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

Alisdair McKay
FRB Minneapolis

Christian K. Wolf
MIT & NBER

This version: October 2021
First version: July 2021

**Abstract:** We show that, in a general family of linearized structural macroeconomic models, knowledge of the dynamic causal effects of contemporaneous and news shocks to the prevailing policy rule is sufficient to: a) construct counterfactuals under alternative policy rules; and b) recover the optimal policy rule corresponding to a given loss function. Under our assumptions, the derived counterfactuals and optimal policies are robust to the Lucas critique. We then discuss strategies for applying these insights when only a limited amount of 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, Gabriel Chodorow-Reich, Peter Ganong, Yuriy Gorodnichenko, Kilian Huber, Chen Lian, Geert Mesters, Ben Moll, Chris Moser, Emi Nakamura, Mikkel Plagborg-Møller, Jón Steinsson, Mike Waugh, Iván Werning, and seminar participants at various venues. 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 response to the Lucas (1976) critique of macroeconometric policy evaluation, two dominant methodological approaches emerged. In the Lucas (1980) program, the task of the researcher is to assess the consequences of a change in the systematic policy *rule*, with that change fully understood by the private sector. This assessment is conducted through fully-specified, parametric structural models with deep microfoundations. Empirical evidence on the transmission of policy shocks, say a surprise tightening of the monetary stance, here often plays the role of estimation target for the model (Rotemberg & Woodford, 1997; Christiano et al., 1999). Alternatively, in the Sims (1980, 1982, 1987) program, the researcher instead studies changes in the policy stance that are *not* perceived as corresponding to a change in the systematic rule. For this more modest objective, purely statistical estimates of the effects of policy shocks suffice: researchers can form policy counterfactuals through vector autoregressions or local projections, without needing to rely on a particular parametric structural model (Leeper & Zha, 2003; Sims & Zha, 2006).

In this paper, we propose a third, hybrid approach to forming policy counterfactuals. Like the Lucas program, our goal is the ambitious one of studying changes in policy rules. Rather than relying on a particular parametric structural model, however, our analysis instead begins with a general, linear data-generating process. We then impose one key restriction: that policy shapes private-sector behavior *only* through the current and future expected path of the policy instrument (say, the nominal rate). Importantly, once linearized, many of the structural models popular in the Lucas program—from simple real business cycle or New Keynesian models (e.g., Baxter & King, 1993; Woodford, 2011) to those with rich consumer and firm heterogeneity (e.g., Kaplan et al., 2018; Ottonello & Winberry, 2020)—fit into this environment. We then prove that, conditional on this structure, purely statistical estimates of the causal effects of contemporaneous *and news* policy shocks are sufficient to construct the desired policy counterfactuals. This identification result offers a bridge between the two existing approaches: it reveals that impulse responses for many policy *shocks* (as in the Sims program) are sufficient statistics for the effects of changes in systematic policy *rules* (as in the Lucas program). The cost of this generality are large informational requirements: our result requires empirical evidence not just on a single policy shock, but on many different ones. Our second contribution is to offer two ways of operationalizing our identification result in the empirically relevant case of limited evidence on policy shock transmission.

The first part of the paper establishes the identification result. We consider an econometrician living in an economy consistent with our structural assumptions. The economy is closed with some fixed, baseline policy rule, and the econometrician would like to a) predict the behavior of the economy under alternative rules and b) find the optimal rule corresponding to some externally set loss function.[1] We further assume that the prevailing policy rule is subject to random shocks. Using standard semi-structural time-series methods, the econometrician can estimate the dynamic causal effects of such policy shocks (see Ramey, 2016, for a survey). Our identification result states that, if the econometrician has successfully estimated the effects of *all* possible contemporaneous and news shocks to the policy rule, then she can in fact construct her desired counterfactuals: she can predict the effects of a change in rule, and she can recover optimal rules in the form of a forecast targeting criterion for a given loss function. 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 policy rule. For example, a prevailing dovish monetary policy rule can be mapped into an alternative hawkish rule by adding a suitable sequence of contractionary interest rate shocks. Importantly, and differently from the standard Sims program, our identification result requires estimates of the effects of policy *news* shocks in order to account for changes in private-sector expectations. It is only this full set of policy shock impulse responses that serve as valid sufficient statistics for the effects of a change in the systematic policy rule.

How general is the setting of our identification result? As already emphasized, in addition to linearity, our key restriction is that the policy rule affects private sector behavior only through the current and expected future path of the policy instrument. To illustrate this assumption, consider for example a simple linearized Euler equation,

$$c_t = -\sigma \left(i_t - \mathbb{E}_t \pi_{t+1}\right) + \mathbb{E}_t c_{t+1}, \tag{1}$$

where $c_t$ is consumption, $i_t$ is the nominal rate of interest, $\pi_t$ is inflation, and $\sigma$ is the intertemporal elasticity of substitution. Policy decisions affect households through the implied movements of the nominal rate $i_t$; conditional on the dynamics of $i_t$, any further properties of

[1] To be clear, our identification results are silent on the shape of the objective function. Explicit, 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.

the policy rule determining $i_t$—for example the extent to which the central bank leans against inflation—are irrelevant.[2] This sufficiency of the policy instrument is a property shared by many of the linearized models used in the Lucas program, from simple analytical business-cycle models to those with many frictions and shocks (e.g., Christiano et al., 2005; Smets & Wouters, 2007) and rich microeconomic heterogeneity.[3] Perhaps the most popular class of models violating our restriction is those featuring an asymmetry of information between the policymaker and the private sector, as in Lucas (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 on models, our structural assumptions also limit the set of policy counterfactuals to which our method can be applied. Our approach can be used compare different cyclical stabilization policies (e.g., monetary or fiscal policy rules for business-cycle policy), but due to linearity it is less well-suited to study policies that alter the economy's steady state (e.g., changes in the inflation target or in the long-run fiscal system).

The main challenge to implementing our approach is that existing empirical evidence on policy shocks is limited. Recall that our identification result requires the econometrician to estimate the causal effects of the full menu of possible contemporaneous and news shocks to the prevailing policy rule. For example, in the context of monetary policy, she would need to know the effects of shocks to interest rates at every single point along the yield curve. Such fine-grained, maturity-by-maturity evidence is not available. In the second part of the paper we present two alternative strategies for dealing with this lack of data.

Our first strategy is to focus on a narrower set of counterfactuals—those for which we have sufficient evidence. Suppose that the econometrician can estimate the causal effects of some (small) set of policy shocks. Then, by our identification result, we can construct counterfactuals for all alternative policy rules that deviate from the prevailing rule in a way consistent with the empirically identified shocks. The more shocks we observe, the richer the deviations from the prevailing rule that we can entertain, and so the richer the set of nested counterfactuals. By the same token, we can find the optimal policy rule within this

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

[3]Note that all of these models still feature cross-equation restrictions in the style of Hansen & Sargent (1980). The decision rules estimated by Hansen & Sargent relate outcomes at date $t$ to data available at $t$, and so are generally shaped by the policy rule. We instead allow the entire expected future path of the policy instrument to appear in decision rules. This way of writing the equilibrium relations gives us the separation of policy and non-policy blocks at the heart of our results.

spanned set of policy shock causal effects. The more evidence is available, the closer is this restricted optimal rule to the full optimal rule.

We provide a practical illustration of these insights with an application to monetary policy counterfactuals. Our starting point is the causal effect of a contractionary investment-specific technology shock under the actually observed monetary policy reaction function, estimated using the shock series of Ben Zeev & Khan (2015). This shock resembles a classic supply shock, with inflation rising, output falling, and the central bank leaning against the increase in inflation. We would then like to learn about the counterfactual propagation of this shock under a) an alternative policy rule that aggressively stabilizes output and b) the optimal policy rule corresponding to a "dual mandate"-type loss function with policymaker preferences over average inflation, consistent with the recent review of the Federal Reserve's policy framework.[4] To construct the two desired counterfactuals, we follow Christiano et al. (1999) and Gertler & Karadi (2015) to learn about the dynamic causal effects of persistent and short-lived changes in the federal funds rate, respectively. We then leverage our theoretical results to explore the counterfactuals a) and b) in this identified subspace.

Our second strategy is to impose additional structure in order to extrapolate from the causal effects of the policy shocks that we did observe to those that we did not. Mathematically, we face a matrix completion problem: we require the full set of policy causal effects (an infinite-dimensional linear map), yet only have evidence on some specific shocks (i.e., some weighted averages of columns of the map). Our solution is to parameterize the causal effect maps using theory-guided structural assumptions, and then estimate the parameters of these maps from the policy shocks that we do observe. This procedure is quite similar in spirit to estimation via impulse-response matching (as in Christiano et al., 2005), but with one crucial difference—we show that, for the purposes of completing our impulse response maps, it may well suffice to specify partial model *blocks* rather than entire general equilibrium models. For example, for our monetary policy application, it suffices to assume that output and inflation are linked through a dynamic Phillips curve relationship. This restriction narrows the class of models to which our results apply, but still leaves much structure unspecified, including in particular the entire demand block of the economy.

Returning to the application, we use our restrictions to estimate the implementable space of inflation-output pairs, and from there construct counterfactuals for a) an alternative rule

[4]We use a flexible average inflation targeting loss function, similar to the one used by Svensson (2020). Our loss function is quadratic in deviations of output from trend and deviations of a 5-year average of inflation from target, with equal weights on the two components.

that completely stabilizes output and b) the optimal average inflation targeting policy rule. As expected, the results are similar to the identified subspace analysis, if somewhat smoother and more accurately estimated, due to the additional structure that we impose.

LITERATURE. Our emphasis on policy shock impulse responses as sufficient statistics connects with previous approaches to policy counterfactual analysis. Relative to standard policy shock impulse-response matching in the Lucas program (e.g. Christiano et al., 2005), we emphasize that evidence on *many* policy shocks obviates the need to specify a particular parametric structural model. Relative to the Sims program, we show that policy *news* shocks provide the missing link to form systematic policy rule counterfactuals without running afoul of the Lucas critique. Within the Sims program, counterfactuals are usually constructed using a sequence of *unanticipated* policy shocks that enforce an alternative policy rule along the equilibrium path (Sims & Zha, 2006; Bernanke et al., 1997; Eberly et al., 2019; Antolin-Diaz et al., 2021). This approach will be credible if the private sector is unlikely to detect the change in regime (Leeper & Zha, 2003), but it is less appropriate to analyze the effects of announced changes in regime such as the Federal Reserve's recent change in the systematic policy framework (Powell, 2020).[5]

Our work also relates to more recent contributions to policy counterfactual analysis. Beraja (2020) similarly forms counterfactuals without relying on particular parametric structural 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 shock propagation. Barnichon & Mesters (2021) use policy shock impulse responses to *test* the optimality of a given policy rule. We show that under mild structural assumptions such policy shock impulse responses can in fact be used to a) form valid counterfactuals for changes in rules and b) fully characterize optimal policy rules.

Finally, we build on recent advances in solution methods for dynamic general equilibrium models. At the heart of our analysis lies the fact that equilibria in such models can be characterized by matrices of impulse response functions (see 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 the same arguments as in our identification re-

[5] Kocherlakota (2019) similarly considers the problem of a policymaker trying to make the optimal decision *conditional on a given policy regime*. His game-theoretic analysis considers a private sector that does not change its strategy, thus providing a theoretical rationalization of the Sims program.

sults 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, thus forcing us to confront the fact that such evidence is limited.

OUTLINE. The remainder of the paper proceeds as follows. Section 2 presents our main identification results. Sections 3 and 4 then discuss two approaches to dealing with realistic data limitations, and apply our results to construct monetary policy counterfactuals for investment-specific technology shocks. Section 5 concludes.

# 2 Dynamic causal effects & policy counterfactuals

This section contains our core identification results. We begin by presenting a simple static version of our argument in the context of a standard small-scale New Keynesian model, and then extend the argument to a general class of linearized dynamic models.

Throughout the paper, we formulate our analysis in linearized perfect-foresight economies. Due to certainty equivalence, the equilibrium dynamics of a linear model with uncertainty will coincide with the solution to such a linearized perfect-foresight economy. We thus emphasize that all results presented below extend without any change to models with aggregate risk solved using conventional first-order perturbation techniques.[6]

## 2.1 A simple example

We begin with an illustration of our identification argument in the context of a particular structural model—the linearized three-equation New Keynesian model (see Galí, 2015).

For $t = 0, 1, 2, \ldots$, the perfect-foresight model is described by the following equations:

$$y_t = -\sigma(i_t - \pi_{t+1}) + y_{t+1} \tag{2}$$

$$\pi_t = \kappa y_t + \beta \pi_{t+1} + \varepsilon_t^s \tag{3}$$

$$i_t = \phi_\pi \pi_t + \nu_t, \tag{4}$$

where $y_t$ is output, $\pi_t$ is inflation, $i_t$ is the nominal interest rate, $\varepsilon_t^s$ is a cost-push shock,

[6] 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 $\nu_t$ is a policy shock. The first two equations describe the behavior of the private sector, while the last equation is the monetary policy rule. Underlying this linear model is a set of structural assumptions that micro-found the parameters of the linearized economy. For our purposes, the crucial property of these micro-foundations is that the coefficients in the private-sector equations are independent of the policy rule—$\sigma$, $\kappa$ and $\beta$ as well as the cost-push shock process are all unaffected by changes in the policy rule (i.e., $\phi_\pi$). Equivalently, private-sector behavior—that is, the two relations (2) and (3)—is affected by policy only through the current and expected future path of the policy instrument $i_t$.

To simplify the analysis as much as possible, we assume that shocks are perfectly transitory, so the system (2) - (4) becomes static, with $y_t = \pi_t = i_t = 0$ for $t \geq 1$.[7]

Objects of interest. We wish to characterize the behavior of the economy in response to the cost-push shock $\varepsilon_t^s$ under policy rules different from (4). For this example we will focus on the following two counterfactuals. First, we would like to know the behavior of $\{y_t, \pi_t, i_t\}$ in response to the cost-push shock under the alternative policy rule

$$i_t = \tilde{\phi}_\pi \pi_t \tag{5}$$

where $\tilde{\phi}_\pi \neq \phi_\pi$. Second, for a policymaker with a known loss function of the form

$$\lambda_\pi \pi_t^2 + \lambda_y y_t^2, \tag{6}$$

we would like to recover the optimal policy rule and characterize optimal output, inflation and interest rate responses to the cost-push shock $\varepsilon_t^s$. In particular, we would like to recover these counterfactuals for a private sector understanding the change in rule—that is, we would like to answer the ambitious question of the Lucas program.

Empirical evidence. Consider an econometrician who observes data generated from the model (2) - (4) under the baseline monetary policy rule (4). Using conventional semi-structural methods (Ramey, 2016), and with enough data, she can perfectly recover the impulse response matrices $\Theta_{x,\varepsilon,\phi_\pi}$ and $\Theta_{x,\nu,\phi_\pi}$—i.e., the impulse responses of $x = \{y, \pi, i\}$ to

[7] We furthermore as usual assume that the Taylor principle holds for all policy rules considered here, so the system has a unique bounded solution with the claimed properties.

shocks $\varepsilon$ and $\nu$ under the baseline rule $\phi_\pi$. They are given as

$$\begin{pmatrix} y_t \\ \pi_t \\ i_t \end{pmatrix} = \underbrace{\begin{pmatrix} -\frac{\sigma\phi_\pi}{1+\kappa\sigma\phi_\pi} \\ \frac{1}{1+\kappa\sigma\phi_\pi} \\ \frac{\phi_\pi}{1+\kappa\sigma\phi_\pi} \end{pmatrix}}_{\equiv\Theta_{x,\varepsilon,\phi_\pi}} \times \varepsilon_t^s + \underbrace{\begin{pmatrix} -\frac{\sigma}{1+\kappa\sigma\phi_\pi} \\ -\frac{\kappa\sigma}{1+\kappa\sigma\phi_\pi} \\ \frac{1}{1+\kappa\sigma\phi_\pi} \end{pmatrix}}_{\equiv\Theta_{x,\nu,\phi_\pi}} \times \nu_t. \tag{7}$$

Our main result is that knowledge of $\{\Theta_{x,\varepsilon,\phi_\pi}, \Theta_{x,\nu,\phi_\pi}\}$ is in fact sufficient to construct the two desired counterfactuals. That is, knowledge of the causal effects of the cost-push and policy shocks, $\varepsilon_t^s$ and $\nu_t$, under some baseline policy rule is actually enough to construct counterfactual impulse responses to $\varepsilon_t^s$ under either the alternative rule (5) or optimal policy for the loss function (6).

a) *Alternative policy rules.* We begin with our first counterfactual. To construct the counterfactual, we are going to design a monetary shock $\nu_t$ that maps the baseline rule (4) into the alternative rule (5). By definition of $\Theta_{x,\varepsilon,\phi_\pi}$ and $\Theta_{x,\nu,\phi_\pi}$, such a shock—together with the equilibrium aggregates $\{\tilde{y}_t, \tilde{\pi}_t, \tilde{i}_t\}$ that it implies—must satisfy the following system:

$$\tilde{y}_t = \Theta_{y,\varepsilon,\phi_\pi} \times \varepsilon_t^s + \Theta_{y,\nu,\phi_\pi} \times \nu_t \tag{8}$$

$$\tilde{\pi}_t = \Theta_{\pi,\varepsilon,\phi_\pi} \times \varepsilon_t^s + \Theta_{\pi,\nu,\phi_\pi} \times \nu_t \tag{9}$$

$$\tilde{i}_t = \Theta_{i,\varepsilon,\phi_\pi} \times \varepsilon_t^s + \Theta_{i,\nu,\phi_\pi} \times \nu_t \tag{10}$$

$$\tilde{i}_t = \tilde{\phi}_\pi \tilde{\pi}_t \tag{11}$$

(8) - (10) are the impulse responses to the shock tuple $\{\varepsilon_t^s, \nu_t\}$ under the baseline rule, and (11) states that the new policy rule (5) holds. In words, we set the shock $\nu_t$ to enforce the new policy rule, imposing that the mapping from shocks to equilibrium aggregates is consistent with the impulse response coefficients under the old rule in $\Theta_{x,\varepsilon,\phi_\pi}$ and $\Theta_{x,\nu,\phi_\pi}$.

The key result for our purposes is that the solution to our system (8) - (11) is exactly the same as the solution one would obtain by solving the model's structural equations (2) - (3) together with the new rule (5). This claim is easily verified. First, the structural solution is $\tilde{\pi}_t = (1+\kappa\sigma\tilde{\phi}_\pi)^{-1}\varepsilon_t^s$, which we obtain by replacing $\phi_\pi$ with $\tilde{\phi}_\pi$ in (7). Alternatively, solving (8) - (11) for $\nu_t$ gives

$$\nu_t = -\frac{\left(\tilde{\phi}_\pi - \phi_\pi\right)\Theta_{\pi,\varepsilon,\phi_\pi}}{\left(\tilde{\phi}_\pi - \phi_\pi\right)\Theta_{\pi,\nu,\phi_\pi} - 1} \times \varepsilon_t^s$$

and substituting into (9) and rearranging yields

$$\begin{aligned}\tilde{\pi}_t &= -\frac{\Theta_{\pi,\varepsilon,\phi_\pi}}{\left(\tilde{\phi}_\pi - \phi_\pi\right)\Theta_{\pi,\nu,\phi_\pi} - 1} \times \varepsilon_t^s = \frac{(1+\kappa\sigma\phi_\pi)^{-1}}{\left(\tilde{\phi}_\pi - \phi_\pi\right)(1+\kappa\sigma\phi_\pi)^{-1}\kappa\sigma + 1} \times \varepsilon_t^s \\ &= \frac{1}{1+\kappa\sigma\tilde{\phi}_\pi} \times \varepsilon_t^s,\end{aligned}$$

exactly as claimed, and similarly for output $\tilde{y}_t$. The intuition is simple. By construction, the policy shock $\nu_t$ is selected so that, in equilibrium, the new policy rule (5) holds. Private sector behavior, however, depends on the policy rule *only* to the extent that it affects the value of the policy instrument $i_t$. With $i_t$ set exactly as in the equilibrium under the new policy rule, it is immediate that all other equilibrium aggregates will also take the same values as in that counterfactual equilibrium.

b) *Optimal policy.* Next we consider optimal policy for a policymaker with preferences described by (6). The conventional, fully structural approach of treating the behavioral relations (2) - (3) as constraints yields the optimal implicit policy rule

$$\pi_t + \frac{\lambda_y}{\kappa\lambda_\pi} y_t = 0. \tag{12}$$

Equation (12) together with the Phillips curve (3) pins down optimal inflation-output pairs. To derive the same rule and optimal outcomes using the measured causal effects, consider the alternative problem of choosing the best *deviation* $\nu_t$ from the baseline rule to minimize the policymaker loss. The solution to this problem is characterized by the first-order condition

$$\lambda_\pi \pi_t \Theta_{\pi,\nu,\phi_\pi} + \lambda_y y_t \Theta_{y,\nu,\phi_\pi} = 0 \tag{13}$$

Substituting the definitions of $\Theta_{\pi,\nu,\phi_\pi}$ and $\Theta_{y,\nu,\phi_\pi}$, we find that (13) reduces to (12), as claimed. Finally, combining this rule with the shock impulse responses $\Theta_{\pi,\varepsilon,\phi_\pi}$ and $\Theta_{y,\varepsilon,\phi_\pi}$, it is straightforward to verify that we obtain the same optimal output-inflation pairs as those computed via the standard optimal policy problem.

As above, the key insight is that—because only the level of the policy instrument matters—we can equivalently think of counterfactual policy rules as shocks to the baseline rule. By observing these shocks, the econometrician recovers the implementable space of targets (here $y$-$\pi$ pairs). Given a loss function, knowledge of this implementable space is enough to recover the optimal policy rule and response to $\varepsilon_t^s$.

Summary & outlook. We have seen that, in the particular structural environment (2) - (4), estimates of the causal effects of policy *shocks* are sufficient to predict the behavior of the economy under a change in policy *rule*. The next subsections investigate the generality of this result. We will see that some assumptions can be relaxed—notably, neither the restriction of a one-period economy nor the particular structure of the three-equation model are at all necessary—while others have to be maintained—both linearity and our restrictions on policy rule feedback to private-sector expectations are central.

The analysis will proceed exactly in parallel to the simple example of this section. The key addition relative to the intuition from the simple model will be that, in a general dynamic environment, our identification arguments will require information on impulse response *paths* for both contemporaneous policy shocks as well as policy *news* shocks.

## 2.2 Model & objects of interest

We consider a linearized perfect foresight 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 model economy is summarized by the equilibrium system

$$\mathcal{H}_w \boldsymbol{w} + \mathcal{H}_x \boldsymbol{x} + \mathcal{H}_z \boldsymbol{z} + \mathcal{H}_\varepsilon \boldsymbol{\varepsilon} = \boldsymbol{0} \tag{14}$$

$$\mathcal{A}_x \boldsymbol{x} + \mathcal{A}_z \boldsymbol{z} + \boldsymbol{\nu} = \boldsymbol{0} \tag{15}$$

$w$ and $x$ are $n_w$- and $n_x$-dimensional vectors of endogenous variables, $z$ is a $n_z$-dimensional vector of policy instruments, $\varepsilon$ is a $n_\varepsilon$-dimensional vector of exogenous structural shocks, and $\nu$ is an $n_z$-dimensional vector of policy shocks. The distinction between $w$ and $x$ is that all variables in $x$ are observable to the policymaker and econometrician alike, while the variables in $w$ are not. 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 $t$.[8] Our key assumption is that the maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ do not depend in any way 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 (15) giving $n_z$ restrictions on the policy instruments for each $t$. Note in particular that the policy shock *sequence* $\boldsymbol{\nu}$ contains the full menu of possible contemporaneous ($t = 0$) and

[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.

news ($t > 0$) shocks to the policy rule (15).

Given $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}\}$, an equilibrium is a set $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}\}$ that solves (14) - (15). We assume that the baseline policy 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* (15) *induces a unique and determinate 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 $\{\boldsymbol{w}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \boldsymbol{\nu}), \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \boldsymbol{\nu}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \boldsymbol{\nu})\}$. Most interest will center 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 $\{\boldsymbol{w}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$.

Discussion & scope. Our results in the remainder of this paper will apply to *any* structural model that can be written in the general form (14) - (15). As emphasized before, in addition to linearity, the key property of the model for our purposes is that policy matters for the non-policy block *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 Lucas program 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 consistent with our assumptions. The three-equation model of Section 2.1 is an obvious case: Euler equation (2) and Phillips curve (3) are the policy-invariant private-sector block (14), and (4) is the policy rule (15). Appendix B.1 shows the specific linear maps that translate the model into the form of (14)-(15). For a slightly richer example, consider the heterogeneous-agent New Keynesian (HANK) model of Wolf (2021). That model consists of New Keynesian Phillips Curve (NKPC),

$$\boldsymbol{\pi} = \kappa \boldsymbol{y} + \beta \boldsymbol{\pi}_{+1} + \boldsymbol{\varepsilon}^s, \tag{16}$$

a consumer demand block (or IS relation),

$$\boldsymbol{y} = \mathcal{C}_y \boldsymbol{y} + \mathcal{C}_\pi \boldsymbol{\pi} + \mathcal{C}_i \boldsymbol{i} + \mathcal{C}_\tau \boldsymbol{\tau} + \boldsymbol{\varepsilon}^d, \tag{17}$$

a government budget constraint,

$$\mathbf{0} = \bar{\tau}\boldsymbol{\tau} + \bar{b}(1+\bar{i})(\boldsymbol{i}_{-1} - \boldsymbol{\pi}) \tag{18}$$

and a monetary policy rule,

$$\boldsymbol{i} = \phi_i \boldsymbol{i}_{-1} + (1-\phi_i)(\phi_\pi \boldsymbol{\pi} + \phi_y \boldsymbol{y}) + \boldsymbol{\nu}, \tag{19}$$

where $\pi$ is inflation, $y$ is output, $i$ is the nominal rate of interest, $\tau$ denotes transfers (which by (18) adjust to balance the government budget), $(\varepsilon^s, \varepsilon^d)$ are supply and demand shocks, and $\nu$ is the monetary policy shock.[9] The NKPC is as in the three-equation model, while the coefficient matrices in (17) are derived from aggregating the partial equilibrium household consumption decisions and thus again do not depend on policy rules. This HANK model fits into our structure with $w = \tau$, $x = (\pi, y)$, $z = i$ and $\varepsilon = (\varepsilon^s, \varepsilon^d)$, (16) - (18) as the block (14), and (19) as the policy rule (15).[10] With only slightly more elaborate versions of the same line of reasoning, it is straightforward to see that, once linearized, even workhorse estimated business-cycle models—such as Christiano et al. (2005) or Smets & Wouters (2007)—as well as recent quantitative HANK models—such as Auclert et al. (2020) or McKay & Wieland (2021)—fit into our structure. Finally, as we discuss in Appendix B.2, 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 results will generally not yield *globally* valid policy counterfactuals. Second, the policy invariance assumption embedded in the equilibrium system (14) - (15) 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 HANK model of Wolf (2021) sketched above,

[9] The subscripts $+1$ and $-1$ denote time paths shifted forward or backward one period, respectively.

[10] The only actual policy choice here is the nominal rate $i$. Lump-sum taxes—which passively adjust to balance the budget—are thus part of the policy-invariant block (14).

changes in the long-run tax-and-transfer system will affect the coefficient matrices in (17), so such policies are necessarily outside the purview of our analysis.[11] Third, some models—even after linearization—do not feature a separation of policy and non-policy blocks as in (14) - (15). 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 B.3, 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. Here, the response coefficient $\theta$ depends on the the policy rule for nominal demand growth as it affects the interpretation of changes in the island-level price, thus breaking our separation between the two model blocks.

In the remainder of this paper we will throughout impose the structural assumptions embedded in our framework (14) - (15).

Objects of interest. We want to learn about two sets of policy rule counterfactuals.

a) *Arbitrary alternative rules.* Consider an alternative policy rule

$$\tilde{\mathcal{A}}_x \boldsymbol{x} + \tilde{\mathcal{A}}_z \boldsymbol{z} = \boldsymbol{0} \tag{20}$$

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

**Assumption 2.** *The policy rule in* (20) *induces a unique and determinate equilibrium. That is, the infinite-dimensional linear map*

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

*is invertible.*

[11] Formally, the coefficient matrices in (17) are derivative matrices for an aggregate consumption function evaluated at the model's steady state. Changes in the tax-and-transfer function change the steady state and so also the coefficient matrices.

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 the exogenous shock path $\boldsymbol{\varepsilon}$?

b) *Optimal policy.* Consider a policymaker with a quadratic loss function of the form

$$\mathcal{L} = \sum_{i=1}^{n_x} \lambda_i \boldsymbol{x}_i' W \boldsymbol{x}_i \tag{21}$$

where $i$ indexes the $n_x$ distinct (observable) macroeconomic aggregates collected in $x$, $\lambda_i$ denotes policy weights, and $W$ is a symmetric positive-definite matrix.[12] 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* (21) *subject to the non-policy constraint* (14) *has a unique solution.*

We are 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 $\boldsymbol{z}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$?

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 all of the required information can in principle be recovered from data generated under the baseline policy rule.

## 2.3 Identification: impulse responses as sufficient statistics

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 (14) - (15) as

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

[12]Our focus on a separable quadratic loss functions is in line with standard optimal policy analysis, but not essential. As shown in Appendix A.1, 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 (15).

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 (15) with parameters $\mathcal{A}$. We will partition it as

$$\Theta_{\mathcal{A}} \equiv \begin{pmatrix} \Theta_{w,\varepsilon,\mathcal{A}} & \Theta_{w,\nu,\mathcal{A}} \\ \Theta_{x,\varepsilon,\mathcal{A}} & \Theta_{x,\nu,\mathcal{A}} \\ \Theta_{z,\varepsilon,\mathcal{A}} & \Theta_{z,\nu,\mathcal{A}} \end{pmatrix}. \tag{22}$$

All of our identification results will require knowledge of $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$—the *full* sets of dynamic causal effects of the policy shocks $\boldsymbol{\nu}$. That is, the econometrician needs to know the effects of every possible current and future (announced) deviation from the prevailing policy rule onto the policy instruments $z$ as well as the (observable) endogenous variables $x$ (i.e., all of the arguments of the policy function and the policymaker loss). Furthermore, to construct counterfactual paths that correspond to a given shock sequence $\boldsymbol{\varepsilon}$, the researcher also needs to know the dynamic causal effects of that shock sequence under the baseline policy rule, $\{\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$.

These informational requirements are the natural dynamic generalization of those for the simple model in Section 2.1. First, since the model is now dynamic, a given policy shock now generates entire *paths* of impulse responses, corresponding to the rows of the $\Theta$'s. Second, rather than a single shock, we now need to know causal effects corresponding to the full menu of possible contemporaneous and news shocks $\boldsymbol{\nu}$—the columns of the $\Theta$'s.

a) *Alternative Policy Rules.* We begin with the first object of interest—policy counterfactuals after a shock sequence $\boldsymbol{\varepsilon}$ under an alternative policy rule.

**Proposition 1.** *Suppose that $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ and $\{\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$ are known. Then, for any alternative policy rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$ that induces a unique, determinate equilibrium, we can recover the policy counterfactuals $\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ as the unique solution to the system*

$$\begin{pmatrix} I & \mathbf{0} & -\Theta_{x,\nu,\mathcal{A}} \\ \mathbf{0} & I & -\Theta_{z,\nu,\mathcal{A}} \\ \tilde{\mathcal{A}}_x & \tilde{\mathcal{A}}_z & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) \\ \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) \\ \mathbf{0} \end{pmatrix}. \tag{23}$$

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

$$\begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \tilde{\mathcal{A}}_x & \tilde{\mathcal{A}}_z \end{pmatrix} \begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = \begin{pmatrix} -\mathcal{H}_\varepsilon \\ \mathbf{0} \end{pmatrix} \boldsymbol{\varepsilon} \tag{24}$$

By Assumption 2 we know that (24) 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, consider instead the alternative system (23). Since (14) also holds under the initial policy rule, and since the last line of (23) imposes the new policy rule, it follows that any $(\boldsymbol{x}, \boldsymbol{z})$ that are part of a solution of (23) are also part of a solution of (24). Since by assumption (24) has a unique solution, it follows that the system (23) is solved by at most one set of paths $(\boldsymbol{x}, \boldsymbol{z})$.

It remains to establish that the system (23) has a solution. For this consider the candidate tuple $\{\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}), \boldsymbol{\nu} = (\tilde{\mathcal{A}}_x - \mathcal{A}_x)\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) + (\tilde{\mathcal{A}}_z - \mathcal{A}_z)\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})\}$. 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 (24), it follows that they are also a solution to the system

$$\begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \mathcal{A}_x & \mathcal{A}_z \end{pmatrix} \begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = - \begin{pmatrix} \mathcal{H}_\varepsilon \boldsymbol{\varepsilon} \\ (\tilde{\mathcal{A}}_x - \mathcal{A}_x)\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) + (\tilde{\mathcal{A}}_z - \mathcal{A}_z)\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) \end{pmatrix} \tag{25}$$

But by Assumption 1 we know that the system (25) 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 (22) that the candidate tuple also solves (23), completing the argument. □

(23) is the dynamic generalization of the simple static system (8) - (11). 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 perturbation that mimics the equilibrium instrument path under the new instrument rule. But since the first model block (14) 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 the full system is dynamic, we need contemporaneous *and* news shocks to the baseline rule in order to be able to mimic an arbitrary alternative rule.

b) *Optimal policy.* The second identification result concerns optimal policy.

**Proposition 2.** *Consider a policymaker loss function* (21)*, and suppose that* $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ *are known. Then we can recover the optimal policy rule* $\{\mathcal{A}_x^*, \mathcal{A}_z^*\}$ *as*

$$\begin{aligned} \mathcal{A}_x^* &= \left(\lambda_1 \Theta'_{x_1,\nu,\mathcal{A}} W, \lambda_2 \Theta'_{x_2,\nu,\mathcal{A}} W, \ldots, \lambda_{n_x} \Theta'_{x_{n_x},\nu,\mathcal{A}} W\right), && (26) \\ \mathcal{A}_z^* &= \mathbf{0}. && (27) \end{aligned}$$

*If* $\{\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$ *are also known, then we can furthermore recover counterfactuals for the shock path* $\boldsymbol{\varepsilon}$ *under the optimal policy rule,* $\boldsymbol{x}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$ *and* $\boldsymbol{z}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$*, through Proposition 1.*

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

$$\begin{aligned} \mathcal{H}_w'(I \otimes W)\boldsymbol{\varphi} &= \mathbf{0} & (28) \\ (\Lambda \otimes W)\boldsymbol{x} + \mathcal{H}_x'(I \otimes W)\boldsymbol{\varphi} &= \mathbf{0} & (29) \\ \mathcal{H}_z' W \boldsymbol{\varphi} &= \mathbf{0} & (30) \end{aligned}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots)$ and $\varphi$ is the multiplier on (14). By Assumption 3 we know that the system (28) - (30) together with (14) 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 (21) subject to (14) - (15). This second problem gives the first-order conditions

$$\begin{aligned} \mathcal{H}_w'(I \otimes W)\boldsymbol{\varphi} &= \mathbf{0} & (31) \\ (\Lambda \otimes W)\boldsymbol{x} + \mathcal{H}_x'(I \otimes W)\boldsymbol{\varphi} + \mathcal{A}_x' W \boldsymbol{\varphi}_z &= \mathbf{0} & (32) \\ \mathcal{H}_z'(I \otimes W)\boldsymbol{\varphi} + \mathcal{A}_z' W \boldsymbol{\varphi}_z &= \mathbf{0} & (33) \\ W \boldsymbol{\varphi}_z &= \mathbf{0} & (34) \end{aligned}$$

where $\varphi_z$ is the multiplier on (15). It follows from (34) that $\boldsymbol{\varphi}_z = \mathbf{0}$. But then (31) - (33) together with (14) determine the same unique solution as before, and $\boldsymbol{\nu}$ adjusts residually to satisfy (15). 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

$$\begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = \Theta_{\mathcal{A}} \times \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\nu} \end{pmatrix} \tag{35}$$

The problem of minimizing (21) subject to (35) gives the optimality condition

$$\sum_{i=1}^{n_x} \lambda_i \Theta_{x_i,\nu,\mathcal{A}}' W \boldsymbol{x}_i = 0 \tag{36}$$

By the equivalence of the policy problems, it follows that (36) is an optimal policy *rule*, taking the form (26) - (27). Finally, the second part of the result follows from Proposition 1 since (36) is just a special example of a policy rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$.

$\square$

The optimal policy rule in (26)-(27) is the dynamic analogue of the static rule in (13).[13] The intuition is as before: since we know the (now dynamic) causal effects of every possible policy perturbation $\boldsymbol{\nu}$ on the policymaker targets $\boldsymbol{x}$, we in fact 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.

The identification results in Propositions 1 and 2 offer a bridge between the Lucas and Sims programs: they show that, under our structural assumptions, impulse responses to contemporaneous and news policy *shocks*—objects estimable using the techniques of the Sims program—are sufficient statistics for the Lucas program objective of predicting the effects of changes in systematic policy *rules*.

Aside: relative impulse responses. Our statements of Propositions 1 and 2 rely on the absolute impulse responses $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$. Both results, however, in fact only really require information on relative dynamic causal effects: if, for example, the first impulse response map $\Theta_{x_1,\nu,\mathcal{A}}$ is invertible, then the proofs of both results apply without any change using the weaker informational requirement $\{\tilde{\Theta}_{x,\nu,\mathcal{A}}, \tilde{\Theta}_{z,\nu,\mathcal{A}}\}$, where $\tilde{\Theta}_{x_i,\nu,\mathcal{A}} \equiv \Theta_{x_i,\nu,\mathcal{A}} \times \Theta^{-1}_{x_1,\nu,\mathcal{A}}$ and $\tilde{\Theta}_{z_i,\nu,\mathcal{A}} \equiv \Theta_{z_i,\nu,\mathcal{A}} \times \Theta^{-1}_{x_1,\nu,\mathcal{A}}$. Intuitively, both for counterfactual rules of the form (20) as well as for optimal policy, the only information required by the econometrician are the relative (or normalized) implementable spaces of policy targets and instruments $x$ and $z$. Our connection of theory and measurement in Section 4.2 will heavily leverage this observation.

## 2.4 Quantitative illustration

We complement our theoretical discussion of the sufficiency of impulse response functions with a numerical illustration in the context of a quantitative HANK model. The purpose of

[13] 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):

$$\sum_{i=1}^{n_x} \lambda_i \Theta'_{x_i,\nu,\mathcal{A}} W \mathbb{E}_t \left[\boldsymbol{x}_i\right] = 0 \tag{37}$$

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, (37) must apply to *revisions* of policymaker expectations at each $t$.

this section is to provide a visual representation of our results in the context of a model that is—unlike the simple case of Section 2.1—neither static nor solvable in closed-form.

We use the HANK model of Wolf (2021), sketched in Section 2.2 and with details of the model parameterization relegated to Appendix B.4. We first of all solve the model with a baseline policy rule of

$$i_t = \phi_\pi \pi_t \tag{38}$$

for $\phi_\pi = 1.5$. Using this solution we compute the policy causal effects $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ and the impulse responses $\{\boldsymbol{x}_\mathcal{A}(\boldsymbol{\varepsilon}), \boldsymbol{z}_\mathcal{A}(\boldsymbol{\varepsilon})\}$ to a contractionary cost-push shock $\varepsilon^s$. Then, following Propositions 1 and 2, we use those impulse responses to construct policy counterfactuals.

a) *Alternative policy rules.* For our first experiment, we would like to learn about the behavior of output and inflation under an alternative policy rule

$$i_t = \phi_i i_{t-1} + (1 - \phi_i)(\phi_\pi \pi_t + \phi_y y_t) \tag{39}$$

for $\phi_i = 0.9$, $\phi_\pi = 2$ and $\phi_y = 0.5$. We will perform this calculation in two ways.

First, we make use of the structural equations of the model: we simply replace the baseline policy rule with the alternative rule and then re-solve the model. The cost-push shock impulse responses under the baseline rule and the counterfactual new rule are displayed as the grey and orange lines in Figure 1.

Next, we use Proposition 1 to equivalently construct the desired counterfactual without knowledge of the structural equations of the model. We do so using $\{\boldsymbol{x}_\mathcal{A}(\boldsymbol{\varepsilon}), \boldsymbol{z}_\mathcal{A}(\boldsymbol{\varepsilon})\}$ and $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$—the dynamic causal effects of the fundamental shock and of policy shocks generated under the prevailing baseline rule (38). We feed these inputs into (23) to solve for $\boldsymbol{x}$, $\boldsymbol{z}$ and $\boldsymbol{\nu}$. The dark blue lines in the left and middle panels of Figure 1 show that, as expected, the solution is identical to the one from the structural solution of the model. The right panel then shows the sequence of shocks $\boldsymbol{\nu}$ that maps the baseline prevailing rule into the new rule. Since the new rule is more accommodating, the sequence of shocks is persistently negative (i.e., the shocks are expansionary).

b) *Optimal policy.* Our second experiment studies optimal policy under a dual mandate loss function

$$\mathcal{L} = \lambda_\pi \boldsymbol{\pi}'\boldsymbol{\pi} + \lambda_y \boldsymbol{y}'\boldsymbol{y} \tag{40}$$

with $\lambda_\pi = \lambda_y = 1$. We again start by solving for the optimal policy using conventional methods: we derive the policy rule corresponding to the first-order conditions (28) - (30), Alternative Policy Rule, HANK Model

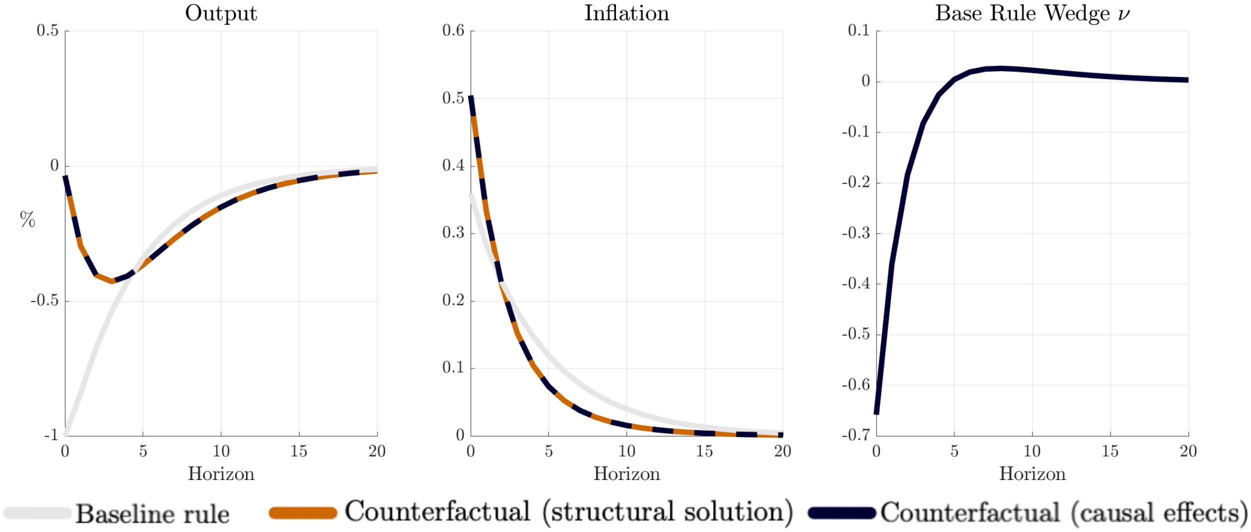


**Figure 1:** Output and inflation impulse responses together with the equivalence shock wedge $\boldsymbol{\nu}$ (see (23)) for the HANK model with policy rules (38) and (39). The impact output contraction under the prevailing baseline rule is normalized to $-1\%$.

solve the model given that policy rule, and report the result as the orange lines in the left and middle panels of Figure 2. 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 (38) tightens too much.

We then instead use Proposition 2 to equivalently recover the optimal policy rule and the corresponding cost-push shock impulse responses. We begin with the optimal rule itself. By (36), 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 dynamic causal 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 (23), again requiring only knowledge of the causal effects of policy. 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 2 shows the optimal policy

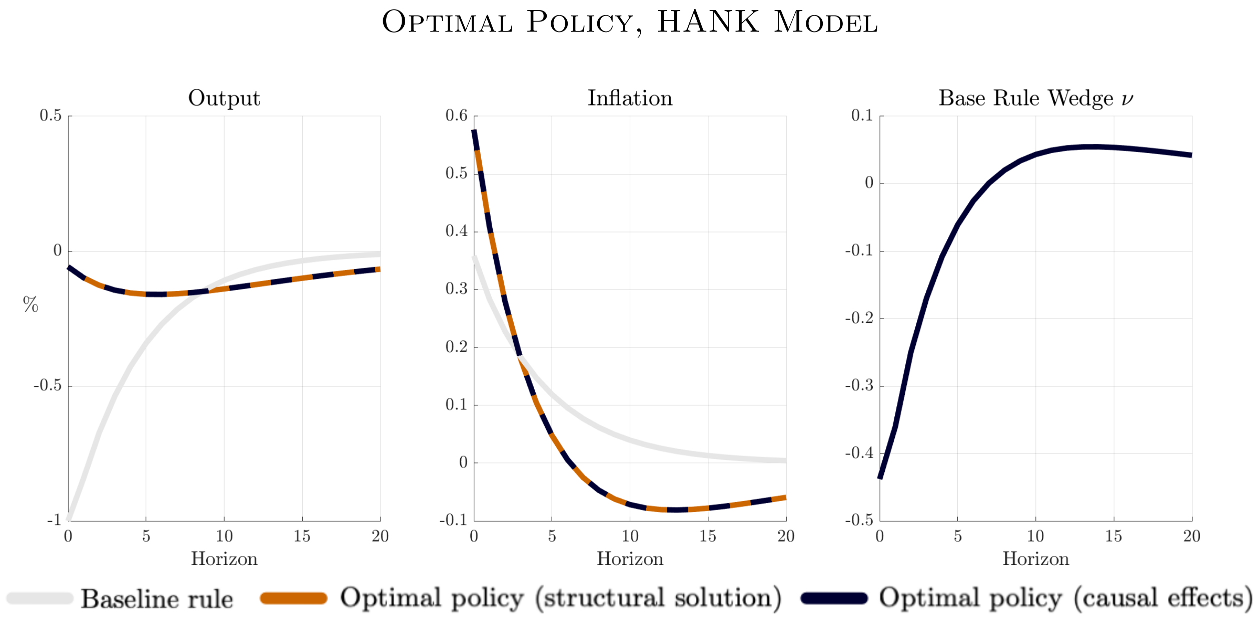


**Figure 2:** Output and inflation impulse responses together with the equivalence shock wedge $\boldsymbol{\nu}$ (see (23)) for the HANK model with policy rules (38) and the optimal policy given by (40). The impact output contraction under the prevailing baseline rule is normalized to $-1\%$.

as a deviation $\boldsymbol{\nu}$ from the prevailing baseline rule. The optimal rule leans less against the inflationary cost-push shock than the baseline rule (38), so again the required policy "shock" is persistently negative (i.e., expansionary).

## 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 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.

In the remainder of this paper we present ways to operationalize this insight. The essential hurdle faced by our approach is that its informational requirements are extremely high: we would need evidence on the dynamic causal effects of the full menu of all possible contemporaneous and news policy shocks—evidence that is clearly not available, for any policy instrument. We will present two ways of dealing with this challenge. First, in Section 3, we show that, if a researcher was able to estimate the dynamic causal effects of a limited number of policy shocks, then she can still construct counterfactuals and find optimal rules in the subspace of changes to the policy rule spanned by those observed shocks. Second, in Section 4, we discuss strategies that use restrictions coming from economic theory to map the available partial evidence into the required full menu of dynamic causal effects. Throughout, we apply our results to analyze the aggregate effects of investment-specific technology shocks under counterfactual monetary policy rules.

# 3 Counterfactuals in identified subspaces

In this section we consider the empirically relevant case of a researcher that is only able to estimate the causal effects of a finite (small) number of particular shocks to the policy rule.

Any single empirically identified shock $s$ corresponds to some particular path $\boldsymbol{\nu}_s$ of departures from the baseline policy rule. Figure 3 depicts two cases: the first identified shock induces a persistent, gradually decaying deviation from the baseline rule, while for the second shock the largest departures from the rule occur not immediately but at some future time (here at $t = 5$). We suppose that a researcher has access to $n_s$ distinct identified policy shocks and their causal effects, and we denote those causal effects by $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$, where the columns of the $\Omega$'s correspond to weighted averages of the full dynamic causal effect maps $\Theta$, with weights for the $s$th identified shock given by $\boldsymbol{\nu}_s$.[14] Anticipating our main empirical application, and in line with Figure 3, we may think of the researcher as knowing the effects of *particular* persistent and delayed anticipated interest rate movements, but not those for *any possible* pattern of interest rate adjustments.[15]

The remainder of this section proceeds as follows. First, in Section 3.1, we show that the estimated causal effects can still be used to construct counterfactuals for *some* alternative policy rules—rules that can be written as the prevailing baseline rule plus linear combinations of the $n_s$ distinct identified shocks. Second, in Section 3.2, we find the *optimal* rule within the policy space spanned by the identified shocks. We illustrate both sets of counterfactuals with applications to monetary policy transmission, leveraging dynamic causal effect estimates for

[14] Our discussion in this section focusses on the finite-shock case, so $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$ have a small number of columns. In any empirical application, those linear maps of course also have a finite number $T$ of rows. We do not pay much attention to this limitation since we consider shocks and counterfactual policies with sufficiently short-lived dynamics, making the maximal truncation horizon immaterial.

[15] Figure 3 in fact shows the mean estimated interest rate response paths for the two monetary policy shocks in our empirical application. We can interpret these paths as giving shock weights $\boldsymbol{\nu}_s$ after premultiplication by $\Theta_{z,\nu,\mathcal{A}}^{-1}$—i.e., we rotate the shocks so that they correspond one-to-one to movements in the policy instrument $z$. As long as $\Theta_{z,\nu,\mathcal{A}}$ is invertible, this rotation is without loss of generality.

Identified Policy Shock Paths, Illustration

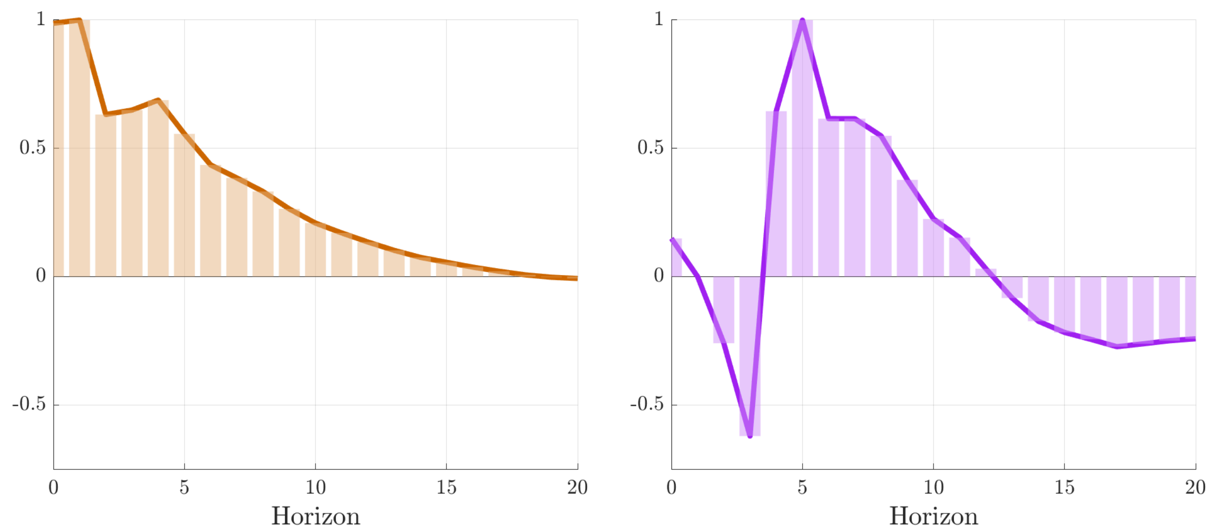


**Figure 3:** Two possible shock paths $\boldsymbol{\nu}_s$: a gradual, persistent departure from the rule (orange, left panel), and a news shock (purple, right panel), with bars indicating the shock weights $\boldsymbol{\nu}_s$.

two particular, distinct monetary policy shocks: those of Christiano et al. (1999) and Gertler & Karadi (2015).

## 3.1 Counterfactual rules

With the researcher observing $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$, the proof of Proposition 1 now only works for *particular* alternative policy rules—those that satisfy the restriction

$$\tilde{\mathcal{A}}_x(\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{x,\mathcal{A}} \times \boldsymbol{s}) + \tilde{\mathcal{A}}_z(\boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{z,\mathcal{A}} \times \boldsymbol{s}) = 0 \tag{41}$$

for some linear combination of the identified shocks $\boldsymbol{s} \in \mathbb{R}^{n_s}$. That is, we must be able to replicate the alternative rule as the prevailing baseline rule plus some linear combination of the $n_s$ particular observed shocks, with the weights given by $\boldsymbol{s}$. Equivalently, the new rule must deviate from the prevailing one in response to shocks $\boldsymbol{\varepsilon}$ in a direction that is consistent with the causal effects identified by the available policy shocks. Naturally, the larger $n_s$, the larger this identified subspace, and so the more policies satisfy (41).

If a researcher is interested in a rule outside of the spanned subspace, then one way forward is to find the best possible fit using the actually empirically observed shocks. For example, under a simple quadratic loss function for deviations from the (unattainable) target

Policy Counterfactual for Investment Shocks, Identified Subspace

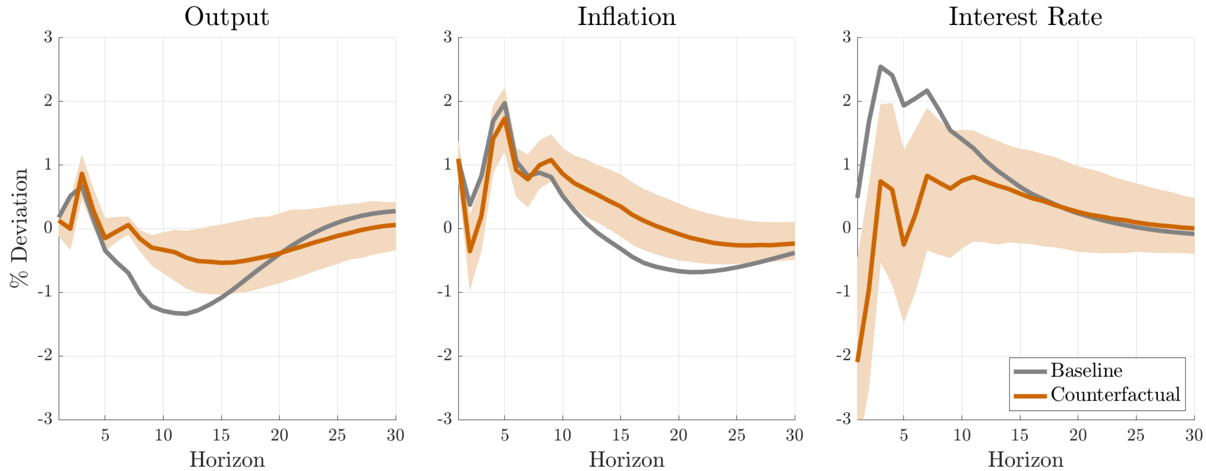


**Figure 4:** Output, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (grey) and the best feasible approximation to a rule that stabilizes output (orange). The shaded areas correspond to 16th and 84th percentile confidence bands.

counterfactual policy rule, the best-fitting shock vector would be given as

$$\boldsymbol{s} = -\left[\left(\tilde{\mathcal{A}}_q \Omega_{q,\mathcal{A}}\right)' \times \left(\tilde{\mathcal{A}}_q \Omega_{q,\mathcal{A}}\right)\right]^{-1} \times \left[\left(\tilde{\mathcal{A}}_q \Omega_{q,\mathcal{A}}\right)' \times \tilde{\mathcal{A}}_q \times \boldsymbol{q}_{\mathcal{A}}(\boldsymbol{\varepsilon})\right] \tag{42}$$

where $\boldsymbol{q} = (\boldsymbol{x}', \boldsymbol{z}')'$. Whether or not any given desired counterfactual rule is (at least approximately) contained within the space spanned by the empirically observed shocks is an inherently application-dependent question.

Application. We illustrate this approach with an application to investment-specific technology shocks. Our object of interest is the behavior of output and inflation following such a technology shock and under a counterfactual monetary policy rule that aggressively stabilizes output fluctuations. We present the main results here, and relegate empirical implementation details to Appendix C.1.

Our approach requires two inputs. First, we need to know the effects of the shock of interest under the prevailing baseline policy rule. To estimate these effects we use the investment-specific technology shock identified by Ben Zeev & Khan (2015). This shock corresponds to a short-lived, unanticipated change in the relative price of investment goods. Second, we need the effects of some (ideally rich) menu of different monetary policy shocks. We consider two of the most popular examples of such monetary policy shocks: the recursively identified shock of Christiano et al. (1999), and the high-frequency identification of Gertler & Karadi (2015). The dynamic response of nominal interest rates differs quite substantially across those two identifications schemes: gradual and long-lived for Christiano et al., and relatively transitory for Gertler & Karadi. Indeed, in our illustrative figure from before (Figure 3), the left panel shows the interest rate path corresponding to the Christiano et al. shock, while the right panel gives the *difference* between the two rate paths, corresponding to a monetary policy news shock. With these two estimates in hand, we can then follow (42) to construct the best possible approximation to a rule that aims to perfectly stabilize aggregate output.

Figure 4 presents our results. Under the prevailing baseline rule (grey), the policymaker leans against the inflationary effects of the shock, further pushing down aggregate real activity. Overall, the shock resembles a classic supply shock with inflation rising and output falling. Under our counterfactual rule (orange), monetary policy is much more accommodating, keeping output relatively close to trend throughout, but of course at the cost of persistently elevated inflation. By our identification results, any structural model of the general form (14) - (15) and consistent with our empirical estimates of monetary transmission will invariably agree with those counterfactuals for a change in the systematic policy rule.

## 3.2 Optimal policy

For optimal policy, we follow the same steps as in the proof of Proposition 2 to now consider the problem of minimizing the policymaker loss function (21) within the identified subspace of policy changes. This problem gives the optimality condition

$$\sum_{i=1}^{n_x} \lambda_i \Omega'_{x_i,\mathcal{A}} W \boldsymbol{x}_i = 0 \tag{43}$$

(43) can be interpreted in two ways. First, it gives $n_s$ restrictions that *any* solution to the full optimal policy problem must satisfy.[16] Second, it fully characterizes the optimal rule in the $n_s$-dimensional identified subspace of dynamic causal effects. The larger that space is, the more meaningful is the derived constrained optimal policy rule. In particular we by Proposition 2 know that, for $n_s \to \infty$, (43) fully characterizes the optimal policy rule.

[16] Equation (43) is related to Barnichon & Mesters (2021), who propose to use a condition of this form to test the optimality of a given policy. Since their analysis relies on fixed private sector expectations, they do not draw any implications for optimal policy *rules*, unlike our approach.

Optimal Policy for Investment Shocks, Identified Subspace

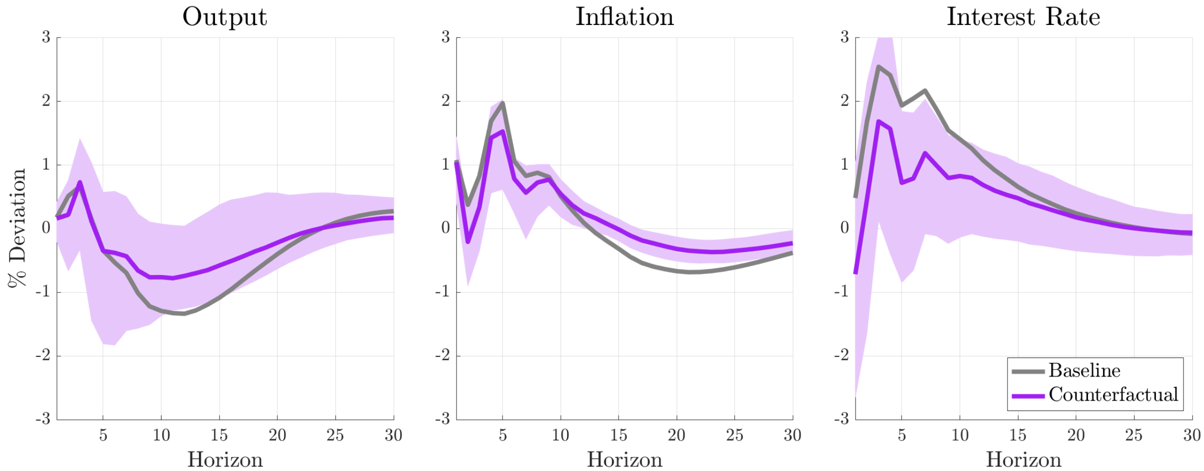


**Figure 5:** Output, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (grey) and the optimal policy rule for a policymaker with preferences over current output and averaged past inflation (see Appendix A.2) within the identified subspace (purple). The shaded areas correspond to 16th and 84th percentile confidence bands.

Application. We illustrate our conclusions for optimal policy rules with another application to the investment-specific technology shocks discussed in Section 3.1. As before, our analysis leverages estimates of the causal effects of monetary policy based on Christiano et al. (1999) and Gertler & Karadi (2015). We now consider a policymaker whose loss function puts equal weight on the deviations of aggregate output from trend and the deviations of a weighted average of current and lagged inflation from target. This loss function is one interpretation of a flexible average inflation targeting framework.[17] We then use (43) to recover the constrained optimal policy rule in the identified subspace as well as the corresponding counterfactual paths of the policy instrument and the two targets.

Figure 5 presents our results. Our analysis reveals that the (constrained) optimal policy rule increases interest rates somewhat less than the observed baseline policy response. As a result, the path for output is somewhat closer to trend than in the baseline and the deviation in inflation is somewhat smaller at longer horizons. On the whole, however, the differences between the implied optimal policy and the baseline policy are fairly small, suggesting that the observed policy response was close to optimal for a policymaker with these preferences.

[17]See Appendix A.2 for a detailed discussion of the corresponding loss function.

# 4 Imposing additional structure

The second approach to implementing our identification result in the face of limited empirical evidence imposes *additional structure* to extrapolate from the dynamic causal effects of the policy shocks that we do observe—the finite-shock maps $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$—to those that we did not observe—the full maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$. Mathematically, we face a matrix completion problem: we know (or at least can estimate) certain linear combinations of the columns of the dynamic causal effect maps, and would like to fill in the remaining columns.

One natural solution is to impose parametric structure on the full causal effect maps—that is, write $\Theta = \Theta(\vartheta)$ for some parameter vector $\vartheta$—and then estimate $\vartheta$ through estimates of the individual shock impulse responses $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$, which themselves are known functions of $\Theta$ and so $\vartheta$. If $\Theta$ is parameterized through a particular fully-specified structural model, then this approach simply amounts to model estimation via impulse response matching, as done routinely in the Lucas program (Christiano et al., 2005). The Sims program on the other hand constructs its counterfactuals only using empirical evidence on a single policy shock, and nothing else (Sims & Zha, 2006). We will show that, under certain structural assumptions on $\Theta$, these traditional Sims program counterfactuals in fact can also be interpreted as achieving the Lucas program objective of studying a change in the policy rule. Section 4.1 elaborates on this connection between familiar approaches and our perspective on impulse response extrapolation. Our novel contribution will then follow in Section 4.2: we will show that, for several interesting counterfactuals, the specification of partial model *blocks*—rather than an entire general equilibrium model—may well suffice to give the structure on $\Theta$ required to operationalize our theoretical identification result.

## 4.1 Two familiar special cases

We begin by re-interpreting two popular existing approaches to policy counterfactuals as two particular strategies of completing the causal effect maps $\Theta$ from evidence on individual identified policy shocks.

Impulse response matching. The Lucas program constructs counterfactuals using micro-founded general equilibrium models. One popular technique for estimating such a model is impulse response matching (e.g. Rotemberg & Woodford, 1997; Christiano et al., 2005)—the researcher chooses the parameters of the model to match the estimated impulse response functions $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$ for an *individual* identified policy shock as closely as possible.

Through the lens of this paper, such model estimation via impulse response matching may be interpreted as a particular approach of mapping causal evidence on a single identified shock into the effects of all other (unobserved) shocks. The parametric model provides structure on $\Theta$ that is indexed by the model parameters $\vartheta$, and those model parameters are chosen to ensure agreement with the empirically estimated dynamic causal effects of policy shocks. Thus, in this context of policy counterfactual analysis, the only purpose of the model is to extrapolate from (estimated) individual columns into the rest of $\Theta$.

Counterfactuals as repeated surprises. Researchers in the Sims program begin by estimating the dynamic causal effects of a *contemporaneous* policy shock $\nu_0$ to the prevailing policy rule. 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, Sims & Zha (2006) propose to subject the economy to a *sequence* of contemporaneous policy shocks that enforce the desired instrument path in equilibrium. This approach answers the traditional Sims program question of predicting counterfactuals without the public perceiving a change in policy regime. Alternatively, however, this approach may also be interpreted as answering the more ambitious Lucas program question under auxiliary structural assumptions—assumptions that put further structure on $\Theta$. In particular, when translated to our notation, the implied structure is that the maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ are lower-triangular, with the columns $j \geq 2$ equal to a time-shifted version of the first column:

$$\Theta_{q,\nu,\mathcal{A}} = \begin{pmatrix} \Theta_{q,\nu,\mathcal{A}}(1,1) & 0 & 0 & \dots \\ \Theta_{q,\nu,\mathcal{A}}(2,1) & \Theta_{q,\nu,\mathcal{A}}(1,1) & 0 & \dots \\ \Theta_{q,\nu,\mathcal{A}}(3,1) & \Theta_{q,\nu,\mathcal{A}}(2,1) & \Theta_{q,\nu,\mathcal{A}}(1,1) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad q \in \{x, z\} \tag{44}$$

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 VAR. Intuitively, 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).

A structure like that in (44) is 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 (44) holds. In such environments, the conventional

Sims program also correctly answers the more ambitious question of predicting the effects of changes in systematic policy rules.[18] Typical (rational-expectations) macroeconomic models with forward-looking agents, on the other hand, allow for an important role of news shocks (e.g. Schmitt-Grohé & Uribe, 2012) and so are inconsistent with (44). In such environments, using the structure in (44) to predict the effects of changes in systematic policy rules will invariably run afoul of the Lucas critique.

## 4.2 Putting structure on the output-inflation trade-off

The special cases we have just discussed reflect two extremes: one branch uses a fully specified, parametric structural model to analyze a change in policy rule, while the other is appealingly agnostic about the detailed structure of the economy, but is either uninformative about systematic rule changes or requires strong restrictions on private sector expectations. We now demonstrate a different, hybrid approach to imposing structure on $\Theta$: we will show that, for several interesting policy counterfactuals, it is sufficient to impose structure that, on the one hand, leaves many aspects of the model economy unspecified, yet on the other hand does not need to assume the absence of private sector anticipation effects.[19]

Objects of interest. We consider an econometrician interested in the behavior of aggregate output and inflation under policy rules of the form

$$\mathcal{A}_\pi \boldsymbol{\pi} + \mathcal{A}_y \boldsymbol{y} = \mathbf{0} \tag{45}$$

Note that (45) nests contemporaneous as well as average inflation targeting, nominal GDP targeting, as well as strict output and inflation stabilization. Thus, knowledge of counterfactual outcomes under (45) will in particular pin down our two desired policy counterfactuals in Sections 3.1 and 3.2.

By our results in Section 2, knowledge of $\Theta_{\pi,\nu,\mathcal{A}}$ and $\Theta_{y,\nu,\mathcal{A}}$ is sufficient to construct counterfactuals for rules like (45). Our key insight is that, for structural models that feature a Phillips curve relationship, that Phillips curve provides all the structure we need.

[18] Furthermore, if agents are quite but not perfectly inattentive (as for example in Auclert et al., 2020), then this approach may deliver a reasonably accurate approximation to correct policy counterfactuals.

[19] We present yet another example of possible structure on $\Theta$ in Appendix A.4.

Structure via dynamic Phillips curves. Using our perfect-foresight notation of Section 2, we can write a general Phillips curve relationship as

$$\boldsymbol{\pi} = \Pi_y \times \boldsymbol{y} + \Pi_\varepsilon \times \boldsymbol{\varepsilon}. \tag{46}$$

Here $\Pi_y$ is the linear map summarizing the structural relationship between inflation and leads and lags of output, up to (non-policy) shocks $\Pi_\varepsilon \times \boldsymbol{\varepsilon}$. For example, in the very simple New Keynesian model of Section 2.1, $\Pi_y$ would take the form

$$\Pi_y = \begin{pmatrix} \kappa & \kappa\beta & \kappa\beta^2 & \dots \\ 0 & \kappa & \kappa\beta & \dots \\ 0 & 0 & \kappa & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{47}$$

The crucial implication of (46) is that, conditional on policy shocks $\boldsymbol{\nu}$, the co-movements of output and inflation are fully characterized by the map $\Pi_y$:

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

In words, we can map output impulse responses into inflation impulse responses (and vice-versa) using only $\Pi_y$. It then follows from our discussion of relative impulse responses in Section 2.3 that knowledge of $\Pi_y$ is sufficient to construct counterfactuals for alternative policy rules of the general form (45), as required.[20] We have thus reduced the problem of extrapolating across columns of the two maps $\Theta_{\pi,\nu,\mathcal{A}}$ and $\Theta_{y,\nu,\mathcal{A}}$ to the simpler one of learning only about the single map $\Pi_y$.

We now proceed in analogy with standard impulse response matching exercises. Given some theory-guided, parametric structure on $\Pi_y$, we propose to estimate those parameters (and thus all of $\Pi_y$) through empirical evidence on identified policy shocks—that is $\Omega_{\pi,\mathcal{A}}$ and $\Omega_{y,\mathcal{A}}$. Rather than requiring a full structural general equilibrium model to match all *absolute* impulse responses, however, here we only require one part of a model to be consistent with empirically observed *relative* impulse responses. Importantly, the implied counterfactuals will be valid independently of any further model details, including preferences, technology,

[20]Strictly speaking, we additionally require the assumption of invertibility of $\Theta_{\pi,\nu,\mathcal{A}}$—that is, the policymaker can implement any possible path of inflation. This assumption 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 A.3.

the nature of expectation formation, and so on—as long as two structural models agree on the Phillips curve map $\Pi_y$ they also agree on the desired policy counterfactuals.

Application. We now use this method to analyze the same two counterfactual responses to investment-specific technology shocks that we considered in Sections 3.1 and 3.2, but now leveraging the extrapolated full impulse response map $\Pi_y$ rather than being restricted to the identified subspaces of causal effects.

We assume that $\Pi_y$ is derived from a hybrid Phillips curve relationship:

$$\pi_t = \gamma_b \pi_{t-1}^4 + \gamma_f \mathbb{E}_t \left[\pi_{t+4}^4\right] + \kappa y_t + \varepsilon_t \tag{48}$$

where $\pi_{t-1}^4 = \frac{1}{4} \times (\pi_{t-1} + \pi_{t-2} + \pi_{t-3} + \pi_{t-4})$. Appendix A.3 shows the linear map $\Pi_y$ corresponding to this Phillips curve specification. We then estimate the parameters $\{\gamma_b, \gamma_f, \kappa\}$ (and so all of $\Pi_y$) using evidence on identified monetary policy shocks. The econometric challenge is that estimates of $\{\Omega_{\pi,\mathcal{A}}, \Omega_{y,\mathcal{A}}\}$ will not perfectly align with the parametric structure imposed by (48); thus, following Barnichon & Mesters (2020), we simply find the best possible fit. Our estimation uses the identified monetary policy shocks of Gertler & Karadi (2015), already discussed in Section 3.

Given an estimate of $\Pi_y$, we can construct the two desired counterfactuals: output and inflation responses to investment-specific technology shocks under counterfactual policy rules that a) perfectly stabilize output and b) are optimal for a dual mandate policymaker with equal weights on aggregate output and an average of current and lagged inflation. The results, reported in Figure 6, closely echo those of Section 3.[21] First, perfect output stabilization implies persistently elevated inflation relative to the baseline rule outcome. Second, the output and inflation impulse response paths under the optimal average inflation targeting policy are relatively close to observed outcomes, but with somewhat smoother output dynamics. With a Phillips curve of the form (48), we can by Proposition 2 in fact explicitly characterize the optimal policy rule as

$$\lambda_\pi \bar{\Pi}' \bar{\boldsymbol{\pi}} + \lambda_y (\Pi_y')^{-1} \boldsymbol{y} = \mathbf{0} \tag{49}$$

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 Appendix A.2), and $\Pi_y$ displayed in Appendix A.3.

[21]Note that $\Pi_y$ is sufficient to characterize the (relative) implementable space of output and inflation, but does not allow us to solve for the nominal interest rate path that is required to engineer those allocations.

Policy Counterfactuals for Investment Shocks, PC Extrapolation

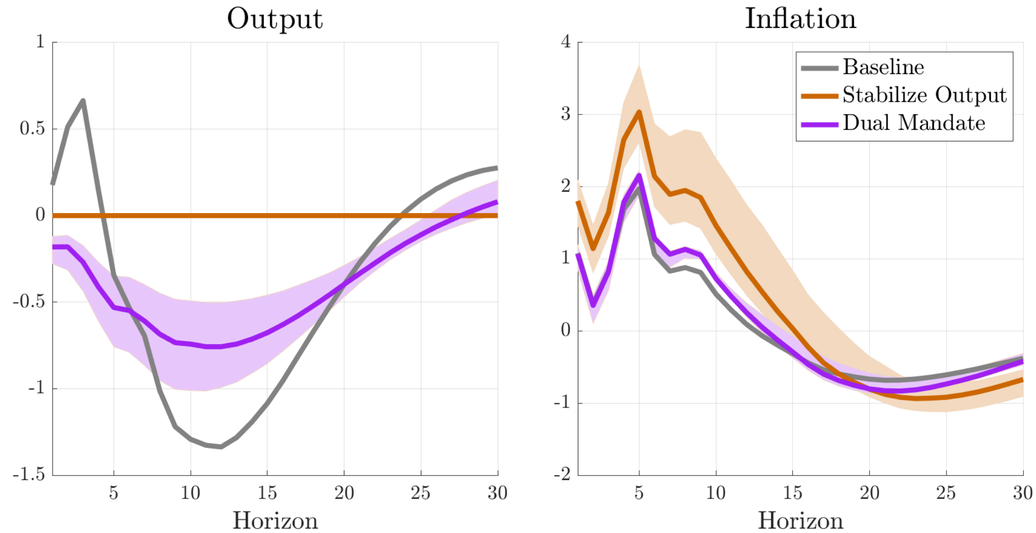


**Figure 6:** Output and inflation impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (grey), a counterfactual rule that perfectly stabilizes output (grey), and the optimal policy rule for a dual-mandate policymaker with preferences over current output and averaged past inflation (see Appendix A.2) (purple). The shaded areas correspond to 16th and 84th percentile confidence bands.

We note that (49) takes the form of an implicit targeting rule (Svensson, 1997): it imposes a set of restrictions that current, lagged and expected future values of output and inflation must satisfy at all times when policy is set optimally.

Overall, it follows from our analysis that any fully specified general equilibrium structural model that (i) fits into the general form (14) - (15), (ii) features a Phillips curve relationship of the form (48) and (iii) is consistent with the empirical monetary policy shock estimates of Gertler & Karadi (2015) will produce the same counterfactuals as in Figure 6, and in particular yield the optimal policy rule (49).

# 5 Conclusions

The standard approach to counterfactual analysis for changes in policy rules relies on fully-specified, dynamic general equilibrium models. Our identification results instead point in a different direction: for valid policy counterfactuals, researchers can estimate dynamic causal effects of policy shocks and combine them to form policy counterfactuals and optimal policy responses. 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.

The main challenge in implementing this strategy is that evidence on an infinitely large set of policy perturbations is unattainable. One natural reaction is simply to try to get as close as possible; viewed in this light, future empirical work should try to identify policy "shocks" that correspond to as many different *paths* of the policy instrument as possible.[22] Every additional piece of empirical evidence will allow researchers to a) expand the space of alternative, counterfactual policy rules that we can analyze and b) find further restrictions that help to more tightly characterize optimal rules. A second response to the limited empirical evidence is to impose more structure, as we do in Section 4. Further research exploring the sets of counterfactuals that can be characterized with only partial model structure (like we do in Section 4.2) would be particularly welcome.

[22] To this end, the "functional VAR" approach of Inoue et al. (2021) seems particularly promising.

# References


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. (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.

Baxter, M. & King, R. G. (1993). Fiscal policy in general equilibrium. *The American Economic Review*, *83*(3), 315–334.

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.

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.

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. (2019). The federal reserve's current framework for monetary policy: A review and assessment. NBER Working Paper.

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.

Hansen, L. P. & Sargent, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and control*, *2*, 7–46.

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

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

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. & 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.

Lucas, R. E. (1980). Methods and problems in business cycle theory. *Journal of Money, Credit and banking*, *12*(4), 696–715.

McKay, A. & Wieland, J. F. (2021). Lumpy durable consumption demand and the limited ammunition of monetary policy. Forthcoming in Econometrica.

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. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Powell, J. H. (2020). Opening remarks: New economic challenges and the fed's monetary policy review. In *Proceedings-Economic Policy Symposium-Jackson Hole*, (pp. 1–18). Federal Reserve Bank of Kansas City.

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

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. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, *48*(1), 1–48.

Sims, C. A. (1982). Policy analysis with econometric models. *Brookings papers on economic activity*, *1982*(1), 107–164.

Sims, C. A. (1987). A rational expectations framework for short-run policy analysis. In Barnett, W. A. & Singleton, K. J. (Eds.), *New approaches to monetary economics*, (pp. 293–308)., Cambridge, UK. Cambridge University Press.

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.

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

Svensson, L. E. (2020). Monetary policy strategies for the federal reserve. *International Journal of Central Banking*.

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.

# A Supplementary results

This appendix presents several supplementary theoretical results. Appendix A.1 begins by extending Proposition 2 to a more general loss function, and Appendix A.2 considers the special case of a loss function over averaged inflation. Appendix A.3 then provides additional information for our Phillips curve structure used in Section 4.2. Finally in Appendix A.4 we present another possible approach to the matrix completion problem.

## A.1 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 (35). 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

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

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 (15), so we do not consider this case any further here.

## A.2 Average inflation targeting loss function

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

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

Here $H$ denotes the maximal (lagged) horizon that enters the inflation averaging, and $\omega_\ell$ denotes the weight on the $\ell$th lag, with $\sum_\ell \omega_\ell = 1$ and $\omega_\ell \geq 0 \; \forall \ell$. For our applications in Sections 3 and 4 we set $H = 20$ 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 represent the loss function of a dual mandate policymaker with preferences over average inflation as

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

For our applications we set $\lambda_\pi = \lambda_y = 1$—an equal weighting of the two mandates. For such a loss function (and setting $W = I$ for simplicity), we find the optimal policy rule as

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

Using the definition of $\Pi_y$ and simplifying, we recover (49).

## A.3 Phillips curve & policy counterfactuals

Consider the augmented Phillips curve (48). 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

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

Policy counterfactuals. Knowledge of $\Pi_y$—together with the assumption that $\Theta_{\pi,\nu,\mathcal{A}}$ is invertible, i.e., any path of inflation is in principle implementable—is sufficient to construct output and inflation counterfactuals corresponding to alternative rules of the general form (45). Formally, 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 $\{\boldsymbol{\pi}, \boldsymbol{y}, \boldsymbol{\nu}\}$.

Invertibility of $\Theta_{\pi,\nu,\mathcal{A}}$. Strictly speaking, our results leveraging $\Pi_y$ impose 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

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

## A.4 Leveraging time invariance

This section presents another possible approach to the matrix completion problem of Section 4, based upon a property of the causal effect maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ that we refer to as asymptotic time invariance.

Definition. Asymptotic time invariance formalizes the idea that the different columns of the causal effect maps are not completely independent objects – for example, impulse

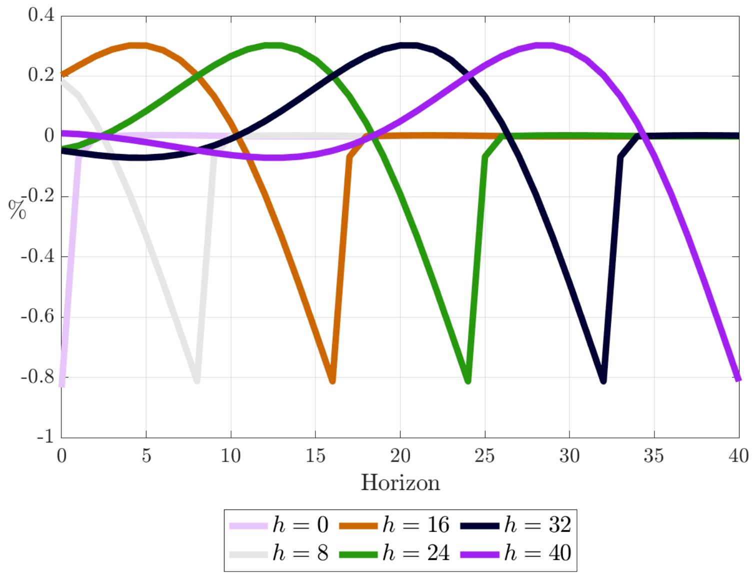


**Figure A.1:** Output impulse responses to contemporaneous and forward guidance monetary policy shocks in the HANK model of Appendix B.4.

responses to a forward guidance shock eight quarters out should be very similar to forward guidance shocks nine quarters out, just shifted by one period. The precise definition is that, for all $s \in \mathbb{N}$,

$$\lim_{t\to\infty} \Theta_{x,\nu,\mathcal{A}}(t+s,t) = \bar{\Theta}_{x,\nu,\mathcal{A}}(s), \quad \lim_{t\to\infty} \Theta_{z,\nu,\mathcal{A}}(t+s,t) = \bar{\Theta}_{z,\nu,\mathcal{A}}(s) \tag{A.2}$$

where $\bar{\Theta}_{x,\nu,\mathcal{A}}$ and $\bar{\Theta}_{z,\nu,\mathcal{A}}$ are two sequences. Figure A.1 provides an illustration of this property in the quantitative HANK model of Section 2.4, 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.

Time invariance as explicit structure. Imposing (A.2) after some finite horizon $H$ reduces the problem of dynamic causal effect identification from an infinite-dimensional one to an $H + 1$-dimensional one. For example, imposing time invariance from horizon $H = 0$

onwards implies that the causal effect matrices have the following particular structure:

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

The sequence $\bar{\Theta}_{q,\nu,\mathcal{A}}$ can be estimated using empirical evidence on a forward guidance shock sufficiently far into the future. (A.3) then provides the mapping from the sequence $\bar{\Theta}_{q,\nu,\mathcal{A}}$ into the entire causal effect map.

# B Model details

This appendix provides details on several of the structural models used in this paper. First, in Appendix B.1, we show how to express the simple three-equation model in our general linearized perfect foresight notation. Appendices B.2 and B.3 then discuss the extent to which our identification results apply to general behavioral models and to models in which the private sector solves a filtering problem. Finally Appendix B.4 gives further details for the HANK model of Sections 2.2 and 2.4.

## B.1 Linear maps for the canonical New Keynesian model

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 (14). Finally the policy rule can be written as

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

which fits into the form of (15).

## B.2 Behavioral models

Our general framework (14) - (15) nests popular behavioral models such as the cognitive discounting framework of Gabaix (2020) or the sticky information of Carroll et al. (2018) or Auclert et al. (2020). 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 $\theta$ 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 (14), but do not affect the separation of policy- and non-policy blocks at the heart of our identification result.

## B.3 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 (14) - (15), 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 $\varepsilon_t^m$ is a policy shock with volatility $\sigma_m$. 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 $\theta$ follows from a filtering problem and is given as

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

with $\sigma_z$ denoting the (exogenous) volatility of idiosyncratic demand shocks and $\sigma_p$ 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 $\theta$ depends on the policy rule coefficient $\phi_y$, breaking our separation assumption.

## B.4 HANK model details & parameterization

The HANK model sketched in Section 2.2 and used for our quantitative illustration in Section 2.4 is exactly the same as in Wolf (2021) (including the parameterization), with only one change: rather than imposing uniform hours worked $\ell_{it} = \ell_t$ for all households $i$, we consider a labor rationing rule that ensures that

$$w_t \ell_{it} e_{it} + d_{it} = e_{it} y_t$$

for all households $i$. That is, the sum of labor income $w_t \ell_{it} e_{it}$ and dividend income $d_{it}$ for all households $i$ just equals aggregate output scaled by household $i$'s productivity. This rationing rule is feasible since $\int_0^1 w_t \ell_{it} e_{it} + \int_0^1 d_{it} = w_t \ell_t + d_t = y_t$, and it allows us to write the perfect foresight consumer demand block as

$$\boldsymbol{c} = \mathcal{C}\left(\boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{i}, \boldsymbol{\tau}, \boldsymbol{\varepsilon}^d\right)$$

Linearizing, we recover (17).

# C Empirical appendix

This appendix elaborates on our empirical estimation. First, in Appendix C.1, we discuss the causal effects estimated for the counterfactuals in Section 3. Second, in Appendix C.2, we briefly review the approach of Barnichon & Mesters (2020) to Phillips curve estimation, used for our counterfactuals in Section 4. Third, in Appendix C.3, we provide results for an alternative specification of the Phillips curve.

## C.1 Shock & policy dynamic causal effects

Our analysis of investment-specific technology shocks follows Ben Zeev & Khan (2015), while our monetary policy shock identification mimics that of (i) Christiano et al. (1999) and (ii) Gertler & Karadi (2015).

Outcomes. We are interested in impulse responses of three outcome variables: output, inflation, and the policy rate. For output, we follow Ramey (2016) and deflate per-capita nominal GDP by the GDP deflator. For inflation, we consider the annualized growth rate of the GDP deflator. All series are obtained from the replication files for Ramey (2016). Finally, we consider the federal funds rate as our measure of the policy rate, obtained from the St. Louis Federal Reserve FRED database. All series are quarterly.

Shocks & identification. We take the investment-specific technology shock series from Ben Zeev & Khan (2015) and the high-frequency monetary policy surprise series from Gertler & Karadi (2015), aggregated to quarterly frequency through simple averaging. Recursive shocks are identified through the estimated VAR itself.

Estimation details. 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, 2021), estimated on a sample from 1969:Q1–2007:Q4. Our estimation of the Gertler-Karadi shock is identical, except for the fact that—because of constraints on when the shock is actually available—the sample runs from 1990:Q1 – 2012:Q4. Finally, for a recursive shock, we return to our baseline long sample period, and now identify a monetary shock as the *last* recursively ordered shock in a system containing output, inflation, and the nominal rate.

Federal Funds Rate, Christiano et al. (1999) vs. Gertler & Karadi (2015)

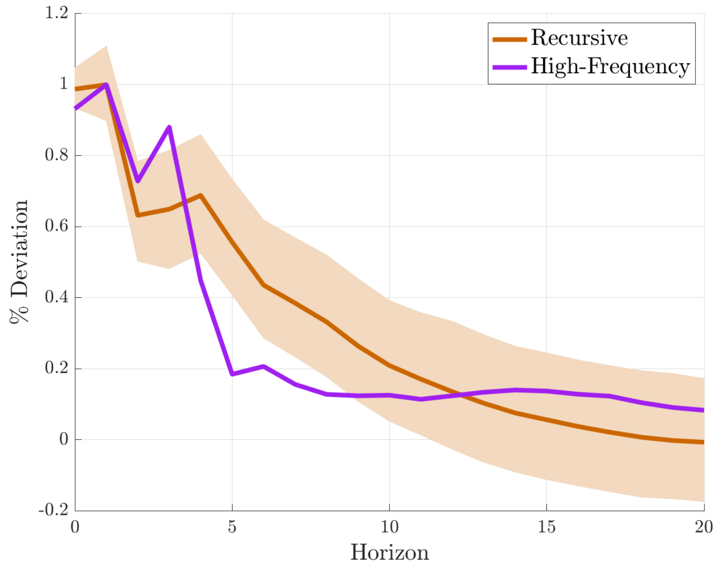


**Figure C.1:** Federal funds rate impulses to the recursive (orange) and high-frequency (purple) monetary policy shocks of Christiano et al. (1999) and Gertler & Karadi (2015), with the peak impulse normalized to 1 per cent.

We estimate all VARs using four lags, a constant, and 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. To do so we separately draw from the different monetary policy model posteriors and then compute the counterfactuals for each draw, thus effectively imposing independence across the estimated VARs.

Interest rate paths. Figure C.1 shows impulse responses of the federal funds rate to the two estimated monetary policy shocks. Consistent with previous work, we find that the recursively identified shock induces much more persistent interest rate movements than a shock identified via high-frequency surprises.[23] Figure 3 in fact uses the displayed mean estimates of interest rate impulse responses to illustrate different possible shock paths: a persistent change in interest rates (equal to the shock of Christiano et al.) in the left panel,

[23]The third well-known example of an identified monetary policy shock—that of Romer & Romer (2004)—induces interest rate movements that are relatively similar to our recursively identified shock, so it adds little to our construction of policy counterfactuals.

Policy Counterfactuals for Investment Shocks, Simple PC Extrapolation

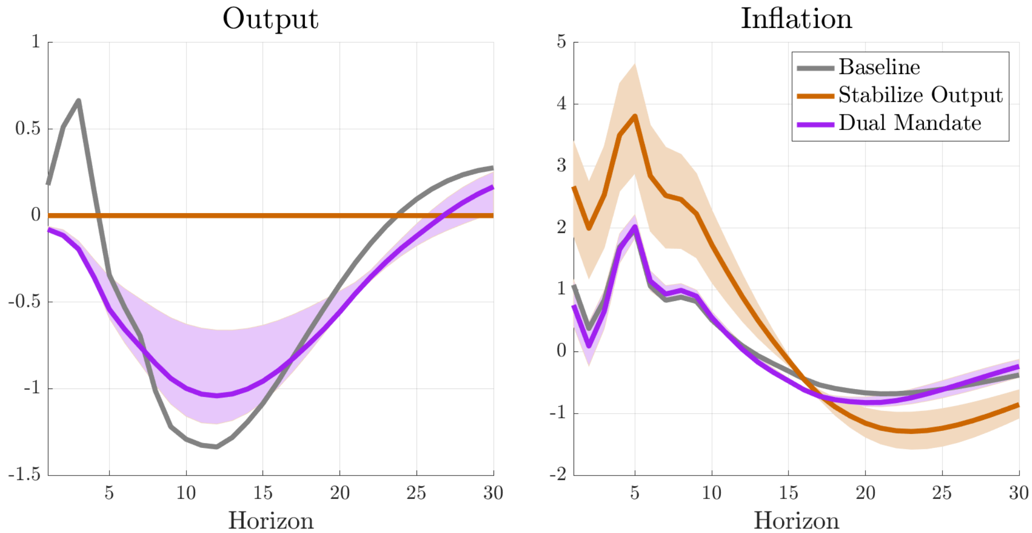


**Figure C.2:** Output and inflation impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (grey), a counterfactual rule that perfectly stabilizes output (grey), and the optimal policy rule for a dual-mandate policymaker with preferences over current output and averaged past inflation (see Appendix A.2) (purple). The shaded areas correspond to 16th and 84th percentile confidence bands.

and the difference between the two shocks—interpretable as an interest rate news shock—in the right panel.

## C.2 NKPC estimation

Barnichon & Mesters (2020) show how to use estimates of monetary policy impulse responses to identify a Phillips curve relationship of the form (48). For our empirical analysis in Section 4.2 we rely on the point estimates and the confidence region corresponding to their Gertler & Karadi analysis (which imposes the additional constraint that $\gamma_f + \gamma_b = 1$), reported in Table IV and Figure V of their paper.

## C.3 Alternative NKPC

For comparison, we also repeat our analysis in Section 4.2 for the simple Phillips curve relationship (47) replacing our general hybrid specification. Results are reported Figure C.2. With this alternative structure inflation moves slightly more and the reversal is faster, reflecting the lack of smoothing due to the absence of backward-looking terms in (47).