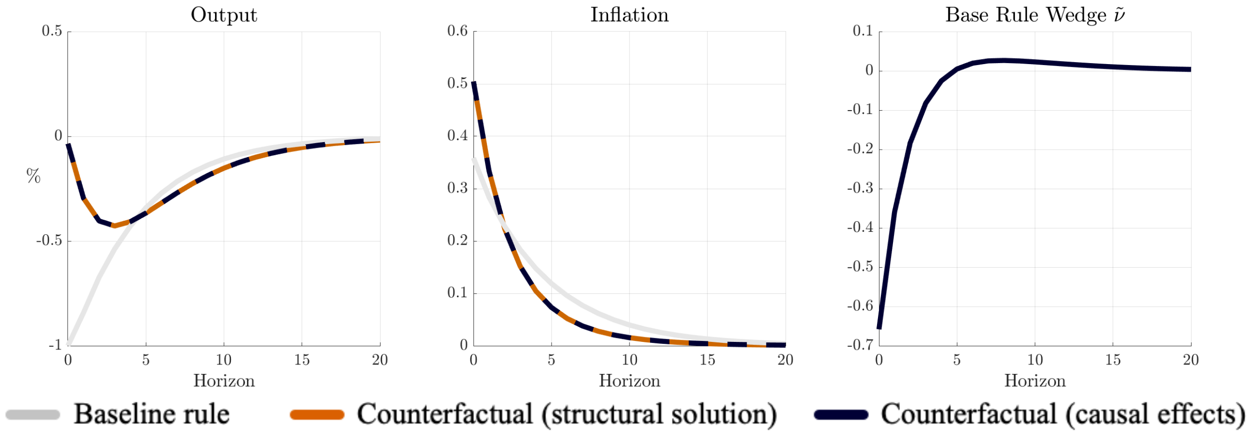

Figure 1: Output and inflation impulse responses for the HANK model with policy rules (28) and (29) together with the policy shock $\tilde{\boldsymbol{\nu}}$ to (28) that mimics (29) (see (13)).

and its relationship to Sims & Zha (2006), Figure 2 constructs policy rule counterfactuals using instead a mix of date-0 and date-$t > 0$ policy shocks. Formally, we consider a researcher with access to the causal effects of the first n_s entries of the policy shock vector $\boldsymbol{\nu}$. Using only the first entry (i.e., the contemporaneous shock $\nu_{0,t}$), she could implement the method of Sims & Zha: she could subject the economy to a new surprise shock $\nu_{0,t}$ at each t chosen so that, at each t, the policy instrument and macro aggregates are related as required by the counterfactual policy rule (29). For $n_s > 1$, we generalize this approach: our researcher uses her n_s policy shocks at each $t \geq 0$ to enforce the desired counterfactual rule not only ex post (as Sims & Zha do with one shock), but also in ex ante expectation for the next $n_s - 1$ periods. We present implementation details for this approach in Appendix A.6.

The main takeaway from the figure is that, as $n_s \to \infty$, the constructed counterfactual converges to the true counterfactual, consistent with our identification results. For $n_s = 1$, on the other hand, the constructed counterfactual (in light grey) is quite far from the truth. Intuitively, the issue is that the contemplated counterfactual policy rule is only imposed ex post but not in ex ante expectation. Since expectations about the future in general affect the present, enforcing the rule through ex post surprises is not the same as switching and committing to a different rule from time $t = 0$ onwards.[12] Moving to $n_s = 2$ shocks enforces

[12] It follows from this discussion that, if the private sector were not at all forward-looking, then one shock would already be enough for Lucas critique-robust counterfactuals, simply because there is no distinction between ex post shocks and news shocks.

Alternative Policy Rule (ex post surprises), HANK Model

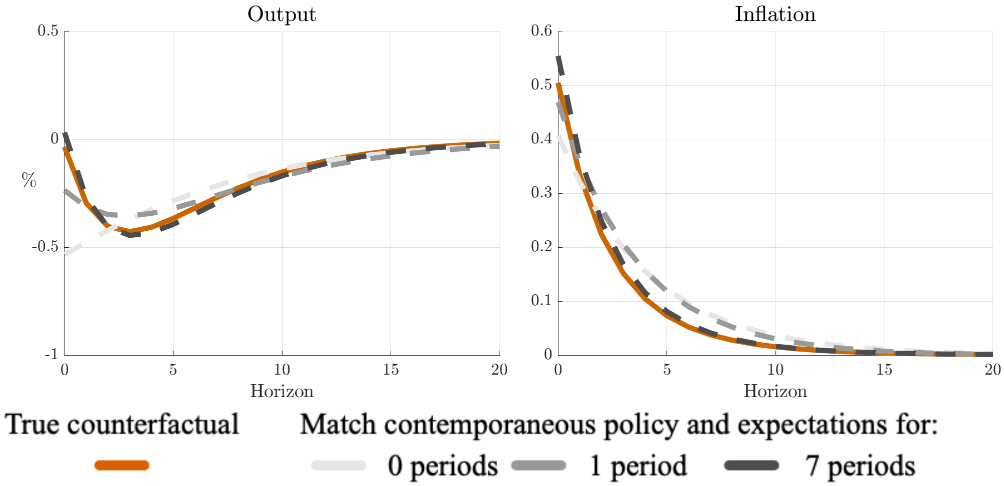

Figure 2: Counterfactual output and inflation impulse responses using ex post surprises. At each date, we solve for the first n_s policy shocks to enforce the counterfactual rule contemporaneously and in expectation for the next $n_s - 1$ periods. Results are for $n_s \in \{1, 2, 8\}$ and for the HANK model with policy rules (28) and (29).

the counterfactual rule also in one-period-ahead expectation, thus reducing the size of the required ex post surprises, reducing the error from incorrect expectations, and so giving a more accurate counterfactual. For $n_s = 8$, the counterfactual rule can be implemented almost perfectly using only date-0 shocks, so ex post surprises and impulse response are close to the theoretical $n_s \to \infty$ limit. This is unsurprising: if the underlying non-policy shock is rather transitory, then differences between the new and old rules will be transitory, and so the mapping between policy rules will mostly rely on knowledge of the causal effects of short-run policy shocks. Intuitively, under both the prevailing as well as the contemplated counterfactual rule, the private sector expects the economy to have returned to steady state anyway at medium horizons, so those medium-run expectations do not distort short-run dynamics. This observation will turn out to be key for our empirical applications in Section 3.

2.5 Discussion

We have demonstrated that, in a quite general family of linearized structural macroeconomic models, impulse responses to policy shocks can serve as "sufficient statistics" for the effects of changes in systematic policy rules. Put differently, our results imply that—under our maintained structural assumptions—the Lucas critique can in principle be circumvented purely through empirical measurement.[13]

In the remainder of this paper we discuss how to operationalize our insights. The main challenge is that our informational requirements are quite high: the population identification result requires evidence on the dynamic causal effects of the full menu of contemporaneous and news policy shocks *at all possible horizons*. Section 3 presents a measurement strategy for the empirically relevant case of researchers with access only to a couple of distinct identified policy shocks. Section 4 then discusses the role of structural modeling in cases where the existing empirical evidence is too limited to apply our empirical method.

3 Counterfactuals with finitely many shocks

This section presents our empirical method for constructing policy counterfactuals with evidence on multiple, but finitely many, distinct policy shocks. Section 3.1 sets the stage by connecting the objects in our identification result to objects that are estimated in practice, Section 3.2 introduces the methodology, and Section 3.3 presents several applications to monetary policy counterfactuals.

3.1 From empirical evidence to our "sufficient statistics"

Empirical researchers have relied on different pieces of identifying information to estimate the effects of policy shocks. For example, the monetary policy shock literature has identified quasi-random variation in policy using a large variety of methods (e.g. Romer & Romer, 2004; Gürkaynak et al., 2005; Gertler & Karadi, 2015; Antolin-Diaz et al., 2021; Inoue & Rossi, 2021), with each exogenous piece likely to load on different shocks $\{\nu_{\ell,t-\ell}\}_{\ell=0}^{\infty}$, thus resulting in different paths of nominal rates.[14] Anticipating our empirical application, Figure 3 provides an illustration showing interest rate paths for two shocks: the left panel corresponding to a transitory rate hike, and the right panel showing a more gradual change.

What is the connection between such empirical evidence and the informational requirements of our "sufficient statistics" identification results? The theoretical discussion of con-

[13] In fact, as we discuss in Appendix A.7, our identification results can in principle even be extended to non-linear models with aggregate risk; as we discuss there, the main change is that our informational requirements increase even further, with the required causal effects of policy shocks now additionally indexed by the state of the economy as well as the magnitude of the policy intervention.

[14] Similarly, the fiscal policy literature has studied both transitory as well as persistent changes in aggregate government purchases (e.g. Mertens & Ravn, 2010; Ramey, 2011; Leeper et al., 2013). Our focus on monetary policy is in keeping with much of the prior literature on policy rule counterfactuals.

Identified Policy Shock Paths, Illustration

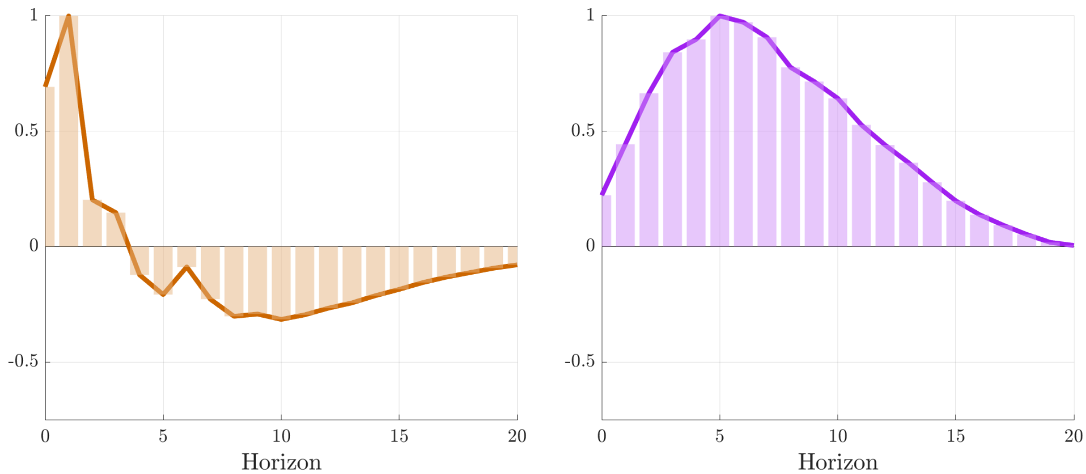

Figure 3: Two possible instrument paths $\boldsymbol{z}(\boldsymbol{\nu}_s)$ corresponding to two different shock paths $\boldsymbol{\nu}_s$, $s = 1, 2$: a short-lived change (orange, left panel) and a gradual, persistent departure from the rule (purple, right panel).

temporaneous and news policy shocks in Section 2 was phrased in terms of policy shocks $\boldsymbol{\nu}$ that perturb the prevailing policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$ horizon by horizon. The proofs of our identification results, however, reveal that what ultimately matters is not the particular shock path $\boldsymbol{\nu}$; rather, the key is that the researcher can predict the counterfactual effects of the policy instrument path associated with some given counterfactual rule. Viewed in this light, we may re-state our identification results as requiring estimates of the effects of all possible policy instrument paths $\boldsymbol{z}$ (rather than all shocks $\{\nu_{0,t}, \nu_{1,t}, \dots\}$).[15] Existing studies give us the dynamic causal effects associated with *particular* paths of the policy instrument, as in Figure 3. Thus, the more shocks are estimated, the larger the space of policy instrument paths whose counterfactual effects we can predict. Our theoretical identification result corresponds to the limit where the estimated shocks span the space of all possible changes in the current and expected future path of the policy instrument. The methodology presented in Section 3.2 discusses how researchers can use the available evidence for *particular* policy instrument paths (like those in Figure 3) to provide a best Lucas critique-robust approximation to the desired policy rule counterfactual.

[15]Formally, what we are discussing here is nothing but a change of basis: we solve for the policy counterfactual not in terms of shocks to some (arbitrary) baseline rule $\{\mathcal{A}_x, \mathcal{A}_z\}$, but directly in terms of policy instrument paths. This switch of basis is without loss of generality as long as the policymaker can implement any possible path of the policy instrument (i.e., the map $\Theta_{z,\nu,\mathcal{A}}$ is invertible).

3.2 Empirical methodology

We consider a researcher that has access to the dynamic causal effects associated with n_s distinct paths of the policy instrument z. We denote those effects by $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$, where each of the n_s columns of the Ω's gives the impulse response to a distinct identified policy shock associated with a distinct path for the policy instrument. Given such lower-dimensional causal effect maps, and given a non-policy shock $\boldsymbol{\varepsilon}$ and a counterfactual rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$, the proof strategy of Proposition 1 will now in general fail. We would need to set

$$\tilde{\mathcal{A}}_x(\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{x,\mathcal{A}} \times \boldsymbol{s}) + \tilde{\mathcal{A}}_z(\boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{z,\mathcal{A}} \times \boldsymbol{s}) = \mathbf{0} \tag{30}$$

where $\boldsymbol{s} \in \mathbb{R}^{n_s}$ denotes weights assigned to the n_s empirically identified policy shocks at date 0. The problem is that this system of T equations (where T is the large maximal transition horizon) in n_s unknowns will generically not have a solution. So how can researchers proceed?

Lucas critique-robust method. Our main proposal is to simply select the weights $\boldsymbol{s}$ on the n_s date-0 shocks to enforce the desired counterfactual rule *as well as possible*. In practice, this means solving a straightforward regression problem:

$$\min_{\boldsymbol{s}} \quad ||\tilde{\mathcal{A}}_x(\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{x,\mathcal{A}} \times \boldsymbol{s}) + \tilde{\mathcal{A}}_z(\boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{z,\mathcal{A}} \times \boldsymbol{s})||. \tag{31}$$

The output of the simple problem (31) is the best approximation to the desired policy counterfactual within the space of empirically identified policy shock paths. By our identification results in Section 2 and because all shocks are dated $t = 0$ (i.e., no ex post surprises), this approach is fully Lucas critique-robust. The richer the evidence on policy shock propagation (i.e., as $n_s \to \infty$), the better this approximate counterfactual will enforce the counterfactual rule, eventually converging to the truth. The important limitation of our approach is that, for small n_s, it will not always be possible to construct an accurate approximation of the desired counterfactual rule: for some contemplated counterfactual rules, the target (30) can be made to hold almost exactly, while for others the implementation error will be large. The usefulness of our proposed method is thus an inherently application-dependent question.

Alternative: a multi-shock refinement of Sims & Zha (2006). In keeping with this paper's overarching focus on robustness to Lucas critique concerns, we will mostly consider results from our baseline method. However, we note that our identification results also suggest a refinement of Sims & Zha (2006)—a refinement that relies on stronger assumptions than our baseline method, but weaker assumptions than the original one-shock approach proposed by Sims & Zha. Given the popularity of the Sims & Zha (2006) approach we briefly discuss this refinement here.

The idea of the refinement is that, by going from one to multiple identified policy shocks, a researcher can reduce her reliance on ex post (i.e., date $t \geq 1$) policy surprises to enforce the counterfactual policy rule. As we discussed in Section 2.4, the original Sims & Zha method is subject to Lucas critique concerns precisely because of ex post surprises: the counterfactual rule holds at each t, but is not expected to hold from $t+1$ onwards. Our proposed extension of the Sims & Zha method trades off rule accuracy versus ex post surprises in the form of a simple ridge regression, generalizing our baseline method (31). To formally state this approach we require some additional notation. We let $\{\Omega_{x,\mathcal{A}}^{(h)}, \Omega_{z,\mathcal{A}}^{(h)}\}$ denote impulse responses to policy shocks that materialize at horizon h; that is, for $h = 0$ those impulse responses are simply given as $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$, while for $h > 0$ impulse responses at the first $h-1$ horizons are exactly zero, and impulse responses from horizon h onwards are equal to $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$. Now let $\boldsymbol{s}^h \in \mathbb{R}^{n_s}$ denote the weights assigned to the n_s shocks at horizon h. Our refinement of Sims & Zha then solves the following ridge regression problem:

$$\min_{\{\boldsymbol{s}^h\}_{h=0}^{H}} ||\tilde{\mathcal{A}}_x(\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \sum_{h=0}^{H} \Omega_{x,\mathcal{A}}^{(h)} \times \boldsymbol{s}^h) + \tilde{\mathcal{A}}_z(\boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \sum_{h=0}^{H} \Omega_{z,\mathcal{A}}^{(h)} \times \boldsymbol{s}^h)|| + \psi \sum_{h=1}^{H} ||\boldsymbol{s}^h||, \tag{32}$$

where the tuning parameter ψ penalizes ex post policy surprises, and $H \gg 0$ is the maximal shock horizon. For $\psi = \infty$ this method simply reduces to our baseline method, with only the date-0 shocks $\boldsymbol{s}^0$ allowed to be different from zero. For $\psi = 0$ (and large H) the counterfactual rule is instead imposed perfectly ex post as in the original proposal of Sims & Zha, with $n_s = 1$ corresponding exactly to their procedure. For intermediate ψ, the researcher is willing to trade off ex post surprises $\boldsymbol{s}^h$ for $h \geq 1$ in return for higher accuracy in implementing the desired counterfactual policy rule.[16] If those ex post surprises are small enough, then researchers may be willing to accept the expectational errors they entail in return for more accurately imposing the counterfactual rule ex post.

The key appeal of this hybrid method relative to the original proposal of Sims & Zha will be that, precisely because our method uses multiple distinct policy shocks, a (much) better

[16] Rather than smoothly penalizing ex post surprises as in (32), researchers may instead consider using n_s shocks to enforce a given counterfactual rule ex post and in expectation for the next $n_s - 1$ periods, as we did in Figure 2. Unfortunately we have found this method often yields explosive dynamics in actual data—a problem that actually also arises with the original approach of Sims & Zha (2006). We provide further details in Appendix B.5.

fit can be achieved using date-0 shocks alone, with less need to rely on ex post surprises. We will illustrate this observation in our applications.

Optimal policy rules. By Proposition 2, our results also allow researchers to learn about *optimal* counterfactual policy rules, given some exogenously specified loss function. Appendix B.1 shows how to apply both our baseline Lucas critique-robust method as well as the multi-shock refinement of Sims & Zha (2006) to such questions of optimal policy design. Very briefly, the idea is to use date-0 policy shocks to reduce the policymaker loss as much as possible. Our approach thus minimizes the loss function by perturbing the baseline policy response in directions spanned by the empirically identified policy shocks.[17]

3.3 Applications

In this section we apply our empirical strategy to predict the effects of investment-specific technology shocks under various counterfactual monetary policy rules. In particular, our objects of interest are the counterfactual behavior of the output gap, inflation, and the short-term nominal rate. We choose to focus on investment-specific technology shocks since such shocks are widely argued to be one of the main drivers of aggregate business-cycle fluctuations, at least in the U.S. (e.g. see Justiniano et al., 2010; Ramey, 2016).

We proceed as follows: we estimate the inputs required by our methodology, apply the method and present the main results, and then discuss how to interpret those results in light of our theoretical identification results in Section 2. Appendix B provides the details of the empirical implementation.

Inputs. The first input to our analysis are the aggregate effects of the non-policy shock of interest $\boldsymbol{\varepsilon}$ under the prevailing baseline policy rule. To recover those effects we rely on the investment-specific technology news shock series identified by Ben Zeev & Khan (2015)—a shock that induces an anticipated change in the relative price of investment goods. We estimate the propagation of this shock by ordering it first in a recursive Vector Autoregression (VAR) (as recommended in Plagborg-Møller & Wolf, 2021b).

The second input are the causal effects of a menu of different monetary policy shocks. For

[17] This part of our method is related to work by Barnichon & Mesters (2021). Those authors argue that, under quite general conditions, evidence on policy shock impulse responses can be used to *test* the optimality of policy conduct. Our method makes stronger assumptions—notably the separation of the policy and non-policy blocks in (6) - (7)—allowing us to explicitly characterize optimal policy (rules), as in Proposition 2.

this we consider two of the most popular examples of such monetary shocks: the shock series of Romer & Romer (2004) and Gertler & Karadi (2015). Our estimates of the responses of interest rates to those two shocks differ quite substantially: rather short-lived for Romer & Romer, and more gradual for Gertler & Karadi. Indeed, in our illustrative figure from before (Figure 3), the left panel corresponds to the short-run nominal interest rate path identified by the Romer & Romer shock, while the right panel presents the Gertler & Karadi interest rate path. To interpret these estimates, it is instructive to return to the simple monetary policy rule from our illustrative example:

$$i_t = \phi \pi_t + \underbrace{\nu_{0,t}}_{\text{contemp. shock}} + \underbrace{\sum_{\ell=1}^{\infty} \nu_{\ell,t-\ell}}_{\text{news shocks}}$$

We interpret the differences in our estimated rate response paths as indicative of the Romer & Romer and Gertler & Karadi policy IVs loading differentially on contemporaneous versus news policy shocks. While the Romer & Romer shock is short-lived, the Gertler & Karadi shock is well-known to move longer-term rates and is thus more likely to have a larger forward guidance component, consistent with Figure 3.[18] To correctly account for joint uncertainty in the estimation of the two shocks, we study their propagation through a single VAR.

Counterfactual policy results. We use our methodology to construct policy rule counterfactuals for several different alternative rules: output gap targeting; a standard Taylor (1993) rule; a nominal interest rate peg; nominal GDP targeting; and the optimal policy rule corresponding to a loss function with equal weight on the output gap and a weighted average of current and lagged inflation (i.e., average inflation targeting). Our discussion will mostly focus on our preferred method that does not allow any ex post surprises, though we also consider results from the Sims & Zha refinement (for an equal penalty on rule inaccuracy and ex post policy shock surprises, i.e. $\psi = 1$). Finally, the counterfactual implied by the original method of Sims & Zha—using only one of our two shocks—is discussed in Appendix B.5. Throughout, our measure of rule accuracy is the horizon-by-horizon error in enforcing the desired counterfactual rule (i.e., the argument of (31) or (32)).

First, Figure 4 shows our counterfactual results for output gap stabilization. The identified investment technology shock has both a cost-push as well as a negative demand com-

[18] Our finding of persistent changes in rates following the Gertler & Karadi shock is consistent with the estimates reported in their paper (their Tables 1 & 3). Alternative approaches that identify (partial) forward guidance shocks and that we could have used include Antolin-Diaz et al. (2021) or Inoue & Rossi (2021).

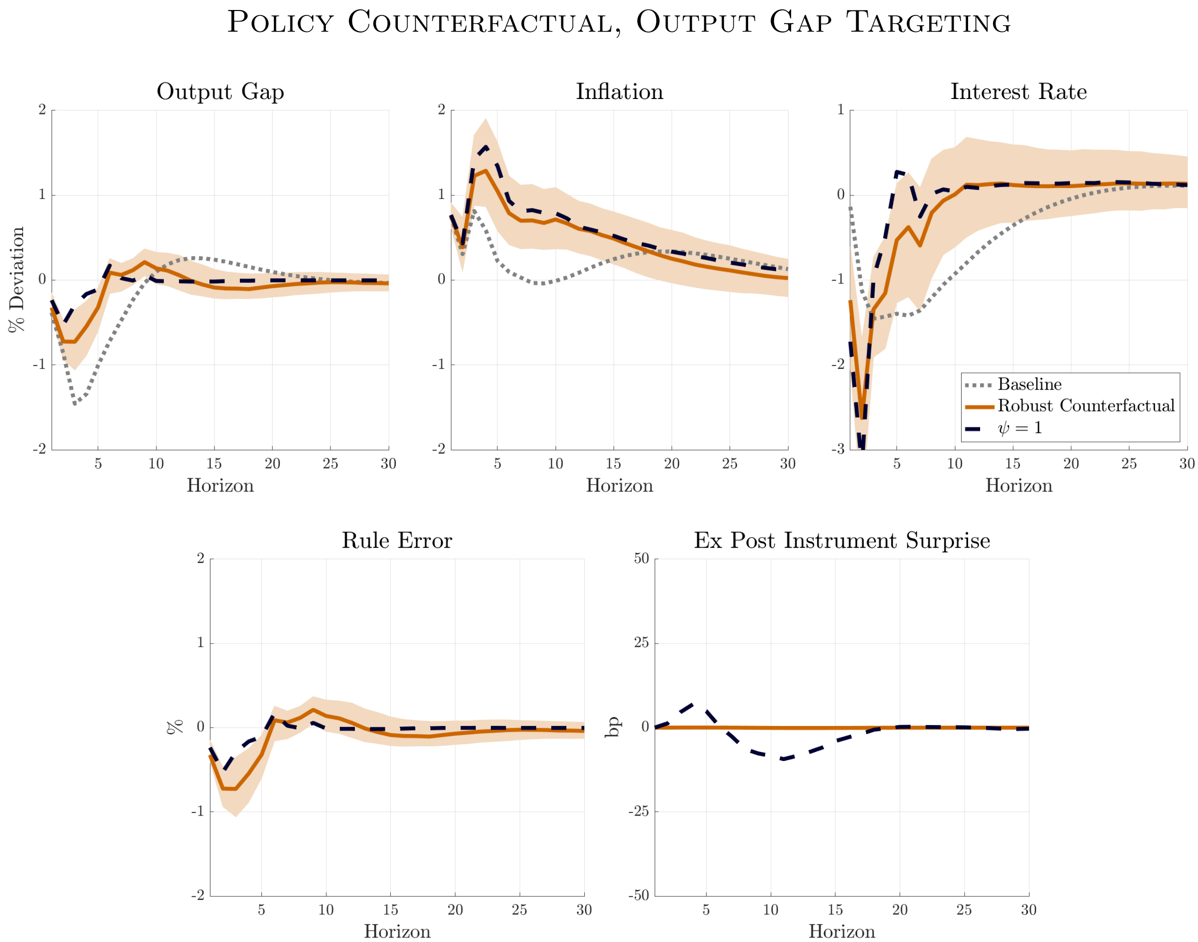

Figure 4: Top panel: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the best feasible approximation to output gap targeting (orange and black dashed), computed following (31) and (32) for $\psi = 1$. Bottom panel: implementation error for the counterfactual rule and ex post nominal interest rate surprise at time t. The shaded areas correspond to 16th and 84th percentile confidence bands.

ponent, consistent with theory (e.g. Justiniano et al., 2010). Under the baseline policy rule (dotted grey), interest rates are cut relatively aggressively, though by not enough to stabilize the output gap; furthermore inflation stays moderately above target. Under our approximation to output gap targeting, rates are cut much more aggressively, essentially stabilizing the output gap from around a couple of quarters after the shock, at the cost of quite persistently higher inflation. The bottom panels and the dashed lines reveal that allowing for some ex post shocks essentially does not change the picture: later shocks do not help with output gap stabilization right at the beginning, but after a couple of quarters the output gap is almost perfectly stabilized anyway using date-0 shocks.

Policy Counterfactual, Taylor Rule

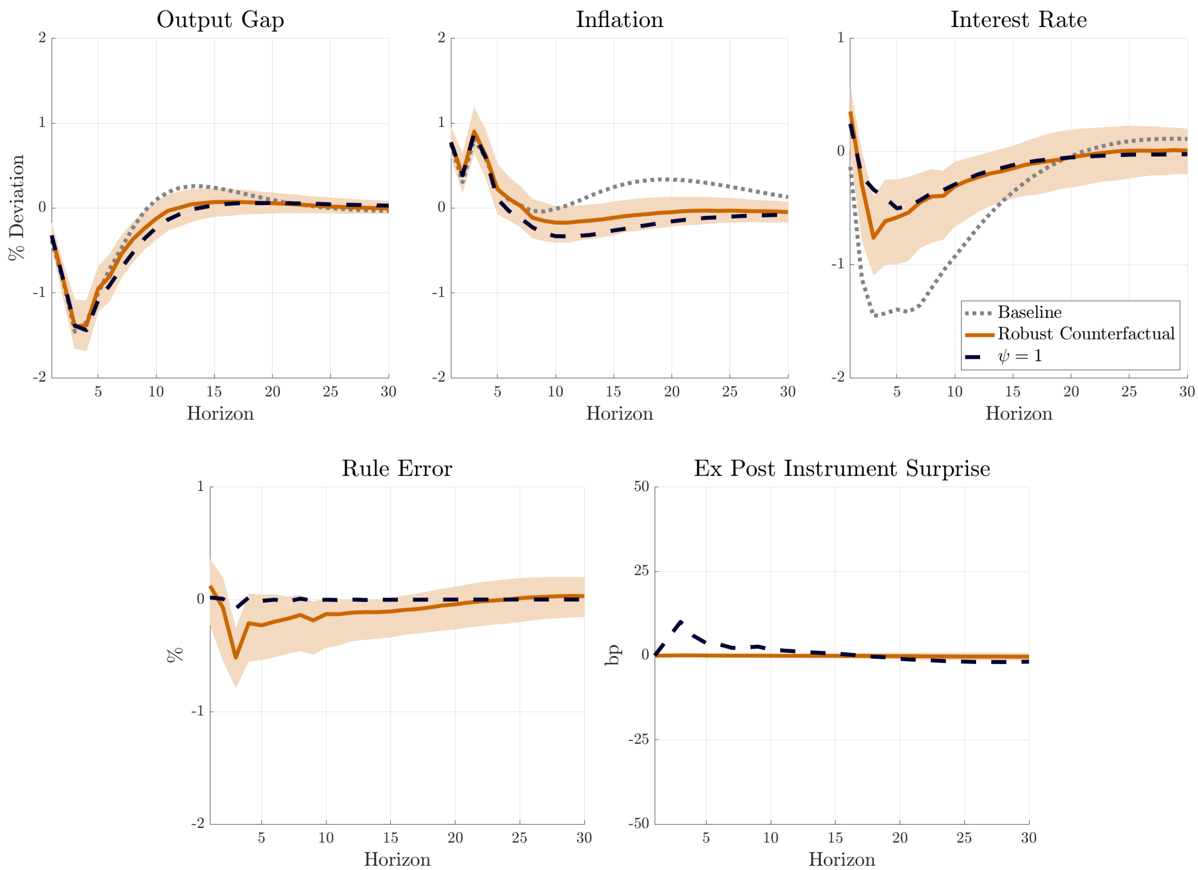

Figure 5: Top panel: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the best feasible approximation to a Taylor rule $\widehat{i}_t = 0.5\widehat{\pi}_t + 0.5\widehat{y}_t$ (orange and black dashed), computed following (31) and (32) for $\psi = 1$. Bottom panel: implementation error for the counterfactual rule and ex post nominal interest rate surprise at time t. The shaded areas correspond to 16th and 84th percentile confidence bands.

Second, Figure 5 shows the results for the rule proposed in Taylor (1993), with response coefficients of 0.5 on aggregate inflation and the output gap. Due to the observed increase in inflation, this rule actually dictates a much less aggressive interest rate cut, resulting in somewhat lower output and inflation at medium horizons. The bottom left panel reveals that the counterfactual rule is imposed relatively well throughout, except at a couple of quarters after the initial shock (where rates are still cut by too much relative to the rule prescription). The black dashed lines furthermore reveal that a moderate ex post interest rate surprise at this point is sufficient to impose the desired rule almost perfectly, with relatively little effect on the implied output gap, inflation, and policy instrument dynamics.

Policy Counterfactual, Optimal AIT Policy Rule

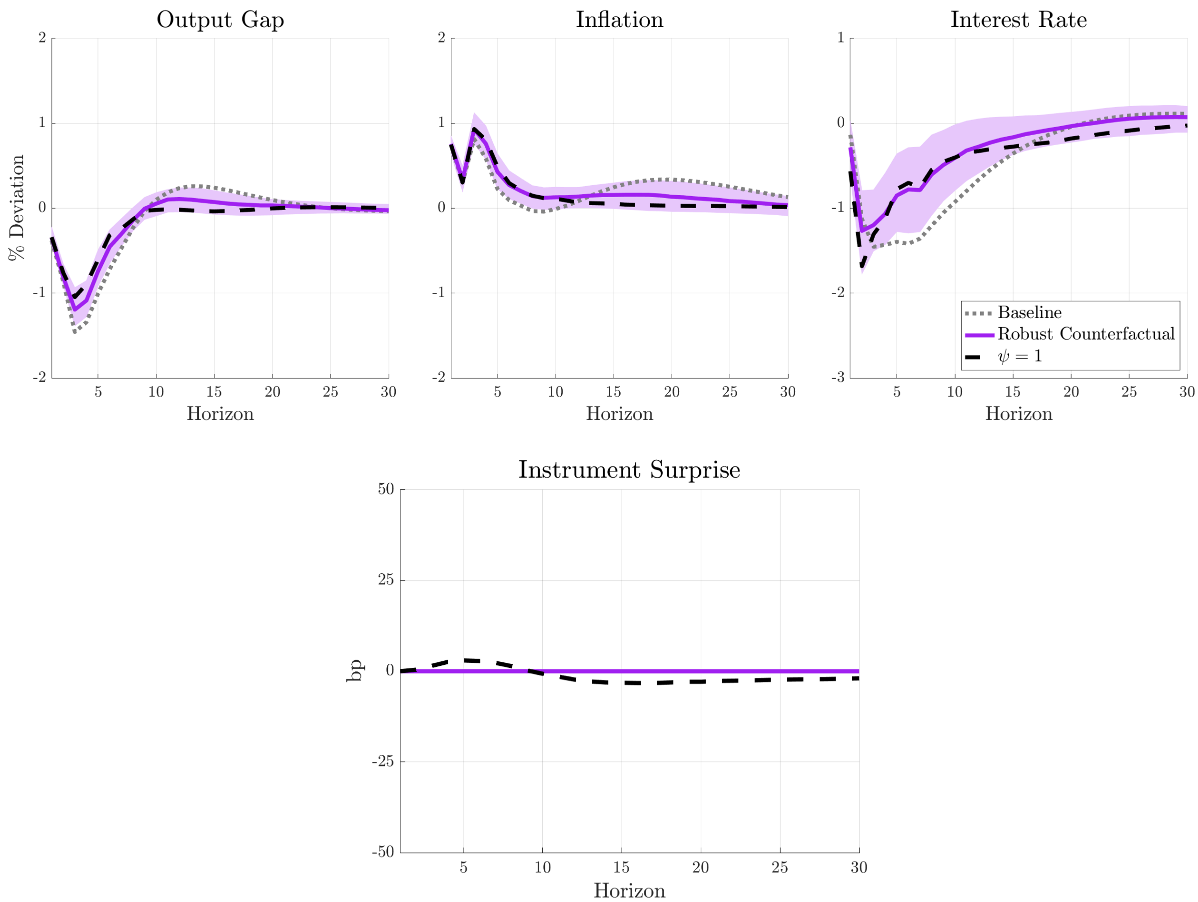

Figure 6: Top panel: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the best feasible approximation to an optimal average inflation targeting monetary policy rule (purple and black dashed), computed as discussed in Appendix B.1 for $\psi = \infty$ and $\psi = 1$. Bottom panel: ex post nominal interest rate surprise at time t. The shaded areas correspond to 16th and 84th percentile confidence bands.

Third, we proceed in the spirit of the recent change in the Federal Reserve's strategy and consider a policymaker with preferences over output and *average* inflation $\bar{\pi}_t$, where[19]

$$\bar{\pi}_t = \sum_{\ell=0}^{K} \omega_\ell \pi_{t-\ell}.$$

[19]Here K denotes the maximal (lagged) horizon that enters the inflation averaging, and ω_ℓ denotes the weight on the ℓth lag, with $\sum_\ell \omega_\ell = 1$ and $\omega_\ell \geq 0 \; \forall \ell$. For our application we set $K = 20$ and $\omega_\ell \propto \exp(-0.1\ell)$. Suitably stacking the weights $\{\omega_\ell\}$, we can define a linear map $\bar{\Pi}$ such that $\bar{\boldsymbol{\pi}} = \bar{\Pi} \times \boldsymbol{\pi}$.

We then represent the loss function of a dual mandate policymaker with preferences over average inflation as

$$\mathcal{L} = \lambda_\pi \bar{\boldsymbol{\pi}}' W \bar{\boldsymbol{\pi}} + \lambda_y \boldsymbol{y}' W \boldsymbol{y}$$

with $\lambda_\pi = \lambda_y = 1$, $W = \text{diag}(1, \beta, \beta^2, \cdots)$ and $\beta = 1/1.01$. Results for our optimal policy counterfactual are displayed in Figure 6. The key takeaway here is that this optimal policy counterfactual differs very little from actually observed outcomes. In other words, there is little room to improve upon the observed allocation by changing policy within the space of policy instrument paths spanned by our two identified policy shocks. Furthermore allowing for ex post surprises does not materially change this conclusion.

Finally, detailed results for our two other counterfactuals are presented in Appendix B.4. We briefly review the main findings here. Fourth, it is challenging to implement a nominal interest peg using only date-0 shocks—in particular at short horizons, nominal interest rates in our best approximation still fall a bit too much. Small ex post surprises, however, are sufficient to almost perfect stabilize rates. In either case, output in this counterfactual contracts by more, and inflation is materially lower at medium horizons. We emphasize that this policy counterfactual corresponds to our version of the classical policy "zero-ing out" exercise routinely implemented in classical VAR analysis through the Sims & Zha approach (e.g. Bernanke et al., 1997; Hamilton & Herrera, 2004; Brunnermeier et al., 2021). Fifth, strict nominal GDP targeting can be implemented quite accurately with only date-0 shocks. Interestingly, this counterfactual looks quite similar to our estimated outcomes under the baseline rule, with rates cut only slightly less aggressively.

Discussion. The results from our applications reveal that existing empirical evidence on policy shocks is already sufficient to tightly restrict policy counterfactuals for several prominent alternative monetary policy strategies. At the same time, our empirical method is clearly not *always* applicable: for some non-policy shocks and some counterfactual rules, it will not be possible to enforce the counterfactual rule accurately. We emphasize that the counterfactuals implied by our baseline method for the investment shock application were relatively accurate precisely because the investment shock is rather transitory, thus only requiring knowledge of the effects of similarly transitory interest rate changes, along the lines of those in Figure 3. More persistent shocks necessarily induce more persistent policy instrument movements and thus would require more empirical evidence on such highly persistent policy shocks (e.g. far-ahead forward guidance). The next section briefly discusses the role of explicit structural modeling in constructing such policy rule counterfactuals.

4 Extrapolating policy shock causal effects

Our identification results in Section 2 provide a new perspective on the role played by structural models in the "Lucas program" (see Christiano et al., 1999). Recall that this program identifies and then matches empirical evidence on a single policy shock, thus pinning down the effects of one particular policy instrument path; the purpose of the model is then to extrapolate from that evidence to all other possible policy instrument paths. Our applications in Section 3 reveal that, in some cases, empirical work has already identified sufficiently rich policy instrument paths to tightly characterize policy counterfactuals, allowing us to sidestep model-based extrapolation. In other cases, however, the existing empirical evidence may not suffice and thus model extrapolation will be needed, at least until empirical measurement improves further.

This section sheds some light on how this extrapolation from one policy shock to others is achieved in "typical" structural macroeconomic models. Our general insight is that the causal effects of policy shocks at different horizons tend to be closely tied together in standard models: that is, information about the effects of policy shocks at one horizon very tightly restricts the level and shape of effects of such shocks at other horizons. Section 4.1 begins with an example, showing that—for a particular but important class of monetary policy counterfactuals—all of the required causal effect extrapolation is actually governed by one partial model block: the Phillips curve. Section 4.2 then offers a general analysis.

4.1 Output-inflation counterfactuals with a partial model

We begin by restriting attention to a particular but important family of systematic policy rule counterfactuals. We consider a researcher interested in the behavior of the output gap and inflation under counterfactual policy rules of the particular form

$$\tilde{\mathcal{A}}_\pi \boldsymbol{\pi} + \tilde{\mathcal{A}}_y \boldsymbol{y} = \mathbf{0} \tag{33}$$

For example, (33) nests traditional flexible inflation targeting, average inflation targeting, nominal GDP targeting, as well as strict output gap and inflation stabilization. Counterfactual rules of the sort (33) are thus of substantial interest.

By our results in Section 2, knowledge of the two causal effect maps $\Theta_{\pi,\nu,\mathcal{A}}$ and $\Theta_{y,\nu,\mathcal{A}}$ is sufficient to construct counterfactuals for alternative policy rules like (33). More precisely, we in fact only require knowledge of *relative* policy shock impulse responses: if $\Theta_{\pi,\nu,\mathcal{A}}$ is invertible, then the proof of our identification result applies without any change using only knowledge of $\Theta_{y,\nu,\mathcal{A}} \times \Theta_{\pi,\nu,\mathcal{A}}^{-1}$. Intuitively, for rules of the form (33), we can effectively treat inflation as the policy instrument and then use the relative (or normalized) causal effects $\Theta_{y,\nu,\mathcal{A}} \times \Theta_{\pi,\nu,\mathcal{A}}^{-1}$ to determine the output path associated with a given inflation path.[20] The simple insight of this section is that, for any structural model that features a Phillips curve relationship between y and π, that Phillips curve already fully pins down the required relative causal effects, completely independently of the rest of the model.

Phillips curves as restrictions on causal effects. Consider a structural model that features a Phillips curve relationship—that is, a link between inflation and leads and lags of the aggregate output gap. Using our perfect-foresight notation of Section 2, we can write a general dynamic Phillips curve relationship as

$$\boldsymbol{\pi} = \Pi_y \times \boldsymbol{y} + \Pi_\varepsilon \times \boldsymbol{\varepsilon}. \tag{34}$$

In this relation, the matrix Π_y governs the link between inflation and the output gap up to (non-policy) shocks $\Pi_\varepsilon \times \boldsymbol{\varepsilon}$. For example, in the textbook three-equation New Keynesian model, Π_y would take the simple form

$$\Pi_y = \begin{pmatrix} \kappa & \kappa\beta & \kappa\beta^2 & \dots \\ 0 & \kappa & \kappa\beta & \dots \\ 0 & 0 & \kappa & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{35}$$

Monetary policy shocks move the economy along this Phillips curve. Therefore the crucial implication of (34) is that the responses of output and inflation induced by policy shocks $\boldsymbol{\nu}$ are related by

$$\Theta_{\pi,\nu,\mathcal{A}} = \Pi_y \times \Theta_{y,\nu,\mathcal{A}}. \tag{36}$$

In words, we can map output gap impulse responses into inflation responses (and vice-versa) using only the matrix Π_y. Knowledge of Π_y^{-1} is thus exactly what is needed to construct counterfactuals for alternative policy rules of the general form (33).

Structural assumptions on a model's Phillips curve tend to imply low-dimensional pa-

[20]The assumption that the policymaker can implement any desired path of inflation is generally satisfied in standard business-cycle models. For example, in the simple model of Section 2.1, it is straightforward to verify that $\Theta_{\pi,\nu,\mathcal{A}}$ is an upper-triangular, invertible matrix. We provide further details in Appendix C.1.

rameterizations of Π_y (e.g. as in (35)). Given such structure, knowledge of the causal effects of one policy shock—informative about one column (or linear combination of columns) of Π_y—will thus identify all of Π_y, thereby pinning down policy rule counterfactuals. This result is thus the first example of the general insight of this section: policy shocks at one horizon are often extremely informative about those at other horizons, with the restrictions across horizons in this particular case fully summarized by the Π_y matrix implied by the parametric form of the Phillips curve.

Applications. We illustrate the usefulness of the above insight with a return to our monetary policy counterfactual experiments from Section 3.3. Three of the experiments that we analyzed there fall into the class of rules (33) considered here: output gap targeting, nominal GDP targeting, and optimal average inflation targeting. We discuss results for the output gap targeting counterfactual here, and relegate details on the other two to Appendix C.1.

We assume that Π_y is derived from an empirically relevant hybrid Phillips curve relationship (see e.g. Mavroeidis et al., 2014)

$$\pi_t = \gamma_b \pi^4_{t-1} + \gamma_f \mathbb{E}_t \left[\pi^4_{t+4}\right] + \kappa y_t + \varepsilon_t \tag{37}$$

where $\pi^4_{t-1} = \frac{1}{4} \times (\pi_{t-1} + \pi_{t-2} + \pi_{t-3} + \pi_{t-4})$. Appendix C.1 shows the linear map Π_y corresponding to this Phillips curve specification. We then estimate the parameters $\{\gamma_b, \gamma_f, \kappa\}$ (and so all of Π_y) using evidence on identified monetary policy shocks. The econometric challenge is that the estimated policy effects $\{\Omega_{\pi,\mathcal{A}}, \Omega_{y,\mathcal{A}}\}$ will not perfectly align with the parametric structure imposed by (37); thus, following Barnichon & Mesters (2020), we simply find the best possible fit. Our estimation uses the shocks of Romer & Romer (2004).

Given an estimate of Π_y, we can construct the desired counterfactual: output gap and inflation impulse responses to investment-specific technology shocks under the counterfactual output gap targeting policy. The results are reported in Figure 7. Note first that the output gap is now stabilized perfectly (rather than approximately, as in Section 3.3).[21] The results show persistently elevated inflation relative to the baseline rule outcome, even more than in our estimates in Section 3.3. While the inflation counterfactual is similar to Figure 4 for short and medium horizons, inflation at longer horizons remains more elevated. This persistence of inflation reflects the strong backward-looking component in our estimated

[21] The assumed invertibility of $\Theta_{\pi,\nu,\mathcal{A}}$ together with the invertible Π_y implies invertibility of $\Theta_{y,\nu,\mathcal{A}}$ via (36), so perfect output gap stabilization is in fact implementable.

Policy Counterfactual via PC Extrapolation, Output Gap Targeting

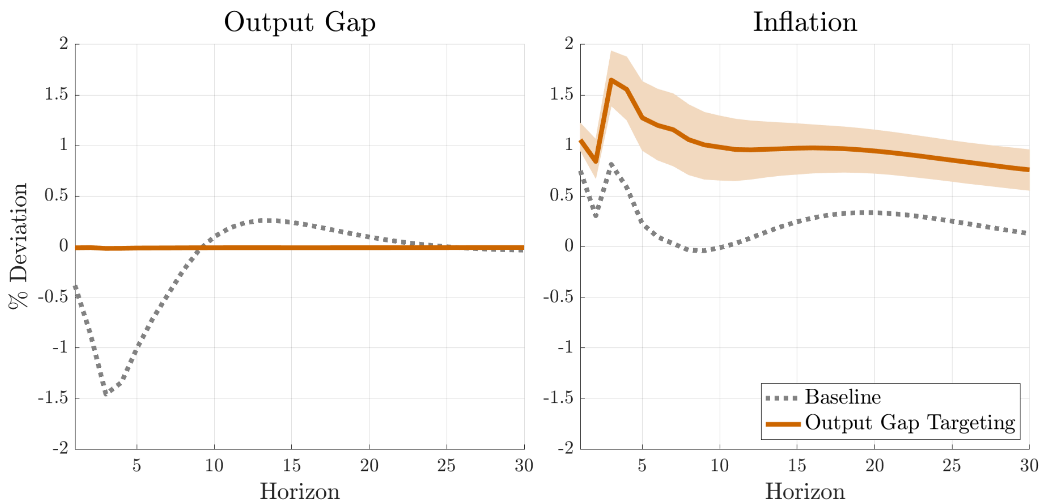

Figure 7: Output gap and inflation impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and a counterfactual rule that perfectly stabilizes the output gap (orange). The shaded areas correspond to 16th and 84th percentile confidence bands.

parametric Phillips curve. Thus, for this experiment, the particular parametric assumptions on the dynamic Phillips curve relationship embedded in (37) exert a quite strong influence at long horizons—and so whether or not these long-horizon counterfactual predictions are credible depends on the validity of those particular parametric assumptions.[22]

Discussion. We have seen in this section that, for a particular (but arguably quite important) family of counterfactual policy rules, "typical" structural business-cycle models achieve the extrapolation of impulse responses from one policy shock to all others in a particular and very restrictive way: the dynamic output-inflation co-movements are governed by the model's Phillips curve. If the researcher is willing to impose the structure of Phillips curve, then knowledge of the causal effects of a single policy shock is a highly informative "identified moment" (Nakamura & Steinsson, 2018) for policy rule counterfactuals. In particular, *any* fully specified general equilibrium structural model that (i) fits into the general form (6) - (7), (ii) features a Phillips curve relationship of the form (37) and (iii) is consistent with the empirical monetary policy shock estimates of Romer & Romer (2004) will produce the same

[22]The estimated backward-looking component in our setting is in fact large enough to imply that, for perfect output gap stabilization, inflation dynamics are non-stationary.

counterfactuals as in Figure 7, independently of any further assumptions on preferences, technology, and expectation formation.

4.2 General informativeness and asymptotic time invariance

Section 4.1 gave a first analytical example of how the causal effects of policy shocks at different horizons may be closely tied together. Economic intuition suggests that this observation may be more general: for example, we may expect the effects of a policy (news) shock h periods from now to be similar to the effects of a shock $h-1$ periods from now, just shifted by one time period. Here we argue that this is indeed the case in standard macro models.

Informativeness in a large-scale structural model. To formalize the idea that policy shock causal effects across different shock horizons may be tied together we build on Andrews et al. (2020). Those authors provide a measure of the informativeness of certain estimable moments—in our case the causal effects of a given identified policy shock—for some counterfactual of interest—in our case policy rule change counterfactuals, which as we have seen depend only on the causal effects of contemporaneous and news policy shocks. Our Phillips curve analysis in Section 4.1 is a special case in which a very small number of estimable moments are fully informative for an important family of counterfactuals. In the general case, our application of the approach of Andrews et al. quantifies informativeness in the form of an R^2 of counterfactual of interest on estimable moments.

Appendix C.2 presents an application to the popular large-scale structural model of Smets & Wouters (2007). We show there that, even though this model is very richly parameterized, the policy shock causal effect maps Θ at the heart of our identification results (approximately) live in small-dimensional subspaces, as formalized by a high R^2 of individual policy shock causal effects for the entirety of Θ. Appendix C.3 then explains why: impulse responses to different news policy shocks are indeed simply time-shifted versions of each other, thus tightly restricting the co-movement of policy shock impulse responses at different horizons.

Discussion. Our interpretation of this "informativeness" analysis is that, at least conditional on conventional macroeconomic model structures, already identified (short-run) policy shocks are highly informative about the effects of other policy shocks and so the universe of policy rule counterfactuals—an "identified moment" result in the spirit of Nakamura & Steinsson (2018) that justifies the practice of model estimation via impulse response matching (as in Christiano et al., 2005). Of course this particular way of extrapolating policy shock impulse responses may simply be a feature of "typical" models, and not of data. Viewed in this light, empirical evidence on further long-run (news) policy shocks would be particularly welcome, both to check the plausibility of model-implied extrapolation and—in our opinion even more usefully—to expand the set of counterfactual rules that can be enforced directly from the empirical estimates as in Section 3.

5 Conclusions

The standard approach to counterfactual analysis for changes in policy rules relies on fully-specified general equilibrium models. Our identification results instead point in a different direction: researchers can estimate the causal effects of distinct policy shocks and combine them to form policy counterfactuals. Importantly, these counterfactuals are valid in a large class of models that encompasses the majority of structural business-cycle models that are currently used for policy analysis.

An important challenge in implementing this strategy is that its informational requirements are high. We showed how to proceed in the empirically relevant case of evidence on multiple but finitely many policy shocks. We illustrated through several examples that empirical evidence is already sufficient to tightly characterize a variety of interesting monetary policy rule change counterfactuals, reducing the need for explicit structural modeling. More generally, a key message of this paper is to emphasize the value of empirical strategies that recover the dynamic causal effects associated with different *time paths* of policy instruments. Every additional piece of empirical evidence on a different policy instrument path will expand the space of counterfactual policy rules that can be analyzed.

References

Andrews, I., Gentzkow, M., & Shapiro, J. M. (2020). On the informativeness of descriptive statistics for structural estimates. *Econometrica*, *88*(6), 2231–2258.

Antolin-Diaz, J., Petrella, I., & Rubio-Ramírez, J. F. (2021). Structural scenario analysis with SVARs. *Journal of Monetary Economics*, *117*, 798–815.

Auclert, A., Bardóczy, B., Rognlie, M., & Straub, L. (2021). Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models. Working Paper.

Auclert, A., Rognlie, M., & Straub, L. (2018). The Intertemporal Keynesian Cross. Technical report, National Bureau of Economic Research.

Auclert, A., Rognlie, M., & Straub, L. (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model. Technical report, National Bureau of Economic Research.

Barnichon, R. & Mesters, G. (2020). Identifying modern macro equations with old shocks. *The Quarterly Journal of Economics*, *135*(4), 2255–2298.

Barnichon, R. & Mesters, G. (2021). Testing Macroeconomic Policies with Sufficient Statistics. Working Paper.

Ben Zeev, N. & Khan, H. (2015). Investment-specific news shocks and US business cycles. *Journal of Money, Credit and Banking*, *47*(7), 1443–1464.

Beraja, M. (2020). Counterfactual equivalence in Macroeconomics. Working Paper.

Bernanke, B. S., Gertler, M., Watson, M., Sims, C. A., & Friedman, B. M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings papers on economic activity*, *1997*(1), 91–157.

Boppart, T., Krusell, P., & Mitman, K. (2018). Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control*, *89*, 68–92.

Brunnermeier, M., Palia, D., Sastry, K. A., & Sims, C. A. (2021). Feedbacks: financial markets and economic activity. *American Economic Review*, *111*(6), 1845–79.

Carroll, C. D., Crawley, E., Slacalek, J., Tokuoka, K., & White, M. N. (2018). Sticky expectations and consumption dynamics. Technical report, national bureau of economic research.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ.*, *1*(1), 451–488.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics*, *1*, 65–148.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, *113*(1), 1–45.

De Groot, O., Mazelis, F., Motto, R., & Ristiniemi, A. (2021). A Toolkit for Computing Constrained Optimal Policy Projections (COPPs). ECB Working Paper.

Eberly, J. C., Stock, J. H., & Wright, J. H. (2020). The federal reserve's current framework for monetary policy: A review and assessment. *International Journal of Central Banking*, *16*(1), 5–71.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., & Watson, M. W. (2007). Abcs (and ds) of understanding vars. *American economic review*, *97*(3), 1021–1026.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., & Schorfheide, F. (2016). Solution and estimation methods for DSGE models. In *Handbook of Macroeconomics*, volume 2 (pp. 527–724). Elsevier.

Gabaix, X. (2020). A behavioral New Keynesian model. *American Economic Review*, *110*(8), 2271–2327.

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

Gertler, M. & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76.

Guren, A., McKay, A., Nakamura, E., & Steinsson, J. (2021). What Do We Learn from Cross-Regional Empirical Estimates in Macroeconomics? *NBER Macroeconomics Annual*, *35*(1), 175–223.

Gürkaynak, R., Sack, B., & Swanson, E. (2005). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking*, *1*(1), 55–93.

Hamilton, J. D. & Herrera, A. M. (2004). Oil shocks and aggregate macroeconomic behavior: The role of monetary policy. *Journal of Money, Credit and Banking*, *36*(2), 265–286.

Hebden, J. & Winker, F. (2021). Impulse-Based Computation of Policy Counterfactuals. Working Paper.

Inoue, A. & Rossi, B. (2021). A new approach to measuring economic policy shocks, with an application to conventional and unconventional monetary policy. Working Paper.

Justiniano, A., Primiceri, G. E., & Tambalotti, A. (2010). Investment shocks and business cycles. *Journal of Monetary Economics*, *57*(2), 132–145.

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, *111*(4), 1092–1125.

Kaplan, G., Moll, B., & Violante, G. (2018). Monetary Policy according to HANK. *American Economic Review*, *108*(3), 697–743.

Kocherlakota, N. R. (2019). Practical policy evaluation. *Journal of Monetary Economics*, *102*, 29–45.

Leeper, E. M., Walker, T. B., & Yang, S.-C. S. (2013). Fiscal foresight and information flows. *Econometrica*, *81*(3), 1115–1145.

Leeper, E. M. & Zha, T. (2003). Modest policy interventions. *Journal of Monetary Economics*, *50*(8), 1673–1700.

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

Lucas, R. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester Conference Series on Public Policy*, volume 1, (pp. 19–46). Elsevier.

Mavroeidis, S., Plagborg-Møller, M., & Stock, J. H. (2014). Empirical evidence on inflation expectations in the new keynesian phillips curve. *Journal of Economic Literature*, *52*(1), 124–88.

McKay, A. & Wieland, J. F. (2021). Lumpy Durable Consumption Demand and the Limited Ammunition of Monetary Policy. Forthcoming in Econometrica.

Mertens, K. & Ravn, M. O. (2010). Measuring the impact of fiscal policy in the face of anticipation: a structural var approach. *The Economic Journal*, *120*(544), 393–413.

Nakamura, E. & Steinsson, J. (2018). Identification in macroeconomics. *Journal of Economic Perspectives*, *32*(3), 59–86.

Ottonello, P. & Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, *88*(6), 2473–2502.

Plagborg-Møller, M. & Wolf, C. K. (2021a). Instrumental variable identification of dynamic variance decompositions. Technical report, National Bureau of Economic Research.

Plagborg-Møller, M. & Wolf, C. K. (2021b). Local projections and VARs estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics*, *126*(1), 1–50.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of macroeconomics*, *2*, 71–162.

Ramey, V. A. & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of political economy*, *126*(2), 850–901.

Romer, C. D. & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084.

Rotemberg, J. J. & Woodford, M. (1997). An optimization-based econometric framework for the evaluation of monetary policy. *NBER macroeconomics annual*, *12*, 297–346.

Schmitt-Grohé, S. & Uribe, M. (2012). What's news in business cycles. *Econometrica*, *80*(6), 2733–2764.

Sims, C. A. & Zha, T. (2006). Does monetary policy generate recessions? *Macroeconomic Dynamics*, *10*(2), 231–272.

Smets, F. & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American economic review*, *97*(3), 586–606.

Stock, J. H. & Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal*, *128*(610), 917–948.

Svensson, L. E. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European economic review*, *41*(6), 1111–1146.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester conference series on public policy*, volume 39, (pp. 195–214). Elsevier.

Wieland, J. F. & Yang, M.-J. (2020). Financial dampening. *Journal of Money, Credit and Banking*, *52*(1), 79–113.

Wolf, C. K. (2020). The missing intercept: A demand equivalence approach. Working Paper.

Wolf, C. K. (2021). Interest Rate Cuts vs. Stimulus Payments: An Equivalence Result. Working Paper.

Woodford, M. (2011). *Interest and prices: Foundations of a theory of monetary policy.* princeton university press.

Online Appendix for:
What Can Time-Series Regressions Tell Us About Policy Counterfactuals?

This online appendix contains supplemental material for the article "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?". We provide (i) supplementary results complementing our theoretical identification analysis in Section 2 as well as implementation details for (ii) our empirical methodology in Section 3 and (iii) the "identified moment" structural analysis of Section 4.

Any references to equations, figures, tables, assumptions, propositions, lemmas, or sections that are not preceded "A."—"C." refer to the main article.

Contents

A Supplementary theoretical results **46**

A.1 Examples of nested models . . . 46

A.2 Filtering problems . . . 48

A.3 More general loss functions . . . 49

A.4 Counterfactual second-moment properties . . . 49

A.5 Optimal policy counterfactual in HANK . . . 51

A.6 Counterfactuals with finitely many shocks . . . 53

A.7 Global identification argument . . . 54

B Details for empirical finite-shock analysis **58**

B.1 Econometric implementation . . . 58

B.2 Data . . . 59

B.3 Shock & policy dynamic causal effects . . . 59

B.4 Supplementary results for our applications . . . 61

B.5 Results for one-shock counterfactuals . . . 63

C Details for Section 4 **66**

C.1 Phillips curve theory & estimation . . . 66

C.2 Informativeness of policy shocks as identified moments . . . 68

C.3 Asymptotic time invariance . . . 74

A Supplementary theoretical results

This appendix provides several results complementing our theoretical identification analysis of Section 2. Appendix A.1 discusses examples of structural models that are nested by our results, Appendix A.2 gives an example of a model that is not, Appendix A.3 extends our optimal policy arguments to more general loss functions, Appendix A.4 provides the details for unconditional second-moment counterfactuals, Appendix A.5 studies optimal policy in our illustrative HANK model, Appendix A.6 shows how we compute counterfactuals with finitely many shocks, and finally Appendix A.7 provides a global identification analysis with even higher informational requirements.

A.1 Examples of nested models

We provide further details on three sets of models: the three-equation New Keynesian model of Section 2.1, a general class of behavioral models, and the HANK model of Section 2.4.

Three-equation NK model. We here state the three-equation model of Section 2.1 in the form of our general matrix system (6) - (7). We begin with the non-policy block. The Phillips curve can be written as

$$\begin{pmatrix} 1 & -\beta & 0 & \dots \\ 0 & 1 & -\beta & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \boldsymbol{\pi} - \kappa \boldsymbol{y} - \boldsymbol{\varepsilon}^s = 0,$$

while the Euler equation can be written as

$$-\sigma \begin{pmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \boldsymbol{\pi} + \begin{pmatrix} 1 & -1 & 0 & \dots \\ 0 & 1 & -1 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \boldsymbol{y} + \sigma \boldsymbol{i} = 0.$$

Letting $\boldsymbol{x} \equiv (\boldsymbol{\pi}', \boldsymbol{y}')'$, we can stack these linear maps into the form (6). Finally the policy rule can be written as

$$\phi_\pi \boldsymbol{\pi} - \boldsymbol{i} + \boldsymbol{\nu} = 0,$$

which directly fits into the form of (7) with $\boldsymbol{z} = \boldsymbol{i}$.

Behavioral model. Our general framework (6) - (7) nests popular behavioral models such as the cognitive discounting framework of Gabaix (2020) or the sticky information set-up of Carroll et al. (2018). We here provide a sketch of the argument for a particular example—the consumption-savings decision of behavioral consumers.

Let the linear map $\mathcal{E}$ summarize the informational structure of the consumption-savings problem, with entry (t, s) giving the expectations of consumers at time t about shocks at time s. Here an entry of 1 corresponds to full information and rational expectations, while entries between 0 and 1 can capture behavioral discounting or incomplete information. For example, cognitive discounting at rate θ would correspond to

$$\mathcal{E} = \begin{pmatrix} 1 & \theta & \theta^2 & \dots \\ 1 & 1 & \theta & \dots \\ 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

while sticky information with a fraction $1 - \theta$ receiving the latest information could be summarized as

$$\mathcal{E} = \begin{pmatrix} 1 & 1-\theta & 1-\theta & \dots \\ 1 & 1 & 1-\theta^2 & \dots \\ 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Let p denote an input to the household consumption-savings problem (e.g., income or interest rates). In sequence space, we can use the matrix $\mathcal{E}$ to map derivatives of the aggregate consumption function with respect to p, denoted $\mathcal{C}_p$, into their behavioral analogues $\tilde{\mathcal{C}}_p$ via

$$\tilde{\mathcal{C}}_p(t, s) = \sum_{q=1}^{\min(t,s)} [\mathcal{E}(q, s) - \mathcal{E}(q-1, s)]\mathcal{C}_p(t - q + 1, s - q + 1)$$

Behavioral frictions thus merely affect the matrices that enter our general non-policy block (6), but do not affect the separation of policy- and non-policy blocks at the heart of our identification result.

Quantitative HANK model. The HANK model used for our quantitative illustration in Section 2.4 is exactly the same as in Wolf (2021) (including the parameterization, except of course for the monetary policy rule). The non-policy shock $\boldsymbol{\varepsilon}$ is an AR(1) innovation to the model's Phillips curve with persistence 0.8.

A.2 Filtering problems

To illustrate how an asymmetry in information between the private sector and the policy authority can break our separation of the policy and non-policy blocks in (6) - (7) even for a linear model, we consider a standard Lucas (1972) island model with a slightly generalized policy rule. The policy authority sets nominal demand x_t according to the rule

$$x_t = \phi_y y_t + x_{t-1} + \varepsilon_t^m$$

where y_t denotes real aggregate output and ε_t^m is a policy shock with volatility σ_m. The private sector of the economy as usual yields an aggregate supply curve of the form

$$y_t = \theta(p_t - \mathbb{E}_{t-1} p_t)$$

where the response coefficient θ follows from a filtering problem and is given as

$$\theta = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_p^2}$$

with σ_z denoting the (exogenous) volatility of idiosyncratic demand shocks and σ_p denoting the (endogenous) volatility of the surprise component of prices, $p_t - \mathbb{E}_{t-1} p_t$. A straightforward guess-and-verify solution of the model gives

$$p_t = \frac{1}{1+\theta} x_t + \frac{\theta}{1+\theta} x_{t-1}$$

and so

$$\sigma_p^2 = \left(\frac{1}{1+\theta}\right)^2 \text{Var}(\phi_y y_t + \varepsilon_t^m)$$

But since

$$y_t = \frac{1}{1 - \frac{\theta}{1+\theta}\phi_y} \frac{\theta}{1+\theta} \varepsilon_t^m$$

it follows that θ depends on the policy rule coefficient ϕ_y, breaking our separation assumption.

A.3 More general loss functions

Proposition 2 can be generalized to allow for a non-separable quadratic loss function. Suppose the policymaker's loss function takes the form

$$\mathcal{L} = \boldsymbol{x}'Q\boldsymbol{x} \tag{A.1}$$

where Q is a weighting matrix. Following the same steps as the proof of Proposition 2, we can formulate the policy problem as minimizing the loss function (A.1) subject to (25). The first-order conditions of this problem are

$$\Theta'_{\nu,x,A}(Q+Q')\boldsymbol{x} = 0$$

so we can recover the optimal policy rule as

$$\begin{aligned} \mathcal{A}_x^* &= \Theta'_{\nu,x,A}(Q+Q') \\ \mathcal{A}_z^* &= \mathbf{0} \end{aligned}$$

Even outside of the quadratic case, the causal effects of policy shocks on $\boldsymbol{x}$ are still enough to formulate a set of necessary conditions for optimal policy, but in this general case the resulting optimal policy rule will not fit into the linear form (7).

A.4 Counterfactual second-moment properties

Our analysis is largely focussed on constructing counterfactuals for particular non-policy shock paths $\boldsymbol{\varepsilon}$. This is in keeping with much of the empirical policy counterfactual literature that followed the lead of Sims & Zha (2006) (e.g. Bernanke et al., 1997; Eberly et al., 2020; Antolin-Diaz et al., 2021). However, under some additional assumptions, our results can also be used to construct *unconditional* counterfactual second-moment properties—that is, predict how variances and covariances of macroeconomic aggregates would change under a counterfactual rule. This section provides the detailed argument.

Setting. We consider a researcher that observes and is interested in the counterfactual properties of some vector of macroeconomic aggregates $y = (x, z)$—the endogenous outcomes and policy instruments of our main analysis. We assume that, under the prevailing baseline policy rule, this vector of macroeconomic aggregates follows a standard structural vector moving average representation:

$$y_t = \sum_{\ell=0}^{\infty} \Theta_\ell \varepsilon_{t-\ell} = \Theta(L)\varepsilon_t \tag{A.2}$$

where $\varepsilon_t \sim N(0, I)$.[23] We would like to predict the second-moment properties of y_t under some counterfactual policy rule (8).

If the researcher can estimate the causal effects of all shocks ε_t on the outcomes y_t, then the identification argument is trivial: she simply applies Proposition 1 for each individual shock, stacks the resulting impulse responses into a new vector moving average representation $\tilde{\Theta}(L)$, and from here computes the counterfactual second-moment properties. This approach may however not be feasible, as it requires the researcher to be able to correctly disentangle all of the structural shocks driving the macro-economy.

Procedure. Our proposed procedure has three steps. First, the researcher estimates the Wold representation of the observables y_t. Second, using Proposition 1, she maps the impulse responses to the Wold errors into new impulse responses corresponding to the counterfactual policy rule. Third, she stacks those new impulse responses to arrive at a new vector moving average representation, and from this representation constructs a new set of second-moment properties. Our identification result states that, if the vector moving representation (A.2) under the baseline rule is invertible, then this procedure correctly recovers the desired counterfactual second moments.

Identification result. Let $\tilde{\Theta}_\ell$ denote the lag-ℓ impulse responses of the observables y_t to the shocks ε_t under the counterfactual policy rule. The process for y_t under the counterfactual policy rule thus becomes

$$y_t = \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \varepsilon_{t-\ell} = \tilde{\Theta}(L)\varepsilon_t$$

and so the second moments of the true counterfactual process are given by

$$\Gamma_y(\ell) = \sum_{m=0}^{\infty} \tilde{\Theta}_m \tilde{\Theta}'_{m+\ell} \tag{A.3}$$

[23] Given our focus on second moments, the normality restriction is purely for notational convenience (see e.g. Plagborg-Møller & Wolf, 2021b).

Now consider instead the output of our proposed procedure. Let u_t denote the Wold errors under the observed policy rule, and let ε_t^* denote any unit-variance orthogonalization of these Wold errors (e.g., $\varepsilon_t^* = \text{chol}(\text{Var}(u_t))^{-1} \times u_t$). Then y_t under the observed policy rule satisfies

$$y_t = \Psi(L)\varepsilon_t^* = \sum_{\ell=0}^{\infty} \Psi_\ell \varepsilon_{t-\ell}^*$$

where $\varepsilon_t^* \sim N(0, I)$. Under invertibility—i.e., $\Theta(L)$ has a one-sided inverse—we in fact know that $\varepsilon_t^* = P\varepsilon_t$, $\Psi(L) = \Theta(L)P'$, $PP' = P'P = I$. The second step of our procedure gives the counterfactual vector moving average representation

$$y_t = \tilde{\Psi}(L)\varepsilon_t^*$$

where $\tilde{\Psi}(L)$ gives the dynamic causal effects of $\varepsilon_t^* = P\varepsilon_t$ on y_t under the counterfactual rule. But since the causal effects of ε_t under the baseline rule are given as $\tilde{\Theta}(L)$, it follows that we must also have

$$\tilde{\Psi}(L) = \tilde{\Theta}(L)P'$$

But then the implied second-moment properties of y_t are given as

$$\Gamma_y(\ell) = \sum_{m=0}^{\infty} \tilde{\Psi}_m \tilde{\Psi}'_{m+\ell} = \sum_{m=0}^{\infty} \tilde{\Theta}_m P'P\tilde{\Theta}'_{m+\ell} = \sum_{m=0}^{\infty} \tilde{\Theta}_m \tilde{\Theta}'_{m+\ell} \tag{A.4}$$

which is exactly equal to (A.3), completing the argument.

Finally, we emphasize that this identification result inherently rests on the assumption of invertibility. Under invertibility, there is a static one-to-one mapping between true shocks ε_t and Wold errors ε_t^*; thus, if we can predict the propagation of the Wold errors under the counterfactual rule, then we also match the propagation of the true shocks, and so we correctly recover second-moment properties. Under non-invertibility, however, there is no analogous one-to-one mapping, and so it is not guaranteed that second moments will be matched by our procedure.

A.5 Optimal policy counterfactual in HANK

Section 2.4 illustrated the general counterfactual rule identification result in Proposition 1 using a quantitative HANK model. We here do the same for the analogous optimal policy identification result in Proposition 2.

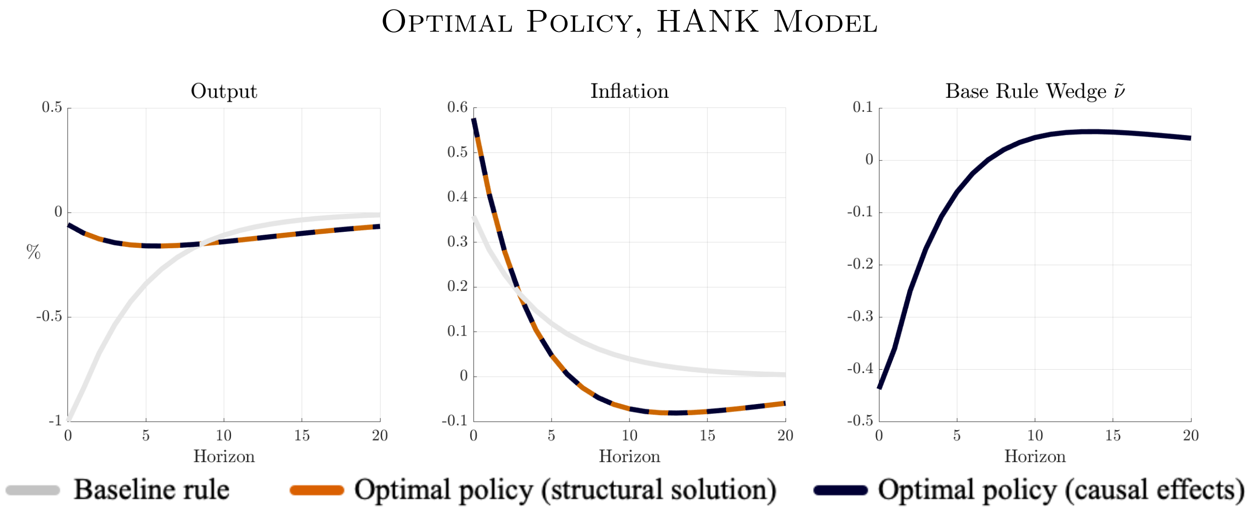

Figure A.1: Output and inflation impulse responses together with the equivalence shock wedge $\tilde{\boldsymbol{\nu}}$ (see (13)) for the HANK model with policy rules (28) and the optimal policy given by (A.5). The impact output contraction under the prevailing baseline rule is normalized to -1%.

We consider a policymaker with a standard dual mandate loss function

$$\mathcal{L} = \lambda_\pi \boldsymbol{\pi}'\boldsymbol{\pi} + \lambda_y \boldsymbol{y}'\boldsymbol{y} \tag{A.5}$$

with $\lambda_\pi = \lambda_y = 1$. As in Section 2.4 we start by solving for the optimal policy using conventional methods: we derive the policy rule corresponding to the first-order conditions (18) - (20), solve the model given that policy rule, and report the result as the orange lines in the left and middle panels of Figure A.1. We see that, at the optimum, the cost-push shock moves inflation by much more than output, consistent with the assumed policy weights and the relatively flat Phillips curve. Compared to this optimal policy, the simple baseline rule of the form (28) tightens too much.

We then instead use Proposition 2 to equivalently recover the optimal policy rule and the corresponding impulse responses. We begin with the optimal rule itself. By (26), the optimal rule is given as

$$\lambda_\pi \Theta'_{\pi,\nu,\mathcal{A}} \boldsymbol{\pi} + \lambda_y \Theta'_{y,\nu,\mathcal{A}} \boldsymbol{y} = 0$$

A researcher with knowledge of the effects of monetary policy shocks on inflation and output, $\{\Theta_{\pi,\nu,\mathcal{A}}, \Theta_{y,\nu,\mathcal{A}}\}$, is able to construct this optimal policy rule. We can then create a counterfactual response to the cost-push shock using (11)-(13), again requiring only knowledge of the causal effects of policy shocks as well as the impulse responses to the cost-push shock under the baseline rule. As expected, the resulting impulse responses—the dark blue lines—are identical to those obtained by explicitly solving the optimal policy problem. Finally, the right panel of Figure A.1 shows the optimal policy as a deviation $\tilde{\boldsymbol{\nu}}$ from the prevailing rule. The optimal rule accommodates the inflationary cost-push shock more than the baseline rule (28), so the required policy "shock" is persistently negative (i.e., expansionary).

A.6 Counterfactuals with finitely many shocks

This section provides further details for our finite-shock counterfactuals constructed in Figure 2. We begin with the one-shock case—the original proposal of Sims & Zha (2006). We then discuss the extension to a general finite number of observed policy shocks.

One Shock (Sims & Zha, 2006). This approach builds policy counterfactuals using empirical estimates of the dynamic causal effects of a single (contemporaneous) policy shock; that is, the researcher knows the first column of the maps in $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$. To predict the behavior of the economy under an alternative path of the policy instrument, the economy is then subjected to a *sequence* of contemporaneous policy shocks $\{\nu_{0,0}, \nu_{0,1}, \nu_{0,2}, \dots\}$ that enforce the desired instrument path in equilibrium. When translated to our notation, this simply corresponds to implementing our identification result not with the true (unknown) causal effect maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$, but instead using

$$\tilde{\Theta}_{q,\nu,\mathcal{A}} \equiv \begin{pmatrix} \Theta_{q,\nu,\mathcal{A}}(1,1) & 0 & 0 & \dots \\ \Theta_{q,\nu,\mathcal{A}}(2,1) & \Theta_{q,\nu,\mathcal{A}}(1,1) & 0 & \dots \\ \Theta_{q,\nu,\mathcal{A}}(3,1) & \Theta_{q,\nu,\mathcal{A}}(2,1) & \Theta_{q,\nu,\mathcal{A}}(1,1) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad q \in \{x, z\} \tag{A.6}$$

where $\Theta_{\bullet}(i,j)$ denotes the (i,j)th entry of a map $\Theta_{\bullet}$. This assumed structure implies that the first column parameterizes the full map—but of course that first column is exactly the impulse response estimated using the time-series regression. With this structure, surprising the economy with a suitable new shock each period is the same as announcing a sequence of contemporaneous and news shocks at $t = 0$ (i.e., our identification result) because the news shocks have no effect until they materialize at which point they are treated as if they were unanticipated.

We note that a structure like that in (A.6) is actually consistent with models populated by fully myopic agents. For example, in a variant of the behavioral New Keynesian model of

Gabaix (2020) with full discounting in both the consumer Euler equation and the firm-side Phillips curve, news shocks have no effect prior to their realization, so the true causal effect maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ in fact have the lower-triangular structure displayed in (A.6).[24] Typical (rational-expectations) macroeconomic models with forward-looking agents, on the other hand, have important expectational channels and so are inconsistent with the assumptions embedded in (A.6). In such environments, using the structure in (A.6) to predict the effects of changes in policy rules will run afoul of the Lucas critique.

Multiple shocks. The original approach of Sims & Zha (2006) leverages the idea that evidence on one policy shock—i.e., any single path $\boldsymbol{\nu}$—is sufficient to enforce any given counterfactual *ex post*. With n_s distinct shocks, the counterfactual rule can be implemented *ex post* as well as in *ex ante* expectation for the next $n_s - 1$ time periods.

To compute the counterfactuals corresponding to this multi-shock case as reported in Figure 2 we proceed as follows. First, at $t = 0$, we solve for the n_s-dimensional vector of policy shocks $\boldsymbol{\nu}^0_{1:n_s} \equiv (\nu^0_0, \ldots, \nu^0_{n_s-1})'$ such that, in response to $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}^0_{1:n_s}$ the counterfactual rule holds at $t = 0$ and is expected to hold for $t = 1, \ldots, n_s - 1$. Output and inflation at $t = 0$ are simply given as the thus-derived impulse responses to $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}^0_{1:n_s}$. Second, at $t = 1$, we solve for the n_s-dimensional vector of shocks $\boldsymbol{\nu}^1_{1:n_s} \equiv (\nu^1_0, \ldots, \nu^1_{n_s-1})'$ such that, in response to the time-0 shocks $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}^0_{1:n_s}\}$ and the time-1 shocks $\boldsymbol{\nu}^1_{1:n_s}$, the counterfactual policy rule holds at $t = 1$ and in expectation for $t = 2, \ldots, n_s$. These impulse responses then give us output and inflation at $t = 1$. Continuing iteratively, we obtain the entire output and inflation impulse responses, as plotted in Figure 2.

A.7 Global identification argument

We here extend our identification results to a general non-linear model with aggregate risk.

Setting. We consider an economy that runs for T periods overall. As in our main analysis, the economy consists of a private block and a policy block. Differently from our main analysis, there is no exogenous non-policy shock sequence $\boldsymbol{\varepsilon}$; rather, there is a stochastic event ω_t each period, with stochastic events drawn from a finite (n_ω-dimensional) set. Let $x_t(\omega^t)$ be the value of the endogenous variables after history $\omega^t \equiv \{\omega_0, \omega_1, \cdots, \omega_t\}$ and let $z_t(\omega^t)$ be the

[24] If agents are quite but not perfectly inattentive (as for example in Auclert et al., 2020), then the one-shock approach may deliver a reasonably accurate approximation to correct policy counterfactuals.

realization of the policy instruments after history ω^t. Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be the full contingent plans for for all $t \in \{0, 1, \cdots, T\}$ and all histories. $\boldsymbol{x}$ and $\boldsymbol{z}$ are vectors in $\mathbb{R}^{n_x \times N}$ and $\mathbb{R}^{n_z \times N}$ respectively, where $N = n_\omega + n_\omega^2 + \cdots + n_\omega^{T+1}$.

We can write the private-sector block of the model as the non-linear equation

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{z}) = \mathbf{0}. \tag{A.7}$$

Similarly, we can write the policy block corresponding to a baseline policy rule as

$$\mathcal{A}(\boldsymbol{x}, \boldsymbol{z}) + \boldsymbol{\nu} = \mathbf{0} \tag{A.8}$$

where the vector of policy shocks $\boldsymbol{\nu}$ is now $n_z \times N$ dimensional. We assume that, for any $\boldsymbol{\nu} \in \mathbb{R}^{n_z \times N}$, the system (A.7) - (A.8) has a unique solution. We write this solution as

$$\boldsymbol{x} = x(\boldsymbol{\nu}), \quad \boldsymbol{z} = z(\boldsymbol{\nu}).$$

We want to construct counterfactuals under the alternative policy rule

$$\tilde{\mathcal{A}}(\boldsymbol{x}, \boldsymbol{z}) = \mathbf{0} \tag{A.9}$$

replacing (A.8). We again assume that the system (A.7) and (A.9) has a unique solution, now written as $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}})$. If we are interested in the counterfactual following a particular path of exogenous events, then we are interested in selections from these vectors.

Proposition A.1. *For any alternative policy rule $\tilde{\mathcal{A}}$ we can construct the desired counterfactuals as*

$$x(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{x}}, \quad z(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{z}} \tag{A.10}$$

where $\tilde{\boldsymbol{\nu}}$ solves

$$\tilde{\mathcal{A}}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) = \mathbf{0}. \tag{A.11}$$

The solution $\tilde{\boldsymbol{\nu}}$ to this system exists and any such solution generates the unique counterfactual $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}})$.

Proof. We construct the solution $\tilde{\boldsymbol{\nu}}$ as

$$\tilde{\boldsymbol{\nu}} \equiv \tilde{\mathcal{A}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}) - \mathcal{A}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}).$$

By the definition of the functions of $x(\bullet)$ and $z(\bullet)$, we know that

$$\begin{aligned}
\mathcal{H}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) &= \mathbf{0} \qquad \text{(A.12)} \\
\mathcal{A}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) + \tilde{\mathcal{A}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}) - \mathcal{A}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}) &= \mathbf{0} \qquad \text{(A.13)}
\end{aligned}$$

Similarly, by the definition of the functions $\tilde{x}(\bullet)$ and $\tilde{z}(\bullet)$, we also know that

$$\begin{aligned}
\mathcal{H}(\tilde{x}(\mathbf{0}), \tilde{z}(\mathbf{0})) &= \mathbf{0} \qquad \text{(A.14)} \\
\tilde{\mathcal{A}}(\tilde{x}(\mathbf{0}), \tilde{z}(\mathbf{0})) &= \mathbf{0} \qquad \text{(A.15)}
\end{aligned}$$

It follows that $\{x(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{x}}, z(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{z}}\}$ is a solution of the system (A.12) - (A.13). By assumption this system has a unique solution, so it must be that $\tilde{\boldsymbol{\nu}}$ satisfies $\{x(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{x}}, z(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{z}}\}$.

We now show that any solution to (A.11) must generate $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}})$. Proceeding by contradiction, consider any other $\tilde{\boldsymbol{\nu}}$ that solves (A.11) and suppose that either $x(\tilde{\boldsymbol{\nu}}) \neq \tilde{\boldsymbol{x}}$ and/or $z(\tilde{\boldsymbol{\nu}}) \neq \tilde{\boldsymbol{z}}$. By definition of the functions $x(\bullet)$ and $z(\bullet)$ together with the property (A.11) we know that

$$\begin{aligned}
\mathcal{H}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) &= \mathbf{0} \\
\tilde{\mathcal{A}}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) &= \mathbf{0}
\end{aligned}$$

and so $(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}}))$ is a solution of (A.7) and (A.9) that is distinct from $(\tilde{x}, \tilde{z})$. But by assumption only one such solution exists, so we have a contradiction.

□

Informational requirements. To construct the desired policy counterfactual for all possible alternative policy rules, we in general need to be able to evaluate the functions $x(\bullet)$ and $z(\bullet)$ for every possible $\boldsymbol{\nu} \in \mathbb{R}^{n_z \times N}$. That is, we need to know the effects of policy shocks of all possible sizes at all possible dates and all possible histories.

To understand how our baseline analysis relaxes these informational requirements, it is useful to proceed in two steps: first removing uncertainty (but keeping non-linearity), and then moving to a linear system.

1. **Non-linear perfect foresight.** For a non-linear perfect foresight economy, we replace our general $(n_x + n_z) \times N$-dimensional system with an $(n_x + n_z) \times T$-dimensional one:

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\varepsilon}) = \mathbf{0}$$

$$\mathcal{A}(\boldsymbol{x}, \boldsymbol{z}) + \boldsymbol{\nu} \quad = \quad \mathbf{0}$$

Because of the lack of uncertainty, other possible realizations of the exogenous events—now denoted $\boldsymbol{\varepsilon}$—do not matter. Proceeding exactly in line with the analysis above, we can conclude that now we need the causal effects of all possible policy shocks $\boldsymbol{\nu} \in \mathbb{R}^{n_z \times T}$ *at the equilibrium path induced by* $\boldsymbol{\varepsilon}$. Thus, since we only care about the actual realized history of the exogenous inputs, the dimensionality of the informational requirements has been reduced substantially.

2. **Linear perfect foresight/first-order perturbation.** Linearity further reduces our informational requirements in two respects. First, because of linearity, to know the effects of every possible $\boldsymbol{\nu} \in \mathbb{R}^{n_z \times T}$, it suffices to know the effects of $n_z \times T$ distinct paths $\boldsymbol{\nu}$ that together span $\mathbb{R}^{n_z \times T}$. Second, estimates given any possible exogenous state path of the economy suffice, simply because the effects of policy and non-policy shocks are additively separable. We have thus reduced the problem to the (still formidable) one of finding the effects of $n_z \times T$ distinct policy shock paths.

B Details for empirical finite-shock analysis

This appendix provides further details for our empirical method and its applications in Section 3. Appendix B.1 begins by elaborating on our baseline empirical method and the Sims & Zha refinement. Appendices B.2 to B.4 then offer supplementary details for our monetary policy rule counterfactuals. Finally, in Appendix B.5, we contrast our results with those obtained from a standard single-shock approach as in Sims & Zha.

B.1 Econometric implementation

We here discuss the practical implementation of our baseline Lucas critique-robust empirical method as well as the refinement of the Sims & Zha (2006) method. Since our robust procedure is a general case of the general ridge regression problem (32) for $\psi = \infty$, we here simply present implementation details for the ridge regression version.

To express the solution to our basic ridge regression problem (32), we stack the policy shocks in the vector $\boldsymbol{s}^H$ and the corresponding causal effects in the matrix $\Omega_{x,\mathcal{A}}^H$. We furthermore let P denote a matrix that is equal to an $(n_s \cdot H) \times (n_s \cdot H)$-dimensional identity matrix except for the first n_s diagonal entries, which are equal to zero. The ridge regression solution is then given as

$$\boldsymbol{s}^H = -\left[\left(\tilde{\mathcal{A}}_x\Omega_{x,\mathcal{A}}^H + \tilde{\mathcal{A}}_z\Omega_{z,\mathcal{A}}^H\right)'\left(\tilde{\mathcal{A}}_x\Omega_{x,\mathcal{A}}^H + \tilde{\mathcal{A}}_z\Omega_{z,\mathcal{A}}^H\right) + \psi P'P\right]^{-1} \times \left[\left(\tilde{\mathcal{A}}_x\Omega_{x,\mathcal{A}}^H + \tilde{\mathcal{A}}_z\Omega_{z,\mathcal{A}}^H\right)'\left(\tilde{\mathcal{A}}_x\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \tilde{\mathcal{A}}_z\boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\right)\right]$$

For our optimal policy counterfactual, we analogously consider the following regularized optimal policy problem:

$$\min_{\boldsymbol{s}^H} \sum_{i=1}^{n_x} \lambda_i \boldsymbol{x}_i' W \boldsymbol{x}_i + \psi||P\boldsymbol{s}^H||$$

such that

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\varepsilon}) + \Omega_{x,\mathcal{A}}^H \boldsymbol{s}^H$$

This gives the optimality conditions:

$$\begin{aligned}(W \otimes \Lambda)\boldsymbol{x} + \boldsymbol{\varphi}_x &= \mathbf{0}\\ -\psi P\boldsymbol{s}^H + (\Omega_{x,\mathcal{A}}^H)'\boldsymbol{\varphi}_x &= \mathbf{0},\end{aligned}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots)$. Solving this system (together with the constraint of the problem) gives our optimal policy counterfactual. In particular, for $\psi = \infty$, we find the optimal counterfactual within the space of identified time-0 policy shock causal effects, without any ex post surprises.

B.2 Data

Our analysis of investment-specific technology shocks follows Ben Zeev & Khan (2015), while our monetary policy shock identification closely mimics that of (i) Romer & Romer (2004) and (ii) Gertler & Karadi (2015).

Outcomes. We are interested in impulse responses of three outcome variables: the output gap, inflation, and the policy rate. For the output gap and inflation, we follow Barnichon & Mesters (2020): we use the detrended real GDP gap (with the underlying trend estimated using the HP filter) as our measure of the output gap, and compute inflation from changes in the core PCE. Finally, we consider the federal funds rate as our measure of the policy rate, obtained from the St. Louis Federal Reserve FRED database. In keeping with much prior work, we also additionally control for commodity prices, with our measure obtained from the replication files of Ramey (2016). All series are quarterly.

Shocks & identification. We take the investment-specific technology shock series from Ben Zeev & Khan (2015) (`bzk_ist_news`), the Romer & Romer (2004) shock series from the replication and extension of Wieland & Yang (2020) (`rr_3`), and the high-frequency monetary policy surprise series from Gertler & Karadi (2015) (`mp1_tc`).[25] When applicable, the shock series are aggregated to quarterly frequency through simple averaging.

B.3 Shock & policy dynamic causal effects

For maximal consistency, we try to estimate all impulse responses within a common empirical specification. For the investment-specific technology shocks, we order the shock measure first in a recursive VAR containing our outcomes of interest (following Plagborg-Møller & Wolf, 2021b), estimated on a sample from 1969:Q1–2007:Q4. For our two monetary policy shocks, we estimate a single VAR in the two shock series, our three outcomes of interest, as well as

[25]Results are very similar if we use the alternative surprise series `ff4_tc` instead.

Policy Shock Impulse Responses, Romer & Romer (2004)

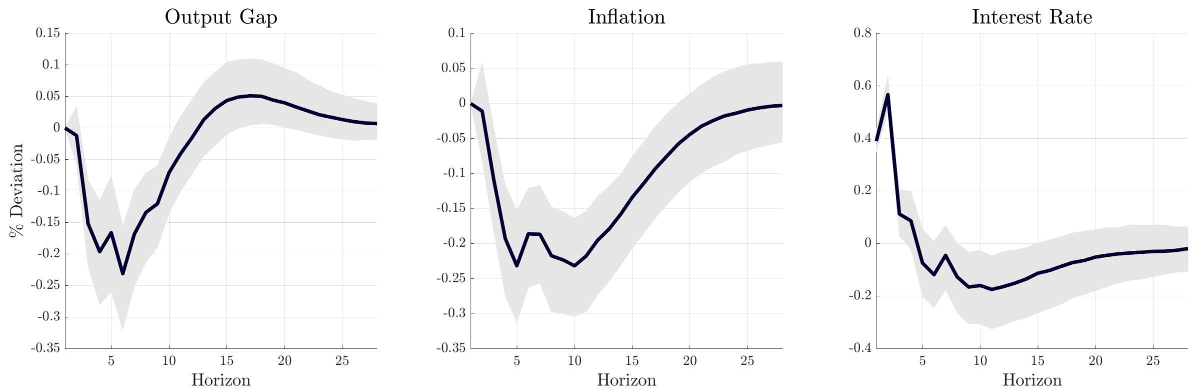

Figure B.1: Impulse responses after the Romer & Romer shock. The grey areas correspond to 16th and 84th percentile confidence bands, constructed using 10,000 draws from the posterior distribution of the reduced-form VAR parameters.

commodity prices, also estimated from 1969:Q1–2007:Q4.[26] For identification, we order the Gertler & Karadi shock first (again consistent with the results in Plagborg-Møller & Wolf (2021b)) and the Romer & Romer last (exactly as in Romer & Romer (2004)).

We use three lags in the technology shock specification, and four lags in the joint monetary policy VAR. We furthermore estimate all VARs with a constant as well as deterministic linear and quadratic trends. For the baseline investment-specific technology shock we fix the OLS point estimates. We then construct policy counterfactuals using our identified monetary policy shocks, taking into account their estimation uncertainty. Since the transmission of both shocks is estimated within a single VAR, we can draw from the posterior and compute the counterfactuals for each draw, thus taking into account joint estimation uncertainty.

Results. The OLS point estimates for the technology shocks of Ben Zeev & Khan (2015) are reported as the grey lines in Figure 4. For monetary policy, the estimated causal effects for our two outcomes of interest as well as the policy instrument are displayed in Figure B.1 (for Romer & Romer) and Figure B.2 (for Gertler & Karadi). The results are broadly in line with prior work: both policy shocks induce the expected signs and magnitudes of the output gap and inflation responses, though the response shapes are quite distinct, consistent with the differences in the induced interest rate paths.

[26]The Gertler & Karadi shock series is only available from 1988 onwards. We thus follow prior work in the macro IV literature (e.g. Känzig, 2021) and set the missing values to zero.

Policy Shock Impulse Responses, Gertler & Karadi (2015)

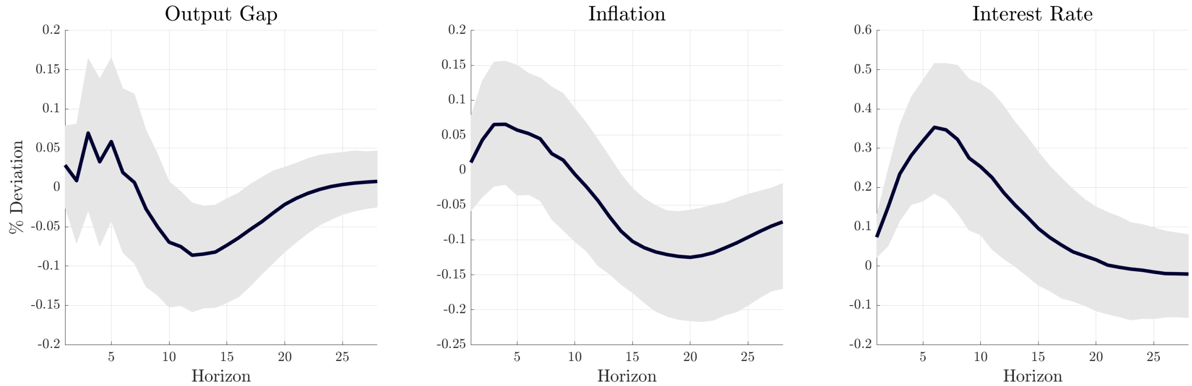

Figure B.2: Impulse responses after the Gertler & Karadi shock. The grey areas correspond to 16th and 84th percentile confidence bands, constructed using 10,000 draws from the posterior distribution of the reduced-form VAR parameters.

B.4 Supplementary results for our applications

In Section 3.3 we presented detailed results for only three of our counterfactuals: output gap targeting, the Taylor rule, and optimal average inflation targeting policy. We here provide the corresponding detailed results for the other two counterfactuals.

Nominal interest rate peg. Results for the nominal interest rate peg are presented in Figure B.3.[27] The counterfactual rule is implemented well from a couple of quarters out onwards, but rates are still cut by too much immediately after the shock. Alternatively, at the cost of a couple of 10 basis point nominal interest rate surprises within the first year after the shock, the interest rate is fixed almost perfectly. Since interest rates are now not cut (as much), the output gap and inflation remain low for a longer period of time. We emphasize that this counterfactual can equivalently be interpreted as giving us the causal effects of the investment technology shock with the effects of systematic monetary feedback solved out in a way that respects the Lucas critique.

[27] As is well-known, such a policy rule will in general not induce a unique equilibrium. In that case our empirical method will yield the equilibrium that corresponds to the same equilibrium selection as under the baseline rule (which was assumed to induce a unique equilibrium). In general this unique equilibrium is the minimal-state-variable equilibrium of the system.

Policy Counterfactual, Interest Rate Peg

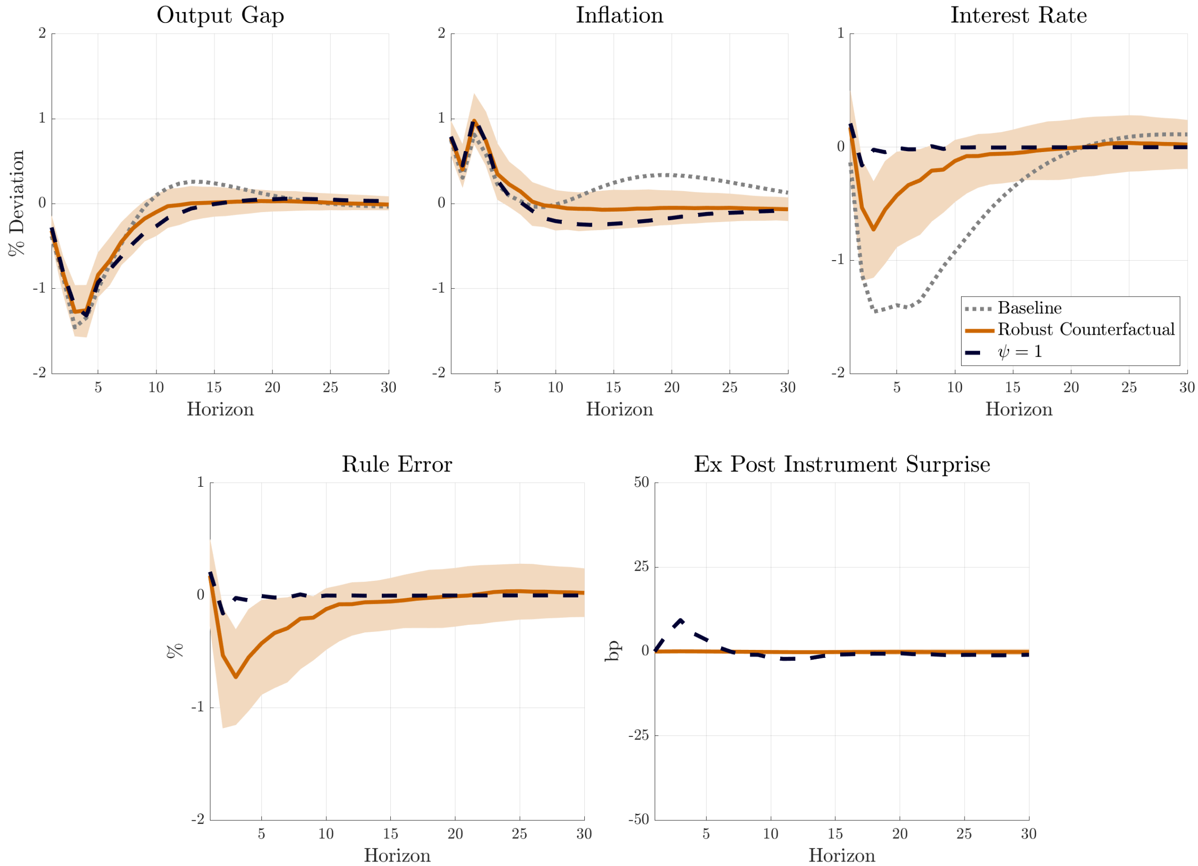

Figure B.3: Top panel: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the best feasible approximation to a nominal interest rate peg (orange and black dashed), computed following (31) and (32) for $\psi = 1$. Bottom panel: implementation error for the counterfactual rule and ex post nominal interest rate surprise at time t. The shaded areas correspond to 16th and 84th percentile confidence bands.

Nominal GDP targeting. Results for nominal GDP targeting are presented in Figure B.4. The counterfactual policy is implicitly defined by the targeting rule

$$\widehat{\pi}_t + (\widehat{y}_t - \widehat{y}_{t-1}) = 0, \quad \forall t = 0, 1, \ldots$$

We find that implementation errors are quite small throughout. Interestingly, the policy instrument path is quite close to the estimated baseline (dotted grey), indicating that nominal GDP is stabilized quite well already under the prevailing rule.

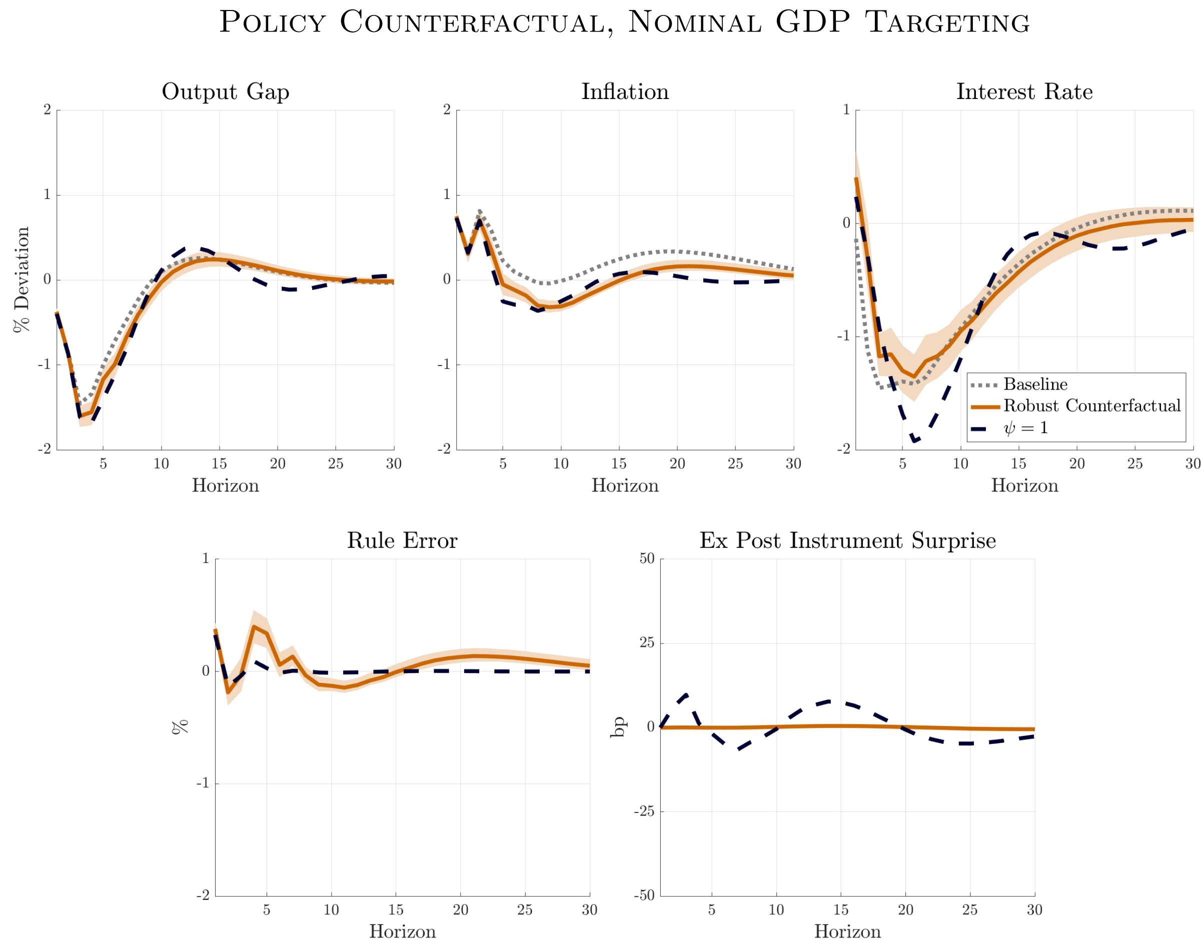

Figure B.4: Top panel: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the best feasible approximation to a nominal GDP targeting (orange and black dashed), computed following (31) and (32) for $\psi = 1$. Bottom panel: implementation error for the counterfactual rule and ex post nominal interest rate surprise at time t. The shaded areas correspond to 16th and 84th percentile confidence bands.

B.5 Results for one-shock counterfactuals

In this section we briefly compare the results from our empirical strategy with that of the canonical one-shock approach of Sims & Zha. As discussed previously, this approach corresponds a special case of our ridge regression strategy with $\psi = 0$, $n_s = 1$, and $H = T - 1$.

Explosive dynamics. Our first observation—already mentioned in Sims & Zha (2006)—is that the one-shock approach to constructing policy counterfactuals may produce explosive dynamics. The argument is simple and easily seen for the particular case of our interest rate peg counterfactual. Here, the sequence of shocks $\boldsymbol{s}^H$ satisfies

$$\boldsymbol{s}^H = -(\Omega^H_{z,\mathcal{A}})^{-1} \times \boldsymbol{z}(\boldsymbol{\varepsilon})$$

where z now indicates the nominal rate and $\Omega^H_{z,\mathcal{A}}$ is a triangular matrix with the vector $\Omega_{z,\mathcal{A}}$ below the main diagonal for each column. $\Omega^H_{z,\mathcal{A}}$ is thus a lower-triangular Toeplitz matrix. Letting $\{a_n\}_{n=0}^{\infty}$ denote the sequence characterizing that Toeplitz matrix, the sequence $\{b_n\}_{n=0}^{\infty}$ characterizing its inverse is given by the recursion $b_0 = 1/a_0$ and

$$b_n = -\frac{1}{b_0} \sum_{r=1}^{n} a_r b_{n-r}$$

It is straightforward to verify numerically that, for the OLS point estimates $\Omega^H_{z,\mathcal{A}}$ corresponding to our Romer & Romer (2004) and Gertler & Karadi (2015) shocks, the resulting inverse sequences $\{b_n\}_{n=0}^{\infty}$ and thus the shocks $\boldsymbol{s}^H$ diverge, thus indeed yielding explosive output gap and inflation dynamics corresponding to an exactly fixed nominal rate of interest.

One shock vs. multiple shocks. To avoid explosive dynamics in the one-shock case, our refinement of the original Sims & Zha procedure with $\psi = 1$ may be implemented for $n_s = 1$ (rather than $n_s = 2$, as in our main analysis). As expected, the counterfactual rule fit in that case is materially worse. For example, for the output gap targeting counterfactual, a one-shock counterfactual based on the Gertler & Karadi fails to achieve high rule accuracy (with a maximal error point estimate of 75 per cent), while the counterfactual based on the Romer & Romer shock requires large ex post nominal interest rate surprises. Figure B.5 illustrates with the results for the one-shock Gertler & Karadi counterfactual.

1-Shock Policy Counterfactual, Output Gap Targeting

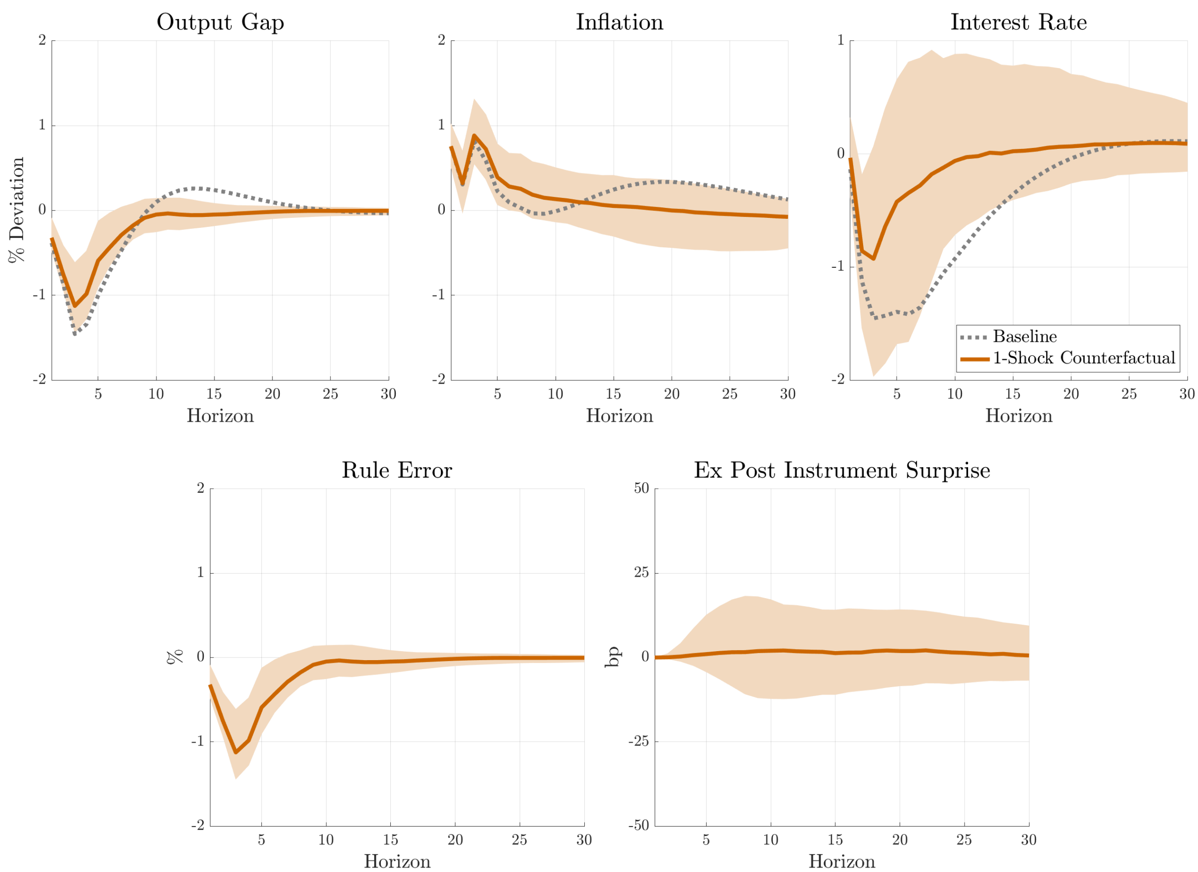

Figure B.5: Top panel: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the best feasible approximation to a output gap targeting (orange and black dashed) using only the shock of Gertler & Karadi and for $\psi = 1$. Bottom panel: implementation error for the counterfactual rule and ex post nominal interest rate surprise at time t. The shaded areas correspond to 16th and 84th percentile confidence bands.

C Details for Section 4

Appendix C.1 begins with our Phillips curve-based counterfactuals. Appendix C.2 then presents our general informativeness discussion and applies it to the well-known model of Smets & Wouters (2007), while Appendix C.3 rationalizes those results through the concept of asymptotic time invariance of impulse response functions.

C.1 Phillips curve theory & estimation

This section provides further details for our theoretical analysis and empirical application in Section 4.1.

RECOVERING POLICY COUNTERFACTUALS FROM Π_y. Knowledge of Π_y—together with the assumption that $\Theta_{\pi,\nu,\mathcal{A}}$ is invertible, i.e., any path of inflation is in principle implementable through policy shocks—is sufficient to construct output and inflation counterfactuals corresponding to alternative rules of the general form (33). Formally, using a change of basis for ν we can recover the desired counterfactual outcomes by solving the system

$$\begin{aligned} \tilde{\mathcal{A}}_\pi \boldsymbol{\pi} + \tilde{\mathcal{A}}_y \boldsymbol{y} &= 0 \\ \boldsymbol{\pi} &= \boldsymbol{\pi}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \boldsymbol{\nu} \\ \boldsymbol{y} &= \boldsymbol{y}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Pi_y^{-1} \boldsymbol{\nu} \end{aligned}$$

for the three unknowns $\{\boldsymbol{\pi}, \boldsymbol{y}, \boldsymbol{\nu}\}$.

Strictly speaking, the above result leveraging Π_y imposes the additional assumption that the monetary policymaker can in principle implement any desired path of inflation. This assumption is routinely satisfied in standard business-cycle models. For example, in our simple model of Section 2.1, it is straightforward to verify that $\Theta_{\pi,\nu,\mathcal{A}}$ is an upper-triangular matrix with

$$\Theta_{\pi,\nu,\mathcal{A}}(i,i) = -\frac{\kappa\sigma}{1+\kappa\sigma\phi_\pi}$$

and $\Theta_{\pi,\nu,\mathcal{A}}(i,j)$ for $i<j$ defined recursively via the system

$$\begin{aligned} \Theta_{y,\nu,\mathcal{A}}(i,j) &= -\sigma(\phi_\pi \Theta_{\pi,\nu,\mathcal{A}}(i,j) - \Theta_{\pi,\nu,\mathcal{A}}(i+1,j)) + \Theta_{y,\nu,\mathcal{A}}(i+1,j) \\ \Theta_{\pi,\nu,\mathcal{A}}(i,j) &= \kappa \Theta_{y,\nu,\mathcal{A}}(i,j) + \beta \Theta_{\pi,\nu,\mathcal{A}}(i+1,j) \end{aligned}$$

Special case: hybrid Phillips curve. Consider the hybrid Phillips curve (37). Along a perfect foresight transition path, we can write this relationship as

$$\underbrace{\begin{pmatrix} 1 & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & 0 & \dots \\ -\frac{1}{4}\gamma_b & 1 & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & \dots \\ -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & 1 & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & \dots \\ -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & 1 & -\frac{1}{4}\gamma_f & -\frac{1}{4}\gamma_f & \dots \\ -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & 1 & -\frac{1}{4}\gamma_f & \dots \\ 0 & -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & -\frac{1}{4}\gamma_b & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}}_{\equiv \Pi_\pi} \times \boldsymbol{\pi} \quad = \quad \kappa \times \boldsymbol{y} \; + \; \boldsymbol{\varepsilon}^s$$

We thus have

$$\Pi_y \equiv \Pi_\pi^{-1} \times \kappa$$

Estimation details. Barnichon & Mesters (2020) show how to use estimates of monetary policy impulse responses to identify a Phillips curve relationship of the form (37). For our empirical analysis in Figure 7 we closely follow their estimation strategy; since we use almost the same data (see Appendix B.2), our estimation results are very similar to theirs. In particular, for our headline results in Figure 7 we also impose the constraint that $\gamma_f + \gamma_b = 1$, so our confidence sets are almost identical to those reported in panel (B) of Figure II in the original article (Barnichon & Mesters, 2020).

Additional counterfactuals. In addition to the output gap targeting counterfactual, we can also use the estimated Phillips curve relationship to revisit the nominal GDP targeting and optimal average inflation targeting policy rule policy rule counterfactuals studied in Section 3.3 and Appendix B.4. We do so in Figures C.1 and C.2. While the nominal GDP targeting counterfactual is very similar to our baseline result in Figure B.4, the optimal average inflation targeting counterfactual induces a somewhat smaller output decline, at the cost of somewhat more persistently elevated inflation. We note furthermore that, by Proposition 2, this particular counterfactual corresponds to the general optimal policy rule

$$\lambda_\pi \bar{\Pi}' \bar{\boldsymbol{\pi}} + \lambda_y (\Pi_y')^{-1} \boldsymbol{y} = \mathbf{0} \tag{C.1}$$

Policy Counterfactual via PC Extrapolation, Nominal GDP Targeting

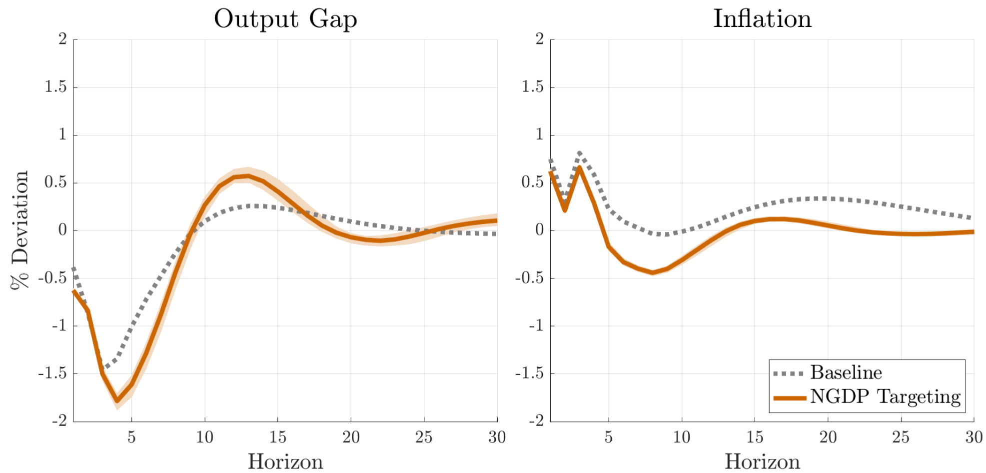

Figure C.1: Output gap and inflation impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and a counterfactual rule that targets nominal GDP (orange). The shaded areas correspond to 16th and 84th percentile confidence bands.

where $\bar{\pi}$ denotes the targeted average of current and lagged inflation and $\bar{\Pi}$ maps inflation into this targeted average, with $\bar{\boldsymbol{\pi}} \equiv \bar{\Pi} \times \boldsymbol{\pi}$ (see Section 3.3). (C.1) takes the form of an implicit targeting rule (Svensson, 1997): it imposes a set of restrictions that current, lagged and expected future values of inflation and the output gap must satisfy at all times when policy is set optimally.

C.2 Informativeness of policy shocks as identified moments

Building on Andrews et al. (2020), we provide a measure of the informativeness of particular estimable moments—the causal effects of certain estimable policy shocks—to the object of interest—structural policy rule counterfactuals. The identification results in Section 2 reveal that policy shock causal effects for enough shocks are sufficient statistics for policy rule changes, while economic intuition suggests that the effects of policy shocks across different horizons should be tightly related. Our analysis in this section confirms this intuition for a particular popular structural model: that of Smets & Wouters (2007).

Local informativeness in a general structural model. We consider a researcher that entertains a particular structural model $\zeta \in Z$, where ζ denotes a vector of the model's structural parameters. As a result of model estimation (or simply through some kind of prior

Policy Counterfactual via PC Extrapolation, Optimal AIT Policy Rule

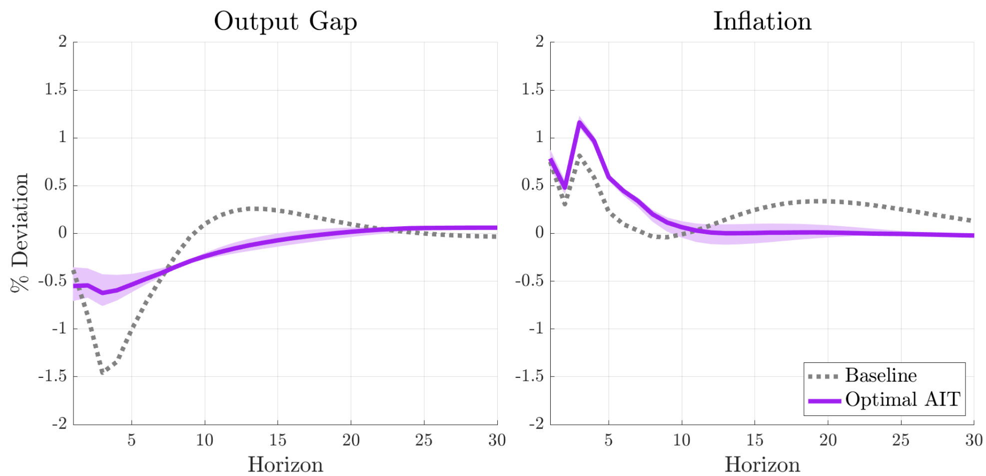

Figure C.2: Output gap and inflation impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the optimal average inflation targeting policy rule (purple). The shaded areas correspond to 16th and 84th percentile confidence bands.

information), the researcher entertains a distribution over that parameter vector:

$$\zeta \sim F(\zeta_0, \Sigma_\zeta).$$

The researcher is interested in some structural counterfactual c given as a function of the model's parameters, $c = c(\zeta)$. We seek to study the (local) informativeness of some other function of the model's parameters, $\gamma = \gamma(\zeta)$, for the counterfactual of interest, in a neighborhood of ζ_0. In the context of this paper c should be interpreted as counterfactual impulse response paths under alternative policy rules, while γ collects certain impulse responses to observable policy shocks.

Our formalization of the notion of informativeness is inspired by—though conceptually distinct from—Andrews et al. (2020).[28] In a neighborhood of ζ_0, the covariance matrix of (c, γ) is given as

$$\Sigma = \begin{pmatrix} \Sigma_c & \Sigma_{c\gamma} \\ \Sigma_{\gamma c} & \Sigma_\gamma \end{pmatrix} = \begin{pmatrix} \frac{\partial c(\zeta_0)}{\partial \zeta} \\ \frac{\partial \gamma(\zeta_0)}{\partial \zeta} \end{pmatrix} \Sigma_\zeta \begin{pmatrix} \frac{\partial c(\zeta_0)}{\partial \zeta}' \\ \frac{\partial \gamma(\zeta_0)}{\partial \zeta}' \end{pmatrix}$$

[28]We compute the exact same measure of informativeness as Andrews et al., (C.2). The interpretation, however, is rather different: Andrews et al. jointly estimate a model as well as descriptive statistics (their γ), while we study the informativeness of certain features of the model (our γ) for others (our c) conditional on the particular estimated model.

For any individual scalar entry $c_i \in c$, we then compute the following measure of the (local) informativeness of γ for c_i:

$$\Delta_i \equiv \frac{\Sigma_{c_i \gamma} \Sigma_\gamma^{-1} \Sigma_{\gamma c_i}}{\Sigma_{c_i}} \in [0, 1] \tag{C.2}$$

The informativeness measure Δ_i answers the following question: how tightly does knowledge of the observables γ restrict the counterfactual c_i? If for example γ contains impulse responses to certain policy shocks, and the counterfactual c_i can be obtained as a linear combination of these shocks (our analysis from Section 3), then $\Delta_i = 1$. If on the other hand c_i depends mostly on policy shocks at other horizons, and the structural model implies little in the way of cross-column restrictions on the impulse response maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$, then Δ_i will be low. Of course, once γ is large enough, we can invert the mapping $\gamma(\zeta)$ to back out ζ and therefore $c(\zeta)$, trivially giving $\Delta_i = 1$. Our question is whether we can have $\Delta_i \approx 1$ for certain small-dimensional yet in principle observable γ. If so, then we would have shown that the model robustly maps the given γ into the same counterfactual irrespective of the particular model parameterization, thus suggesting a robustness in the "identified moment" sense of Nakamura & Steinsson (2018).

Results for Smets & Wouters (2007). We present results for a particular data-generating process: the structural model of Smets & Wouters. We pick this model because it is parameterized flexibly enough to provide a fit to aggregate time series that is competitive with reduced-form VARs; in particular, the output gap and inflation causal effect maps $\{\Theta_{y,\nu,\mathcal{A}}, \Theta_{\pi,\nu,\mathcal{A}}\}$ are affected by 17 distinct structural parameters—our vector ζ. We estimate the model in the usual way using aggregate time series data, and then use the posterior mean and variance-covariance matrix as ζ_0 and Σ_ζ, respectively.

Given the model, it remains to specify the counterfactuals c and the observables γ. Here we proceed as follows. First, for c, we begin by considering the entirety of the output and inflation causal effect maps $\{\Theta_{y,\nu,\mathcal{A}}, \Theta_{\pi,\nu,\mathcal{A}}\}$—i.e., our sufficient statistics for the *universe* of possible systematic rule change counterfactuals. We will later consider counterfactuals for particular shock paths. Second, for γ, we choose the impulse responses corresponding to the two interest rate paths that we used in our empirical applications (displayed in Figure 3). Recall that we collected the output and inflation impulse responses to these two shock paths in the matrices $\{\Omega_{y,\mathcal{A}}, \Omega_{\pi,\mathcal{A}}\}$. We then proceed as follows: for the output causal effect map $\Theta_{y,\nu,\mathcal{A}}$ as the counterfactual c, we select as our observables γ the short- and medium-run

(a) Informativeness for $\Theta_{y,\nu,\mathcal{A}}$

(b) Informativeness for $\Theta_{\pi,\nu,\mathcal{A}}$

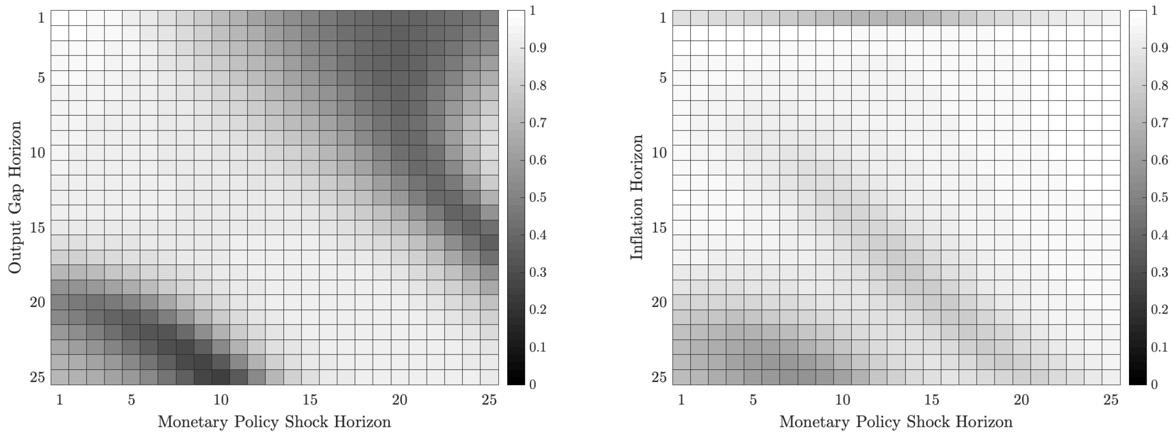

Figure C.3: Output gap and inflation informativeness for monetary policy shocks in the structural model of Smets & Wouters (2007), computed using (C.2) and for the observables γ defined in (C.3).

average responses of output to our two identified policy instrument paths, i.e.,

$$\gamma = \left(\frac{1}{4} \sum_{h=1}^{4} \Omega_{y,\mathcal{A}}(h, \bullet), \frac{1}{12} \sum_{h=5}^{16} \Omega_{y,\mathcal{A}}(h, \bullet) \right) \tag{C.3}$$

We proceed similarly for inflation. We thus in both cases ask the question: how much does knowledge of *only* the average short- and medium-run causal effects of the observed instrument paths onto the outcome of interest—i.e., four numbers—restrict the remainder of the (high-dimensional) policy shock causal effect maps? We here present the main results, and then discuss some further computational details at the end of this section.

The left panel of Figure C.3 shows the informativeness measure Δ_i of our four output gap impulse response moments for the rest of $\Theta_{y,\nu,\mathcal{A}}$, while the right panel does the same for inflation. The heatmaps reveal that informativeness is reasonably high throughout, with averages of 0.75 for the output gap and 0.91 for inflation. Informativeness is particularly high for short-term shocks—corresponding to our two identified instrument paths—and relatively short horizons—corresponding to the averaged impulse responses in our γ's—and decreases away from the main diagonal. Adding a third long-run average to our observables γ, the average Δ's increase to 0.94 and 0.95, respectively.[29]

While Figure C.3 depicts our measure of informativeness for the *entire* causal effect maps,

[29] To be precise, we set $\gamma = \left(\frac{1}{4} \sum_{h=1}^{4} \Omega_{y,\mathcal{A}}(h, \bullet), \frac{1}{8} \sum_{h=5}^{12} \Omega_{y,\mathcal{A}}(h, \bullet), \frac{1}{8} \sum_{h=13}^{20} \Omega_{y,\mathcal{A}}(h, \bullet) \right)$.

INFORMATIVENESS Δ FOR RULE (29) IN SMETS & WOUTERS

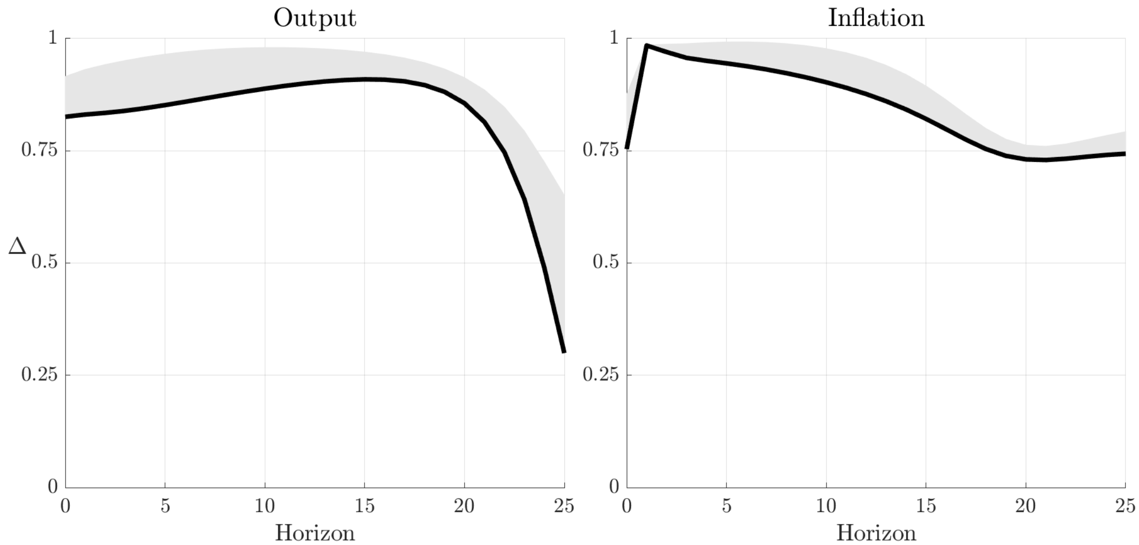

Figure C.4: Time paths of the output gap and inflation informativeness in the structural model of Smets & Wouters (2007) for the counterfactual rule (29), to be implemented following the model's estimated investment-specific technology shock, for γ defined as in (C.3) (solid line) and adding a third, long-run impulse response observable (shaded, see Footnote 29).

in practice counterfactuals for typical business-cycle fluctuations are likely to depend mainly on impulse responses to contemporaneous and a couple of short-run policy news shocks, as for example suggested by our illustrative analysis in Section 2.4 as well as our applications in Section 3.3. Given this observation, we would expect informativeness for such *particular* policy counterfactuals to be even higher than the averages across all of the shock causal effects reported above. We illustrate this conjecture by computing our informativeness measure for a particular shock—the investment-specific technology shock of Smets & Wouters—and a particular counterfactual rule—the rule (29) previously considered in Section 2.4. We pick this rule because it implies substantial nominal interest rate inertia and thus lies outside of the purview of the Phillips curve-based analysis from Section 4.1. Figure C.4 presents the results, plotting horizon-by-horizon informativeness for the desired counterfactual, in solid for our baseline γ (i.e., four observables) and shaded if we add long-run response for each shock. Exactly as expected, the informativeness measures are higher than the averages reported before at short horizons, before falling at longer horizons.

DISCUSSION. Closely building on Andrews et al. (2020), the analysis in this section has introduced a tool that allows researchers to communicate—given their maintained parametric structural model—which moments of the data "drive" their reported policy counterfactuals. Consistent with our identification results coupled with basic economic intuition, we find that impulse responses to identified policy shocks can be highly informative "identified moments" (Nakamura & Steinsson, 2018) for structural policy rule counterfactuals, providing a novel justification to impulse response matching as a way of estimating structural macroeconomic model (Rotemberg & Woodford, 1997; Christiano et al., 2005). Appendix C.3 provides some intuition by discussing the concept of asymptotic time invariance to link impulse responses to policy (news) shocks at different horizons.

Aside: further computational details. Our estimation of the structural model of Smets & Wouters (2007) uses replication codes kindly provided by Johannes Pfeifer.[30] The estimation yields the posterior mode ζ_0 and the variance-covariance matrix Σ_ζ.

We compute the monetary policy shock causal effect maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ by solving the model using sequence-space methods, and then sequentially adding all different contemporaneous and news shocks to the policy rule. We then compute $\frac{\partial \Theta_{x,\nu,\mathcal{A}}}{\partial \zeta}$ and $\frac{\partial \Theta_{z,\nu,\mathcal{A}}}{\partial \zeta}$ using finite-difference methods. Given that all counterfactuals c and observables γ are functions of $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$, we can use these derivative matrices to construct the joint variance-covariance matrix Σ for counterfactual $\hat{c}$ and observables $\hat{\gamma}$.

Our observables γ are chosen using the estimated policy instrument paths from our empirical analysis, plotted in Figure B.1 and Figure B.2. We take the point estimates of the interest rate path, and then at the estimated mode ζ_0 construct the sequence of monetary policy shocks $\boldsymbol{\nu}_{rr}$ and $\boldsymbol{\nu}_{gk}$ that would correspond to the two identified shocks. Our observables γ are then computed from the model-implied output and inflation impulse responses to those two shocks $\boldsymbol{\nu}_{rr}$ and $\boldsymbol{\nu}_{gk}$. The informativeness of the impulse responses to these particular shocks for impulse responses to all other possible shocks is reported in Figure C.3.

Finally, for the particular counterfactual studied in Figure C.4, we consider the investment-specific technology shock estimated by Smets & Wouters. We then, at the model's mode ζ_0, compute the particular monetary policy shock paths $\tilde{\boldsymbol{\nu}}$ that would map the investment-specific technology shock under the baseline rule to its counterfactual propagation under our alternative rule (29). Figure C.4 shows the informativeness of our selected observed impulse responses (i.e., entries of the causal effects of $\boldsymbol{\nu}_{rr}$ and $\boldsymbol{\nu}_{gk}$) for the responses to this $\tilde{\boldsymbol{\nu}}$.

[30]The code is available at `https://sites.google.com/site/pfeiferecon/dynare`.

C.3 Asymptotic time invariance

This section formalizes the intuition that impulse responses to policy shocks at different horizons are tightly linked, using the concept of asymptotic time invariance of impulse responses. This property provides a (partial) rationalization of the high degree of informativeness of individual impulse responses documented in Appendix C.2.

The precise definition of asymptotic time invariance is that, for all $s \in \mathbb{N}$,

$$\lim_{h\to\infty} \Theta_{x,\nu,\mathcal{A}}(h+s,h) = \bar{\Theta}_{x,\nu,\mathcal{A}}(s), \quad \lim_{h\to\infty} \Theta_{z,\nu,\mathcal{A}}(h+s,h) = \bar{\Theta}_{z,\nu,\mathcal{A}}(s) \tag{C.4}$$

where $\bar{\Theta}_{x,\nu,\mathcal{A}}$ and $\bar{\Theta}_{z,\nu,\mathcal{A}}$ are two sequences. Figure C.5 provides an illustration of this property in the model of Smets & Wouters, showing output impulse responses to various different contemporaneous and forward guidance monetary shocks. We see that, for forward guidance shocks far into the future (large shock horizon h), the output impulse responses are left- and right-translations of each other, exactly as expected. In light of this observation, it is not surprising that a small number of identified shock impulse responses are highly informative about the entirety of the policy shock causal effect maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$, as seen in Appendix C.2 for the model of Smets & Wouters.

Asymptotic Time Invariance of IRFs in Smets & Wouters

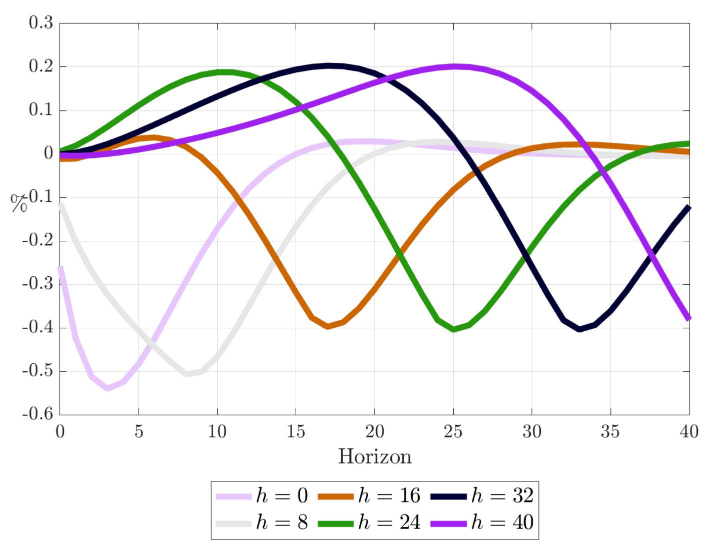

Figure C.5: Output impulse responses to contemporaneous and forward guidance monetary policy shocks in the model of Smets & Wouters (2007).