

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

Alisdair McKay
FRB Minneapolis

Christian K. Wolf
MIT & NBER

This version: February 2022

First version: July 2021

Abstract: We show that, in a general family of linearized structural macroeconomic models, knowledge of the empirically estimable causal effects of contemporaneous and news shocks to the prevailing policy rule is sufficient to: 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, Robert Ulbricht, Mike Waugh, Iván Werning, and seminar participants at various venues. Jackson Mejia provided superb research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

An important function of macroeconomics is to predict the consequences of changes in policy. In 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 policy *rule*, with that change understood by the private sector. The 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., 2005). Alternatively, in the Sims (1980, 1982, 1987) program, the researcher studies changes in the policy stance that are *not* perceived as a change in the systematic rule. For this more modest objective, the Sims program can rely on time series statistical techniques and so remains appealingly agnostic about the deep structure of the economy.¹

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 model, however, we begin with a general linear data-generating process. We then impose one key restriction: that policy affects private-sector behavior *only* through the current and future expected path of the policy instrument (say, the nominal interest rate). Importantly, once linearized, many of the structural models popular in the Lucas program—from 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 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. Intuitively, through a judicious choice of contemporaneous and news policy shocks, we can set private-sector expectations of the policy instrument equal to what they would be under a counterfactual rule. By our assumptions on how policy is allowed to shape private-sector behavior, this is precisely enough to yield a counterfactual that is robust to the Lucas critique. Our second contribution is to offer ways of operationalizing this identification result in the empirically relevant case when only a limited amount of causal evidence on policy news shock transmission is available.

¹For example see Bernanke et al. (1997), Sims & Zha (2006) or Eberly et al. (2020) for examples of counterfactual policy analysis along these lines.

IDENTIFICATION RESULT. The first part of the paper establishes the main identification result. We consider an econometrician living in an economy consistent with our structural assumptions. The economy is closed with some 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.² We furthermore assume that the prevailing policy rule is subject to 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 contemporaneous shocks to the policy rule as well as the effects of news about future deviations from the policy rule *at all horizons*, then those estimates contain all the information she needs to construct her two desired counterfactuals. 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 expected instrument path that it implies—can equivalently be synthesized by adding well-chosen *shocks* to the baseline policy rule. For example, in the eyes of the private sector, a prevailing dovish monetary policy rule subject to a particular sequence of contractionary interest rate shocks is identical to some counterfactual hawkish policy rule without any shocks.

How general is the setting of this identification result? 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 the private-sector block of the textbook New Keynesian model, which consists of the two equations

$$y_t = -\sigma (i_t - \mathbb{E}_t \pi_{t+1}) + \mathbb{E}_t y_{t+1},$$

$$\pi_t = \kappa y_t + \beta \mathbb{E}_t \pi_{t+1},$$

where y_t is the output gap, i_t is the nominal rate of interest, and π_t is inflation. The structural parameters σ , κ , and β are unrelated to the conduct of monetary policy. Policy decisions affect the private sector through the implied movements of the nominal rate i_t ; conditional on the dynamics of i_t , any further properties of the policy rule determining i_t —for example

²Our 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 extent to which the central bank leans against inflation—are irrelevant.³ This sufficiency of the policy instrument is a property shared by many of the linearized models typically used in the Lucas program, including those with many frictions, shocks, and rich microeconomic heterogeneity.⁴ 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 linearity assumption also limits the set of policy counterfactuals to which our method can be applied. Our approach can be used to compare different cyclical stabilization policies (e.g., monetary or fiscal policy rules for business-cycle policy), but is less well-suited to study policies that alter the steady state (e.g., changes in the inflation target or in the long-run fiscal system).

PRACTICAL IMPLEMENTATION. The main challenge to implementing our approach is that existing empirical evidence on the effects of policy shocks is limited. Recall that our identification result requires the econometrician to estimate the causal effects of contemporaneous and news shocks to the prevailing policy rule at all horizons. 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. We discuss two alternative strategies for dealing with this lack of data.

The first strategy is to focus on a narrower set of counterfactuals—those for which we already have sufficient evidence. Suppose that the econometrician can estimate the causal effects of some (perhaps small) set of policy shocks. Then, by our identification result, we can construct counterfactuals for all alternative policy rules under which the policy instrument deviates from the prevailing rule in a way consistent with a linear combination of 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. More generally, even if a given counterfactual rule is not exactly nested, existing evidence

³More 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.

⁴Note 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.

may suffice to approximate it well. The same logic applies to optimal policy analysis: we can always find the (constrained) optimal policy rule within the spanned set of estimated policy shock causal effects.

We illustrate these insights with an application to monetary policy counterfactuals. Our object of interest is the counterfactual propagation of investment-specific technology shocks 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, as in the recent review of the Federal Reserve’s policy framework.⁵ As the inputs to our methodology, we take the causal effects associated with two popular monetary policy shock series: those of Romer & Romer (2004) and Gertler & Karadi (2015). Armed with those two sets of dynamic causal effects, we then construct an approximation to our two desired counterfactuals.

The second strategy addresses counterfactuals under which the policy instrument deviates from the prevailing rule in a way that is inconsistent with the empirically identified shocks. To make progress, this strategy relies on additional structural assumptions to extrapolate from the causal effects of policy shocks that we did observe to those that we did not observe. How much structure one needs to impose depends on the counterfactuals of interest. We show that for an important class of counterfactuals—those for which the counterfactual policy rule can be expressed as an implicit targeting rule for the output gap and inflation—it is sufficient to impose structure on just the supply side of the model. Specifically, a parametric dynamic Phillips curve relationship and evidence on identified policy shocks allow us to form these counterfactuals while remaining agnostic about the demand side of the economy.

Outside of the class of counterfactuals covered by our two approaches, a full general equilibrium structure is in general again necessary, bringing us back to the original Lucas program. We argue that our identification results nevertheless remain useful, for the following reason. While the identification result states that impulse responses to different policy shocks are formal “sufficient statistics” for policy counterfactuals, economic intuition suggests that the effects of policy news shocks across different horizons are likely to be tightly related. Building on Andrews et al. (2020), we provide computational tools to assess whether or not this is so in any given parametric model. We find the answer to be affirmative in the large-scale structural model of Smets & Wouters (2007), suggesting that evidence on the causal

⁵We 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.

effects of individual policy shocks can serve as powerful “identified moments” (Nakamura & Steinsson, 2018) for the entire universe of possible systematic policy rule counterfactuals. This analysis provides a formal argument for impulse response matching as a way to estimate structural general equilibrium models (Christiano et al., 2005).

LITERATURE. Within the Lucas program, some researchers use identified policy shocks as a means to identify the structural parameters of their model (e.g. Rotemberg & Woodford, 1997; Christiano et al., 2005). Our results show that that evidence on *many* policy shocks in principle obviates the need to specify a particular parametric structural model. Within the Sims program, counterfactuals are usually constructed using a sequence of *unanticipated* policy shocks that enforce an alternative policy rule *ex post* along the equilibrium path (Sims & Zha, 2006; Bernanke et al., 1997; Kilian & Lewis, 2011; Eberly et al., 2020) but not in *ex ante* expectation. This approach will be credible if the private sector is unlikely to detect the change in policy regime (Leeper & Zha, 2003; Hamilton & Herrera, 2004; Antolin-Diaz et al., 2021), but is less appropriate to analyze the effects of announced regime changes such as the Federal Reserve’s recent adjustment to its policy framework (Powell, 2020). By using policy news shocks, our approach ensures that the counterfactual also holds *ex ante*, in private-sector expectations. Aligning these expectations allows us to study systematic changes in policy without violating the Lucas critique.

Our work also relates to other recent contributions to counterfactual policy analysis. Be-
raja (2020) similarly forms counterfactuals without relying on particular parametric struc-
tural 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 relatively mild additional struc-
tural 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.⁶

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 char-
acterized 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

⁶Kocherlakota (2019) considers the problem of a policymaker trying to make the optimal decision *condi-
tional 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.

estimable objects. In contemporaneous and independent work, De Groot et al. (2021) and Hebden & Winker (2021) show how to use the same arguments to efficiently *compute* policy counterfactuals by generating impulse responses to policy shocks from a structural model. Our focus is not computational—we aim to calculate policy counterfactuals directly from empirical evidence, 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 systematic 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 in Section 2.1 by presenting a simplified version of our argument in the context of a stylized two-variable, two-period economy, and then in Sections 2.2 to 2.4 extend the logic to a general class of infinite-horizon linearized dynamic models.

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

2.1 A simple example

We begin with an illustration of our identification argument in the simplest possible setting: a two-variable, two-period economy.

MODEL. The two variables of our economy are a private-sector aggregate x_t (e.g., output) and a policy instrument z_t (e.g., the nominal interest rate). They are related through a pair of equations: the private-sector block summarizing optimal behavior of all agents (e.g., an

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

Euler equation) and the policy rule (e.g., a Taylor-type rule). These relations are

$$x_0 = \varepsilon_0 - \beta z_0 + \gamma x_1 \tag{1}$$

$$x_1 = \varepsilon_1 - \beta z_1 \tag{2}$$

as the private-sector block and

$$z_0 = \phi x_0 + \nu_0 \tag{3}$$

$$z_1 = \phi x_1 + \nu_1 \tag{4}$$

as the policy block. The vector $(\varepsilon_0, \varepsilon_1, \nu_0, \nu_1)$ collects exogenous “shocks”, i.e., departures from private-sector optimality and the systematic policy rule at times $t = 0$ and $t = 1$. While our economy features perfect foresight, the model equations may alternatively be interpreted as belonging to a linearized model with aggregate risk, where (1) - (3) hold at $t = 0$, while the $t = 1$ relations (2) and (4) only hold in time-0 *expectation*. This perspective reveals that the shocks ε_1 and ν_1 should be interpreted as news shocks: they may only be realized at $t = 1$, but they are already perfectly anticipated at $t = 0$. For the subsequent analysis, the crucial property of this simple model economy is that the coefficients in the private-sector equations (1) - (2) are independent of the policy rule—i.e., β and γ are unaffected by changes in ϕ . Equivalently, private-sector behavior is affected by policy only through the current and future values of the policy instrument z_t .

Given $(\varepsilon_0, \varepsilon_1, \nu_0, \nu_1)$, an equilibrium is a set (x_0, x_1, z_0, z_1) such that (1) - (4) all hold. We write the solution as $\mathbf{x}_\phi(\boldsymbol{\varepsilon}, \boldsymbol{\nu})$ and $\mathbf{z}_\phi(\boldsymbol{\varepsilon}, \boldsymbol{\nu})$ with boldface denoting a path, e.g., $\mathbf{x} \equiv (x_0, x_1)'$. For our analysis we will fix a particular vector of non-policy shocks $(\varepsilon_0, \varepsilon_1)$. If the policy rule (3) - (4) is followed perfectly (i.e., $\nu_0 = \nu_1 = 0$), then the response of the economy to this vector of shocks is simply given as the solution of the system (1) - (4) for a shock vector $(\varepsilon_0, \varepsilon_1, 0, 0)$. With slight abuse of notation we write these responses as $\mathbf{x}_\phi(\boldsymbol{\varepsilon})$ and $\mathbf{z}_\phi(\boldsymbol{\varepsilon})$.

OBJECT OF INTEREST. We wish to characterize the behavior of the economy in response to the non-policy shocks $(\varepsilon_0, \varepsilon_1)$ not under the baseline policy rule (3) - (4), but instead under some counterfactual alternative policy rule of the form

$$z_0 = \tilde{\phi} x_0 \tag{5}$$

$$z_1 = \tilde{\phi} x_1 \tag{6}$$

where $\tilde{\phi} \neq \phi$. Note that this thought experiment supposes that the private sector perfectly understands the change in rule: the new relation between z and x not only holds at $t = 0$, but is also known to hold at $t = 1$. In the analogous economy with aggregate risk, this model feature corresponds to the new rule being *expected* to hold at $t = 1$. Our object of interest is thus the ambitious question of the Lucas program. The identification result characterizes the information required to construct that counterfactual.

IDENTIFICATION RESULT. We can write an equilibrium of the model under the baseline policy rule as a mapping from exogenous shocks to endogenous outcomes:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} \Theta_{x,\varepsilon,\phi} \\ \Theta_{z,\varepsilon,\phi} \end{pmatrix} \times \boldsymbol{\varepsilon} + \begin{pmatrix} \Theta_{x,\nu,\phi} \\ \Theta_{z,\nu,\phi} \end{pmatrix} \times \boldsymbol{\nu} \quad (7)$$

The matrices $\Theta_{\bullet,\bullet,\phi}$ are 2×2 and contain the mapping from shocks (either $\boldsymbol{\varepsilon}$ or $\boldsymbol{\nu}$) to outcomes (either \mathbf{x} or \mathbf{z}) under the baseline rule ϕ —that is, the Θ 's collect the dynamic causal effects (or impulse responses) of the various exogenous shocks. Closed-form expressions for the entries of these matrices are provided in Appendix A.1. For our purposes here, it for now suffices to note that the impulse responses to the non-policy shock vector $\boldsymbol{\varepsilon}$ are given as

$$x_0(\boldsymbol{\varepsilon}) = \frac{1}{1 + \beta\phi} \left[\varepsilon_0 + \frac{\gamma}{1 + \beta\phi} \varepsilon_1 \right], \quad x_1(\boldsymbol{\varepsilon}) = \frac{1}{1 + \beta\phi} \varepsilon_1 \quad (8)$$

and

$$z_0(\boldsymbol{\varepsilon}) = \frac{\phi}{1 + \beta\phi} \left[\varepsilon_0 + \frac{\gamma}{1 + \beta\phi} \varepsilon_1 \right], \quad z_1(\boldsymbol{\varepsilon}) = \frac{\phi}{1 + \beta\phi} \varepsilon_1 \quad (9)$$

The (unknown) counterfactuals of interest are the impulse responses $\mathbf{x}_{\tilde{\phi}}(\boldsymbol{\varepsilon}) \equiv \Theta_{x,\varepsilon,\tilde{\phi}} \times \boldsymbol{\varepsilon}$ and $\mathbf{z}_{\tilde{\phi}}(\boldsymbol{\varepsilon}) \equiv \Theta_{z,\varepsilon,\tilde{\phi}} \times \boldsymbol{\varepsilon}$, derived instead under the alternative policy rule (5) - (6).

Our main identification result states that, to derive the desired counterfactual of interest, knowledge of three objects is sufficient: first, the impulse responses $\mathbf{x}_{\phi}(\boldsymbol{\varepsilon})$ and $\mathbf{z}_{\phi}(\boldsymbol{\varepsilon})$ to the non-policy shock $\boldsymbol{\varepsilon}$ under the baseline policy rule; and second and third, the impulse responses $\{\Theta_{x,\nu,\phi}, \Theta_{z,\nu,\phi}\}$ of x and z to contemporaneous and news shocks to the baseline policy rule. That is, knowledge of the effects of a menu of *shocks* under some prevailing policy rule is sufficient to predict the effects of a change in the systematic policy *rule*. To prove the result, we look for a vector of policy shocks $\tilde{\boldsymbol{\nu}}$ such that, if the economy under the baseline policy rule were to be subject to a combination of shocks $(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}})$, then the new policy rule (5) - (6) would hold everywhere along the equilibrium transition path. This policy shock vector must

thus satisfy

$$\mathbf{z}_\phi(\boldsymbol{\varepsilon}) + \Theta_{z,\nu,\phi} \times \tilde{\boldsymbol{\nu}} = \tilde{\phi} \times [\mathbf{x}_\phi(\boldsymbol{\varepsilon}) + \Theta_{x,\nu,\phi} \times \tilde{\boldsymbol{\nu}}], \quad (10)$$

a system of two equations in two unknowns, $\tilde{\nu}_0$ and $\tilde{\nu}_1$. Note that the information assumed in our identification result suffices to evaluate (10) and so solve for a $\tilde{\boldsymbol{\nu}}$ such that the relation holds.⁸ Furthermore, given $\boldsymbol{\varepsilon}$ and $\tilde{\boldsymbol{\nu}}$, our informational requirements also suffice to compute the impulse responses of x and z to the shocks $(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}})$ under the base rule ϕ . Relatively tedious algebra, presented in Appendix A.1, gives these impulse responses as

$$x_{0,\phi}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) = \frac{1}{1 + \beta\tilde{\phi}} \left[\varepsilon_0 + \frac{\gamma}{1 + \beta\tilde{\phi}} \varepsilon_1 \right], \quad x_{1,\phi}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) = \frac{1}{1 + \beta\tilde{\phi}} \varepsilon_1 \quad (11)$$

and

$$z_{0,\phi}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) = \frac{\tilde{\phi}}{1 + \beta\tilde{\phi}} \left[\varepsilon_0 + \frac{\gamma}{1 + \beta\tilde{\phi}} \varepsilon_1 \right], \quad z_{1,\phi}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) = \frac{\tilde{\phi}}{1 + \beta\tilde{\phi}} \varepsilon_1 \quad (12)$$

Comparing (8) - (9) and (11) - (12) yields the result: impulse responses to the combination of shocks $(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}})$ under the prevailing rule are identical to impulse responses to the non-policy shock $\boldsymbol{\varepsilon}$ alone but under the counterfactual policy rule. The intuition is simple. By construction, the policy shocks $\tilde{\boldsymbol{\nu}}$ are selected so that, in equilibrium, the new policy rule (5) - (6) holds for $t = 0, 1$. Private sector behavior, however, depends on the policy rule *only* to the extent that it affects the value of the policy instrument z_t . With z_t set exactly as in the equilibrium under the new policy rule, it is immediate that x_t will also take the same values as in that counterfactual equilibrium.

We emphasize that the above argument necessarily requires a judicious choice of *both* the contemporaneous policy shock ν_0 as well as the news policy shock ν_1 . For shocks to the baseline rule to mimic the contemplated counterfactual rule, that counterfactual rule needs to hold both today and (in expectation) tomorrow, which in general is only possible through a combination of contemporaneous and news shocks. If instead we were to only set ν_0 to some value $\check{\nu}_0$ to enforce the counterfactual policy rule at $t = 0$ (while keeping $\check{\nu}_1 = 0$), then we would recover the following solution:

$$x_{0,\phi}(\boldsymbol{\varepsilon}, \check{\boldsymbol{\nu}}) = \frac{1}{1 + \beta\tilde{\phi}} \left[\varepsilon_0 + \frac{\gamma}{1 + \beta\tilde{\phi}} \varepsilon_1 \right], \quad z_{0,\phi}(\boldsymbol{\varepsilon}, \check{\boldsymbol{\nu}}) = \frac{\tilde{\phi}}{1 + \beta\tilde{\phi}} \left[\varepsilon_0 + \frac{\gamma}{1 + \beta\tilde{\phi}} \varepsilon_1 \right] \quad (13)$$

Note that this solution differs from the true counterfactual precisely because the new rule

⁸Our subsequent discussion reveals that, generically, this system indeed has a unique solution for any $\tilde{\phi}$.

does not hold at $t = 1$, feeding back to the present through the forward-looking term γx_1 in the private-sector block. We will later relate the approach of just relying on contemporaneous shocks $\check{\nu}_0$ to existing practice in the semi-structural Sims program.

SUMMARY & OUTLOOK. Using a very simple model example, we in this section formalized the idea that knowledge of the dynamic causal effects of contemporaneous and news policy shocks suffices for structural policy counterfactual analysis. In particular, we argued that this information allows the researcher to map the observed effects of some non-policy disturbance under the prevailing rule into its hypothetical effects under various alternative rules.

The remainder of this section discusses the generality and limitations of this result, and furthermore extends it to optimal policy analysis. We also link the informational requirements to the econometric estimands of conventional semi-structural time series methods (e.g. Ramey, 2016) and in particular discuss the relation of our identification results to common practice in the empirical Sims (1980, 1982, 1987) program.

2.2 Model & objects of interest

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

The economy is summarized by the system

$$\mathcal{H}_w \mathbf{w} + \mathcal{H}_x \mathbf{x} + \mathcal{H}_z \mathbf{z} + \mathcal{H}_\varepsilon \boldsymbol{\varepsilon} = \mathbf{0} \quad (14)$$

$$\mathcal{A}_x \mathbf{x} + \mathcal{A}_z \mathbf{z} + \boldsymbol{\nu} = \mathbf{0} \quad (15)$$

w and x are n_w - and n_x -dimensional vectors of endogenous variables, z is a n_z -dimensional vector of policy instruments, ε is a n_ε -dimensional vector of exogenous structural shocks, and ν is an n_z -dimensional vector of policy shocks. The distinction between w and x is that the variables in x are observable while the variables in w are not; specifically, x contains the outcomes of interest to the econometrician and the arguments of the counterfactual policy rule. The infinite-dimensional linear maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ summarize the non-policy block of the economy, yielding $n_w + n_x$ restrictions for each t .⁹ Our key assumption—echoing the simple model of Section 2.1—is that the maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ do not depend on the

⁹The boldface vectors $\{\mathbf{w}, \mathbf{x}, \mathbf{z}, \boldsymbol{\varepsilon}, \boldsymbol{\nu}\}$ stack the time paths for all variables (e.g., $\mathbf{x} = (\mathbf{x}'_1, \dots, \mathbf{x}'_{n_x})'$), and the linear maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ are conformable.

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 . As in our simple example, we may equivalently interpret this system of equations as pertaining to a linearized economy with aggregate risk, with all equations for $t = 1, 2, \dots$ holding only in time-0 expectation. Entries of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}$ for $t = 1, 2, \dots$ should thus again be interpreted as news shocks. In particular, the policy shock vector $\boldsymbol{\nu}$ collects the full menu of contemporaneous and news shocks to the prevailing policy rule at all horizons, generalizing the two-shock set-up in our simple two-period model.

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

DISCUSSION & SCOPE. Our identification results in Section 2.3 and the empirical analysis in Sections 3 and 4 will apply to *any* structural model that can be written in the general form (14) - (15). In addition to linearity, the key property of the model for our purposes is that policy enters the non-policy block of the economy *only* through the realized path of the policy variables \boldsymbol{z} ; equivalently, in the linearized economy with aggregate risk, policy matters only through its effects on the expected future path of the instrument z . How restrictive are those assumptions?

Our first observation is that many of the explicit, parametric structural models used for counterfactual and optimal policy analysis in the 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. Our first

example is the canonical three-equation model of monetary policy analysis (see Galí, 2015). Translated to our perfect-foresight notation, we can write the non-policy block—that is, the Euler equation and the Phillips curve—as

$$\begin{aligned}\mathbf{y} &= -\sigma(\mathbf{i} - \boldsymbol{\pi}_{+1}) + \mathbf{y}_{+1} + \boldsymbol{\varepsilon}^d \\ \boldsymbol{\pi} &= \kappa\mathbf{y} + \beta\boldsymbol{\pi}_{+1} + \boldsymbol{\varepsilon}^s\end{aligned}$$

where $x = (\mathbf{y}, \boldsymbol{\pi})'$ collects the endogenous private-sector variables and the economy is subject to demand and supply shocks $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}^d, \boldsymbol{\varepsilon}^s)'$.¹⁰ The policy rule specifies the evolution of the policy instrument $z = i$:

$$\dot{\mathbf{i}} = \phi_\pi \boldsymbol{\pi} + \boldsymbol{\nu}$$

Importantly, and as required by our theory, the coefficients in the private-sector equations—here σ , κ and β —are independent of the policy rule. The exact same line of reasoning reveals that 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. For example, in standard HANK-type models, the simple Euler equation of representative-household models is replaced by a more general aggregate consumption function (e.g. Auclert et al., 2018):

$$\mathbf{y} = \mathcal{C}_y \mathbf{y} + \mathcal{C}_\pi \boldsymbol{\pi} + \mathcal{C}_i \mathbf{i} + \boldsymbol{\varepsilon}^d$$

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

While thus clearly quite general, our framework also has important limitations. First, since we leverage certainty equivalence of the linearized model economy, our identification results will generally not yield *globally* valid policy counterfactuals. Second, the policy

¹⁰Throughout, the subscripts +1 and −1 denote time paths shifted forward or backward one period, respectively. Appendix A.1 shows the specific linear maps that translate the three-equation model into our general form (14)-(15).

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 aggregate consumption function sketched above, changes in the long-run tax-and-transfer system will affect the coefficient matrices \mathcal{C}_\bullet , so such policies are necessarily outside the purview of our analysis. Third, some models—even after linearization—do not feature a separation of policy and non-policy blocks as in (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 A.2, the standard Lucas island model induces an aggregate supply relation of the form

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

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

OBJECTS OF INTEREST. We want to learn about two sets of policy rule counterfactuals.

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

$$\tilde{\mathcal{A}}_x \mathbf{x} + \tilde{\mathcal{A}}_z \mathbf{z} = \mathbf{0} \tag{16}$$

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

Assumption 2. *The policy rule in (16) induces a unique and determinate equilibrium. That is, the infinite-dimensional linear map*

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

is invertible.

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

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

$$\mathcal{L} = \sum_{i=1}^{n_x} \lambda_i \mathbf{x}_i' W \mathbf{x}_i \quad (17)$$

where i indexes the n_x distinct (observable) macroeconomic aggregates collected in \mathbf{x} , λ_i denotes policy weights, and W is a symmetric positive-definite matrix.¹¹ 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 \mathbf{z} to minimize the loss function (17) 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 $\mathbf{x}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$ and $\mathbf{z}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$?

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: shock 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

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

¹¹Our focus on a separable quadratic loss functions is in line with standard optimal policy analysis, but not essential. As shown in Appendix A.4, our results extend to the non-separable quadratic case, where the loss is now given by $\mathbf{x}'Q\mathbf{x}$ 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 \mathbf{w} , \mathbf{x} and \mathbf{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

$$\Theta_{\mathcal{A}} \equiv \begin{pmatrix} \Theta_{w,\varepsilon,\mathcal{A}} & \Theta_{w,\nu,\mathcal{A}} \\ \Theta_{x,\varepsilon,\mathcal{A}} & \Theta_{x,\nu,\mathcal{A}} \\ \Theta_{z,\varepsilon,\mathcal{A}} & \Theta_{z,\nu,\mathcal{A}} \end{pmatrix}. \quad (18)$$

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, we require knowledge of 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 rule and the policymaker loss). Furthermore, to construct counterfactual paths that correspond to a given non-policy shock sequence $\boldsymbol{\varepsilon}$, we also require knowledge of the dynamic causal effects of that particular shock sequence under the baseline policy rule, $\{\mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$. We emphasize that, in principle, all of these objects are estimable using data generated under the baseline policy rule: for example, given valid instrumental variables for all the distinct policy shocks $\boldsymbol{\nu}$ as well as a single instrument for the non-policy shock path $\boldsymbol{\varepsilon}$, the required entries of the Θ 's can be estimated using conventional semi-structural time-series methods (see Ramey, 2016, for a review).

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

$$\mathbf{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) = \mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) \equiv \mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{x,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}} \quad (19)$$

$$\mathbf{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) = \mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) \equiv \mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{z,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}} \quad (20)$$

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

$$\tilde{\mathcal{A}}_x [\mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{x,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}}] + \tilde{\mathcal{A}}_z [\mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{z,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}}] = \mathbf{0} \quad (21)$$

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

$$\begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \tilde{\mathcal{A}}_x & \tilde{\mathcal{A}}_z \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} -\mathcal{H}_\varepsilon \\ \mathbf{0} \end{pmatrix} \varepsilon \quad (22)$$

By Assumption 2 we know that (22) has a unique solution $\{\mathbf{x}_{\tilde{\mathcal{A}}}(\varepsilon), \mathbf{z}_{\tilde{\mathcal{A}}}(\varepsilon)\}$. To characterize this solution as a function of observables, suppose instead that we could find a $\tilde{\boldsymbol{\nu}}$ that solves (21). Since (14) also holds under the initial policy rule, and since (21) imposes the new policy rule, it follows that any $(\mathbf{x}_{\mathcal{A}}(\varepsilon, \tilde{\boldsymbol{\nu}}), \mathbf{z}_{\mathcal{A}}(\varepsilon, \tilde{\boldsymbol{\nu}}))$ with $\tilde{\boldsymbol{\nu}}$ solving (21) are also part of a solution of (22). Since by assumption (22) has a unique solution, it follows that the system (21) is solved by at most one $\tilde{\boldsymbol{\nu}}$.

It remains to establish that the system (21) has a solution. For this consider the candidate $\tilde{\boldsymbol{\nu}} = (\tilde{\mathcal{A}}_x - \mathcal{A}_x)\mathbf{x}_{\tilde{\mathcal{A}}}(\varepsilon) + (\tilde{\mathcal{A}}_z - \mathcal{A}_z)\mathbf{z}_{\tilde{\mathcal{A}}}(\varepsilon)$. Since the paths $\{\mathbf{w}_{\tilde{\mathcal{A}}}(\varepsilon), \mathbf{x}_{\tilde{\mathcal{A}}}(\varepsilon), \mathbf{z}_{\tilde{\mathcal{A}}}(\varepsilon)\}$ solve (22), it follows that they are also a solution to the system

$$\begin{pmatrix} \mathcal{H}_w & \mathcal{H}_x & \mathcal{H}_z \\ \mathbf{0} & \mathcal{A}_x & \mathcal{A}_z \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{x} \\ \mathbf{z} \end{pmatrix} = - \begin{pmatrix} \mathcal{H}_\varepsilon \varepsilon \\ (\tilde{\mathcal{A}}_x - \mathcal{A}_x)\mathbf{x}_{\tilde{\mathcal{A}}}(\varepsilon) + (\tilde{\mathcal{A}}_z - \mathcal{A}_z)\mathbf{z}_{\tilde{\mathcal{A}}}(\varepsilon) \end{pmatrix} \quad (23)$$

But by Assumption 1 we know that the system (23) has a unique solution, so indeed the paths $\{\mathbf{w}_{\tilde{\mathcal{A}}}(\varepsilon), \mathbf{x}_{\tilde{\mathcal{A}}}(\varepsilon), \mathbf{z}_{\tilde{\mathcal{A}}}(\varepsilon)\}$ are that solution. It then follows from the definition of $\Theta_{\mathcal{A}}$ in (18) that the candidate $\tilde{\boldsymbol{\nu}}$ also solves (21), completing the argument. \square

It follows from Proposition 1 that, as claimed, we can recover the desired counterfactual as a function of $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ and $\{\mathbf{x}_{\mathcal{A}}(\varepsilon), \mathbf{z}_{\mathcal{A}}(\varepsilon)\}$ alone. The key building block equation (21) is the infinite-horizon analogue of the system (10) from our simple two-period, two-variable example. 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 particular perturbation $\tilde{\boldsymbol{\nu}}$ 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 we now consider an infinite-horizon setting, we require evidence on contemporaneous and all possible future news shocks to the baseline rule in order to be able to mimic an arbitrary alternative policy rule.

b) The second identification result concerns optimal policy.

Proposition 2. *Consider a policymaker with loss function (17). For any $\boldsymbol{\varepsilon}$, the solution to the optimal policy problem is uniquely implemented by the rule $\{\mathcal{A}_x^*, \mathcal{A}_z^*\}$ with*

$$\mathcal{A}_x^* = \left(\lambda_1 \Theta'_{x_1, \nu, \mathcal{A}} W, \lambda_2 \Theta'_{x_2, \nu, \mathcal{A}} W, \dots, \lambda_{n_x} \Theta'_{x_{n_x}, \nu, \mathcal{A}} W \right), \quad (24)$$

$$\mathcal{A}_z^* = \mathbf{0}. \quad (25)$$

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

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

$$\mathcal{H}'_w(I \otimes W)\boldsymbol{\varphi} = \mathbf{0} \quad (26)$$

$$(\Lambda \otimes W)\mathbf{x} + \mathcal{H}'_x(I \otimes W)\boldsymbol{\varphi} = \mathbf{0} \quad (27)$$

$$\mathcal{H}'_z W\boldsymbol{\varphi} = \mathbf{0} \quad (28)$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots)$ and $\boldsymbol{\varphi}$ is the multiplier on (14). By Assumption 3 we know that the system (26) - (28) together with (14) has a unique solution $\{\mathbf{x}^*(\boldsymbol{\varepsilon}), \mathbf{z}^*(\boldsymbol{\varepsilon}), \boldsymbol{\varphi}^*(\boldsymbol{\varepsilon})\}$.

Now consider the alternative problem of choosing deviations $\boldsymbol{\nu}^*$ from the prevailing rule to minimize (17) subject to (14) - (15). This second problem gives the first-order conditions

$$\mathcal{H}'_w(I \otimes W)\boldsymbol{\varphi} = \mathbf{0} \quad (29)$$

$$(\Lambda \otimes W)\mathbf{x} + \mathcal{H}'_x(I \otimes W)\boldsymbol{\varphi} + \mathcal{A}'_x W\boldsymbol{\varphi}_z = \mathbf{0} \quad (30)$$

$$\mathcal{H}'_z(I \otimes W)\boldsymbol{\varphi} + \mathcal{A}'_z W\boldsymbol{\varphi}_z = \mathbf{0} \quad (31)$$

$$W\boldsymbol{\varphi}_z = \mathbf{0} \quad (32)$$

where $\boldsymbol{\varphi}_z$ is the multiplier on (15). It follows from (32) that $\boldsymbol{\varphi}_z = \mathbf{0}$. But then (29) - (31) 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

$$\begin{pmatrix} \mathbf{w} \\ \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \Theta_{\mathcal{A}} \times \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\nu}^* \end{pmatrix} \quad (33)$$

The problem of minimizing (17) subject to (33) gives the optimality condition

$$\sum_{i=1}^{n_x} \lambda_i \Theta'_{x_i, \nu, \mathcal{A}} W \mathbf{x}_i = 0 \quad (34)$$

By the equivalence of the policy problems, it follows that (34) is an optimal policy *rule*, taking the form (24) - (25). Finally, the second part of the result follows from Proposition 1 since (34) is just a special example of a policy rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$.

□

Proposition 2 reveals that, in conjunction with a given policymaker loss function, the information required to construct counterfactuals for arbitrary policy rules also suffices to characterize *optimal* policy rules.¹² The intuition is exactly as before: since we know the causal effects of every possible policy perturbation $\boldsymbol{\nu}$ on the policymaker targets \mathbf{x} , we in particular know the space of those targets that is implementable through policy actions. At an optimum, we must be at the point of this space that minimizes the policymaker loss. As before, it does not matter whether this optimum is attained through some systematic policy rule or through shocks to an alternative rule.

DISCUSSION. 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 that are 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*.

¹²Note 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):

$$\sum_{i=1}^{n_x} \lambda_i \Theta'_{x_i, \nu, \mathcal{A}} W \mathbb{E}_t [\mathbf{x}_i] = 0 \quad (35)$$

where now $\mathbf{x}_i = (x_{it}, x_{it+1}, \dots)'$. 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, (35) must apply to *revisions* of policymaker expectations at each t .

Traditionally, the Sims program has used a different definition of counterfactuals from that in the Lucas program. The Sims program is motivated by the view that some policy actions are not interpreted by the private sector as changes in policy regime and therefore do not lead to important changes in expectations (Sims, 1982). Given this motivation, counterfactuals are constructed by selecting a sequence of *contemporaneous* policy shocks to enforce the counterfactual policy rule along the equilibrium transition path (e.g. Sims & Zha, 2006; Bernanke et al., 1997; Eberly et al., 2020). Discussions of the credibility of this approach focus on whether the private sector would plausibly interpret the required shocks as typical fluctuations in policy and not suspect a change in regime—for example the shocks are not too large (Leeper & Zha, 2003). Our approach differs in that we are using the Lucas program’s definition of the counterfactual in which the private sector is assumed to understand the change in policy regime. We are able to adopt this perspective because we use the policy news shocks to account for the effects of the policy on expectations.

2.4 Visual illustration

We now provide a visual illustration of our results. As our laboratory we use the structural HANK model of Wolf (2021), with details of the model parameterization relegated to Appendix A.1. In this environment we compute policy counterfactuals in two ways: first by just resolving the model with different counterfactual policy rules (as in the Lucas program), and second by leveraging our identification results. To do so, we begin by solving the model with a baseline policy rule of

$$i_t = \phi_\pi \pi_t \tag{36}$$

for $\phi_\pi = 1.5$. Using this solution we then compute the policy causal effect maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ as well as the impulse responses $\{\mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$ to a contractionary cost-push shock ε^s . We emphasize that those causal effects would be estimable for an econometrician living in our model laboratory and with access to valid instrumental variables for the cost-push shock ε^s as well as the policy shocks $\{\nu_0, \nu_1, \dots\}$. Next, 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

$$i_t = \phi_i i_{t-1} + (1 - \phi_i)(\phi_\pi \pi_t + \phi_y y_t) \tag{37}$$

ALTERNATIVE POLICY RULE, HANK MODEL

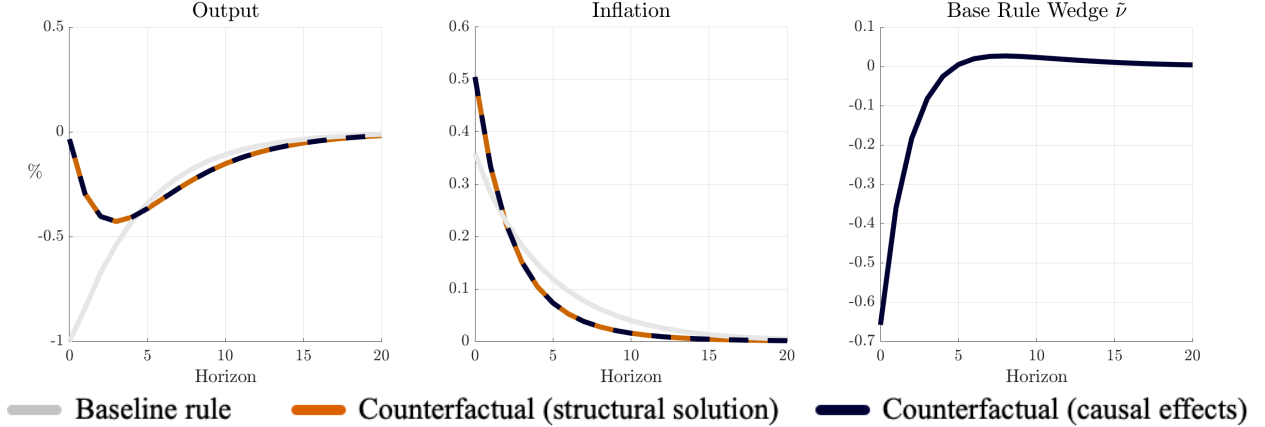


Figure 1: Output and inflation impulse responses together with the equivalence shock wedge $\tilde{\nu}$ (see (21)) for the HANK model with policy rules (36) and (37). The impact output contraction under the prevailing baseline rule is normalized to -1% .

for $\phi_i = 0.9$, $\phi_\pi = 2$ and $\phi_y = 0.5$. As indicated above, we first of all do so by making 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 any knowledge of the structural equations of the model. We do so using $\{\mathbf{x}_A(\boldsymbol{\varepsilon}), \mathbf{z}_A(\boldsymbol{\varepsilon})\}$ and $\{\Theta_{x,\nu,A}, \Theta_{z,\nu,A}\}$ —the dynamic causal effects of the fundamental shock and of policy shocks generated under the prevailing baseline rule (36). We feed these inputs into (19)-(21) to solve for \mathbf{x} , \mathbf{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 $\tilde{\nu}$ 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 standard dual mandate loss function

$$\mathcal{L} = \lambda_\pi \boldsymbol{\pi}' \boldsymbol{\pi} + \lambda_y \mathbf{y}' \mathbf{y} \quad (38)$$

with $\lambda_\pi = \lambda_y = 1$. We again start by solving for the optimal policy using conventional

OPTIMAL POLICY, HANK MODEL

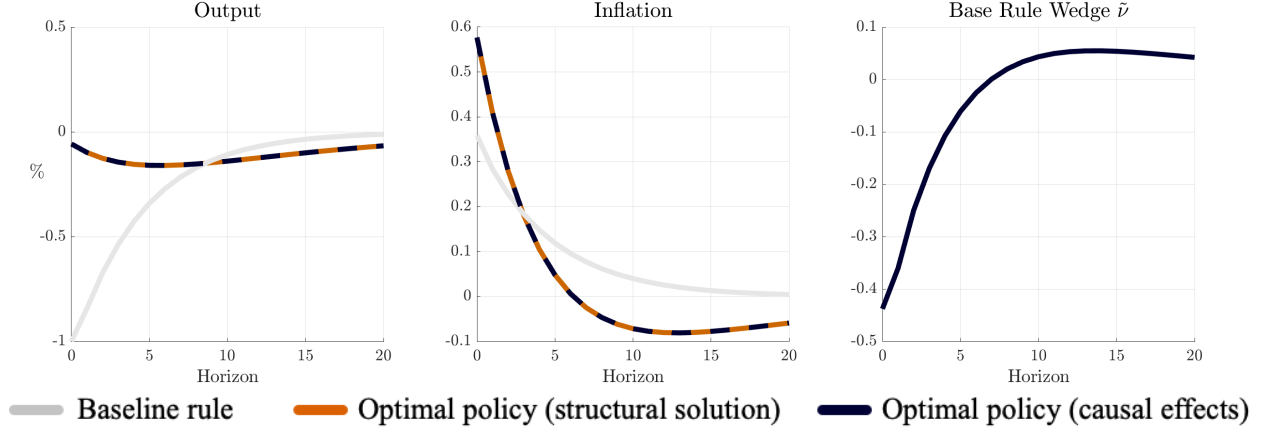


Figure 2: Output and inflation impulse responses together with the equivalence shock wedge $\tilde{\nu}$ (see (21)) for the HANK model with policy rules (36) and the optimal policy given by (38). The impact output contraction under the prevailing baseline rule is normalized to -1% .

methods: we derive the policy rule corresponding to the first-order conditions (26) - (28), 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 (36) 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 (34), 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 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 (19)-(21), again requiring only knowledge of the causal effects of policy shocks as well as the impulse responses to the cost-push shock under the baseline policy rule. As expected, the resulting impulse responses—the dark blue lines—are identical to those obtained by explicitly solving the optimal policy problem. Finally, the right panel of Figure 2 shows the optimal policy

as a deviation $\tilde{\nu}$ from the prevailing baseline rule. The optimal rule accommodates the inflationary cost-push shock more than the baseline rule (36), so again the required policy “shock” is persistently negative (i.e., expansionary).

A striking feature of both counterfactuals is that the required policy rule shock wedge $\tilde{\nu}$ is relatively transitory. This is unsurprising: if the underlying shock is rather transitory, then differences between the new and old rules will be transitory, and so the mapping between policy rules will mostly rely on knowledge of the causal effects of short-run shocks. This observation will turn out to be key for our empirical applications in Sections 3 and 4.

COMPARISON WITH SIMS PROGRAM. To further illustrate the logic of our identification argument, in Appendix Figure A.1 we report counterfactuals for the alternative policy rule (37) computed not using (19)-(21) (as in Figure 1), but instead using a sequence of *contemporaneous* shocks that enforce (37) along the entire equilibrium transition path as has traditionally been done in the Sims program. The figure shows that the two counterfactuals differ substantially due to their different treatments of expectations of future policy.

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 fact, as we discuss in Appendix A.3, our identification results can even be extended to non-linear models with aggregate risk; as we discuss there, the main change is that our informational requirements increase even further, with the required causal effects of policy shocks now additionally indexed by the state of the economy as well as the magnitude of the policy intervention.

In the remainder of this paper we present ways to operationalize our insights. We will throughout focus on our headline identification result for linearized models, simply because those informational requirements are already quite high: we need evidence on the dynamic causal effects of the full menu of contemporaneous and news policy shocks *at all horizons*—evidence that is not available, for any policy instrument. Sections 3 and 4 present two ways of dealing with that challenge while using a realistic amount of information on policy shocks.

3 Counterfactuals in identified subspaces

This section presents our first approach. While researchers may not be able to estimate the effects of contemporaneous and news policy shocks *at all possible horizons*, they can often do so for several *particular* shock paths. Such limited evidence is of course not enough to construct counterfactuals for all possible alternative counterfactual rules, but it may suffice for the counterfactual of interest, or at least a close approximation to it. This approach is a natural generalization of the counterfactuals constructed in the Sims program: rather than using a sequence of unanticipated policy shocks to enforce the counterfactual policy rule *ex post*, the researcher instead uses combinations of several distinct policy shocks to enforce the counterfactual rule as well as possible both *ex post* and *ex ante*, in private-sector expectations. The limitation of our approach is that it only applies to particular counterfactuals, but the appeal is that it allows the researcher to remain agnostic about the deep structure of the economy without running afoul of the Lucas critique.

In Section 3.1, we extend our identification results to the empirically relevant case of a researcher that only observes the dynamic causal effects of a finite (small) number of shocks to the policy rule. In Section 3.2, we apply our results to monetary policy counterfactuals.

3.1 Constrained identification results

Empirical researchers have relied on several different pieces of identifying information to estimate the dynamic causal effects of policy shocks (e.g. see Ramey, 2016, for a review). For example, the monetary policy shock literature has identified the causal effects associated with different paths of nominal interest rates (e.g. Romer & Romer, 2004; Gertler & Karadi, 2015), while the fiscal policy literature has studied both transitory as well as persistent changes in government purchases (e.g. Ramey, 2011). Anticipating our main empirical application, Figure 3 provides an illustration for monetary policy shocks, showing interest rate paths for two possible identified shocks: the left panel corresponding to a transitory rate hike, and the right panel showing a more gradual change.

What is the connection between such empirical evidence and the informational requirements of our identification results? The theoretical discussion of contemporaneous and news policy shocks in Section 2 was phrased in terms of policy shocks ν that perturb the prevailing policy rule horizon by horizon. Ultimately, however, what matters for our results is that we are able to match the current and expected future path of the policy instrument that is associated with a given counterfactual rule, and not what kind of underlying shock path ν this

IDENTIFIED POLICY SHOCK PATHS, ILLUSTRATION

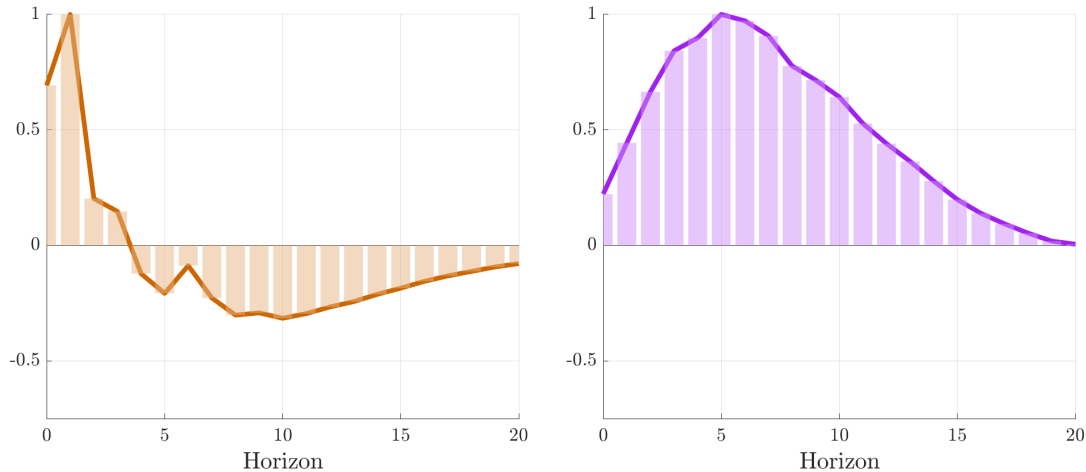


Figure 3: Two possible instrument paths $\mathbf{z}(\boldsymbol{\nu}_s)$ corresponding to two different shock paths $\boldsymbol{\nu}_s$, $s = 1, 2$: a short-lived change (orange, left panel) and a gradual, persistent departure from the rule (purple, right panel).

counterfactual path corresponds to. Viewed in this light, our identification results require the researcher to know the dynamic causal effects associated with every possible path of the policy instrument.¹³ The connection to prior empirical work is now clear: existing studies give us the dynamic causal effects associated with *particular* paths of the policy instrument. For example, in Figure 3, the shaded areas indicate the weights corresponding to the deviation of the policy instrument from its baseline value at each horizon. The remainder of this section asks how such evidence for particular policy instrument paths can be used to arrive at restricted versions of our theoretical identification results.

COUNTERFACTUAL POLICY RULES. We consider a researcher that has access to the dynamic causal effects associated with n_s distinct paths of the policy instrument z . We denote those causal effects by $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$, where the columns of the Ω 's correspond to the n_s distinct identified shocks.¹⁴ Given such lower-dimensional causal effect maps, the proof of

¹³Formally, what we are discussing here is nothing but a change of basis: we solve for the counterfactual not in terms of shocks to some (arbitrary) baseline rule, but directly in terms of policy instrument paths. This switch of basis is without loss of generality as long as the policymaker can implement any possible path of the policy instrument (i.e., the map $\Theta_{z,\nu,\mathcal{A}}$ is invertible).

¹⁴Our 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

Proposition 1 now only works for *particular* pairs of counterfactual policy rules $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$ and non-policy shock paths $\boldsymbol{\varepsilon}$ —those that satisfy the restriction

$$\tilde{\mathcal{A}}_x(\mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{x,\mathcal{A}} \times \mathbf{s}) + \tilde{\mathcal{A}}_z(\mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{z,\mathcal{A}} \times \mathbf{s}) = \mathbf{0} \quad (39)$$

for some linear combination of the identified policy shocks with weights $\mathbf{s} \in \mathbb{R}^{n_s}$. In words, (39) says that the contemplated counterfactual rule must deviate from the prevailing one in response to shocks $\boldsymbol{\varepsilon}$ in a direction that is consistent with the causal effects corresponding to the available policy shocks. Naturally, the larger n_s , the larger the sets of non-policy shock paths and counterfactual policy rules that can be replicated exactly.

Given a counterfactual rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$ and non-policy shock $\boldsymbol{\varepsilon}$, it will not always be possible to find a vector \mathbf{s} such that (39) holds exactly. There are two ways of dealing with this challenge, each involving an approximation in a different place. First, fixing the impulse responses to the non-policy shock $\boldsymbol{\varepsilon}$, the researcher may select the shock weights \mathbf{s} so that the counterfactual rule holds as well as possible; that is, they solve the problem

$$\min_{\mathbf{s}} \quad \|\tilde{\mathcal{A}}_x(\mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{x,\mathcal{A}} \times \mathbf{s}) + \tilde{\mathcal{A}}_z(\mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Omega_{z,\mathcal{A}} \times \mathbf{s})\|. \quad (40)$$

Second, subject to the constraint that the counterfactual policy rule holds perfectly, the researcher may find shock impulse responses $\{\mathbf{x}^\dagger, \mathbf{z}^\dagger\}$ that are as similar as possible to those to $\boldsymbol{\varepsilon}$; that is, they solve

$$\min_{\mathbf{x}^\dagger, \mathbf{z}^\dagger, \mathbf{s}} \quad \|\mathbf{x}^\dagger - \mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon})\| + \|\mathbf{z}^\dagger - \mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\| \quad (41)$$

such that

$$\tilde{\mathcal{A}}_x(\mathbf{x}^\dagger + \Omega_{x,\mathcal{A}} \times \mathbf{s}) + \tilde{\mathcal{A}}_z(\mathbf{z}^\dagger + \Omega_{z,\mathcal{A}} \times \mathbf{s}) = \mathbf{0}.$$

Following either of these strategies, researchers can judge on a case-by-case basis whether our identification results allow them to provide an accurate approximation to the systematic policy rule counterfactual that they are interested in. We provide computational details for both approaches in Appendix B.4.

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 (17) within

sufficiently short-lived dynamics, making the maximal truncation horizon immaterial.

the identified n_s -dimensional subspace of policy changes. This problem gives the optimality condition

$$\sum_{i=1}^{n_x} \lambda_i \Omega'_{x_i, \mathcal{A}} W \mathbf{x}_i = 0 \quad (42)$$

(42) can be interpreted in two ways. First, it gives n_s restrictions that *any* solution to the full optimal policy problem must satisfy.¹⁵ 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 $n_s \rightarrow \infty$, (42) fully characterizes the optimal policy rule.

3.2 Application

We illustrate the identified subspace approach with an application to investment-specific technology shocks. Our object of interest is the behavior of the aggregate output gap, inflation, and interest rates following such a technology shock but under different counterfactual monetary policy rules. We present the main results here, and relegate further empirical implementation details to Appendix B.2.

INPUTS. The first input to our analysis are the aggregate effects of the non-policy shock of interest ε under the prevailing baseline policy rule. To recover those effects we rely on the investment-specific technology shock series identified by Ben Zeev & Khan (2015). This shock corresponds to a short-lived, unanticipated change in the relative price of investment goods. Importantly, technology shocks of this sort have been widely argued to be important drivers of aggregate cyclical fluctuations. As such, their effects under possible counterfactual policy rules are of particular interest. We estimate the propagation of this shock by ordering it first in a recursive Vector Autoregression (VAR) (Plagborg-Møller & Wolf, 2021).

The second input are the dynamic causal effects of some—ideally rich—menu of different monetary policy shocks. For this we consider two of the most popular examples of such monetary shocks: the shock series of Romer & Romer (2004) and Gertler & Karadi (2015). The dynamic responses of interest rates differ quite substantially across those two identifications schemes: rather short-lived for Romer & Romer, and gradual for Gertler & Karadi. Indeed, in our illustrative figure from before (Figure 3), the left panel corresponds to the interest

¹⁵Equation (42) 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.

POLICY COUNTERFACTUAL FOR INVESTMENT SHOCKS, IDENTIFIED SUBSPACE

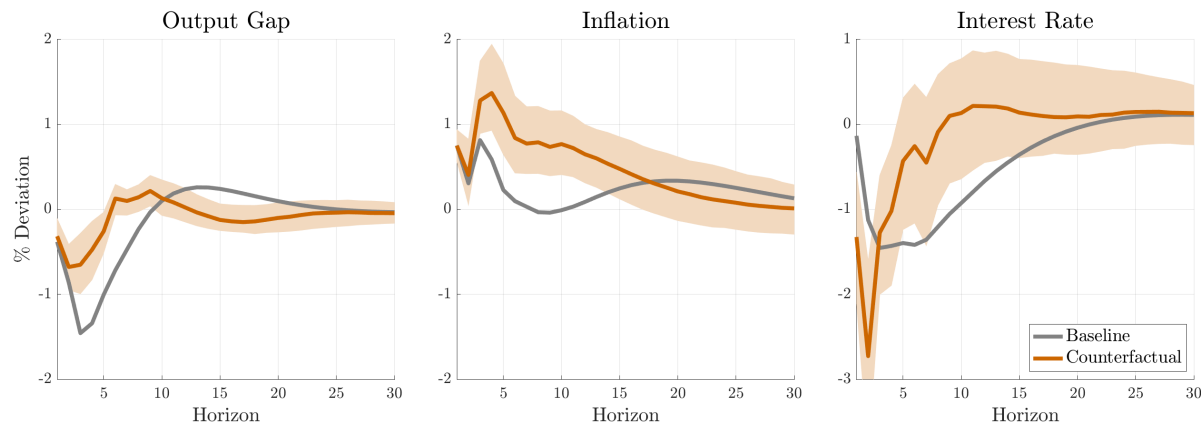


Figure 4: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (grey) and the best feasible approximation to a rule that stabilizes output (orange), computed following (40). The shaded areas correspond to 16th and 84th percentile confidence bands.

rate path identified by the Romer & Romer shock, while the right one presents the Gertler & Karadi shock path. To correctly account for joint uncertainty in the estimation of the two shocks, we study their propagation through a single VAR.

COUNTERFACTUAL POLICY RULES. We wish to predict the dynamics of the output gap, inflation and interest rates in response to a contractionary investment technology shock, but under a counterfactual policy rule that sets the output gap to zero at all times. Figure 4 presents our results. The identified investment technology shock has both a cost-push as well as a negative demand component, consistent with theory (e.g. Justiniano et al., 2010). Under the baseline policy rule (grey), interest rates are cut relatively aggressively, though by not enough to stabilize the output gap; furthermore inflation stays moderately above target. Under our best approximation to the counterfactual rule (orange) computed from (40), interest rates are instead cut much more aggressively, essentially stabilizing the output gap from around one year after the shock, at the cost of higher inflation.

The counterfactual analysis in Figure 4 gives an example of a pair of counterfactual policy rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$ and non-policy shock ε that is almost perfectly contained in the identified subspace for horizons greater than a year. Given the well-documented lags in monetary policy transmission, it is furthermore rather unlikely that *any* monetary intervention could perfectly close the output gap at short horizons, so our counterfactual is likely a close approximation

OPTIMAL POLICY FOR INVESTMENT SHOCKS, IDENTIFIED SUBSPACE

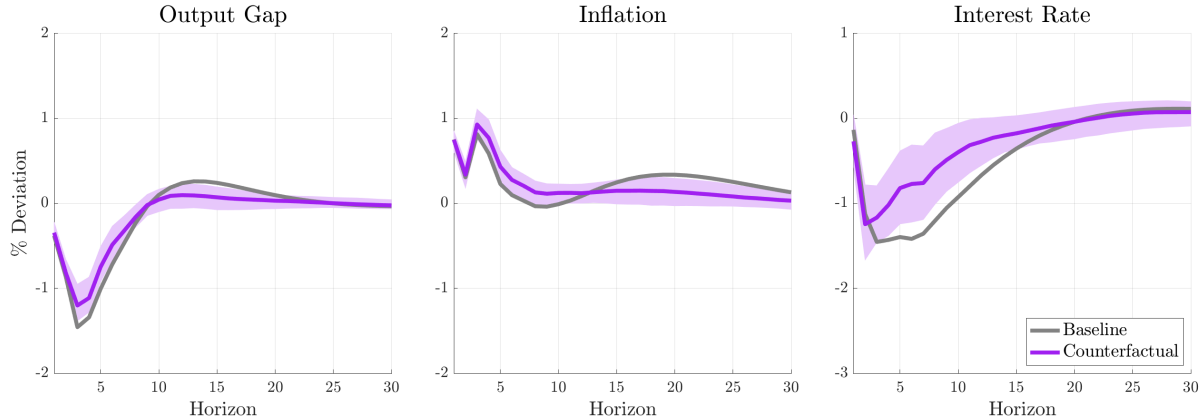


Figure 5: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (grey) and the optimal policy rule for a policymaker with preferences over current output and averaged past inflation (see Appendix A.4) within the identified subspace (purple). The shaded areas correspond to 16th and 84th percentile confidence bands.

to a rule that closes the output at all times as well as possible.¹⁶

By our identification results, any structural model of the general form (14) - (15) will necessarily agree with these predicted counterfactual effects if it is consistent with our off-the-shelf evidence on policy shocks. Specifically, if the model is consistent with our empirical estimates of the effects of monetary policy shocks and the counterfactual policy rule implies the alternative path of policy shown in the right panel of Figure 4, then the output gap and inflation counterfactuals will necessarily be those shown in Figure 4.

OPTIMAL POLICY. We consider a policymaker whose loss function puts equal weight on the output gap as well as 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 (42) 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 policymaker targets.

Figure 5 presents the results. It reveals that—at least for the policymaker preferences that we assume—the observed output gap and inflation paths are very close to the con-

¹⁶For further reference, we in Appendix B.2 display the best approximation to the shock paths $\{\mathbf{x}(\boldsymbol{\varepsilon}), \mathbf{z}(\boldsymbol{\varepsilon})\}$ that can be perfectly offset using our identified monetary policy shocks, computed following (41).

¹⁷See Appendix A.4 for a detailed discussion of the loss function.

strained optimum: given our empirical knowledge of the inflation-output allocations that are implementable through policy actions, there was little reason to set the policy instrument differently. We again emphasize that any structural model that fits into the framework (14) - (15) and is consistent with our empirical monetary policy estimates would necessarily agree with this normative conclusion.

SUMMARY. Our analysis in this section has revealed through examples that existing empirical evidence on distinct policy shock paths is already rich enough to study particular sets of structural policy rule counterfactuals without running afoul of the Lucas critique. On the other hand, the results in this section are approximations to the counterfactuals that are constrained by the available empirical evidence on policy shocks. One way forward is to seek new sources of evidence on policy shocks that expand the space of counterfactuals that we can construct. In the next section we consider another way forward: using additional structural assumptions to fill in the missing information on policy news shocks.

4 Policy shock causal effects as “identified moments”

In Section 3 we discussed the kinds of counterfactuals that can be constructed through the (finite-dimensional) subspace of estimated policy causal effects. If researchers are interested in counterfactuals that are (far) outside of this subspace, then they will need to rely on additional structural assumptions—typically in the form of fully specified models—to extrapolate from the causal effects of the policy shocks that we did observe to those that we did not. The typical Lucas program example of this approach is model estimation through impulse response matching (Rotemberg & Woodford, 1997; Christiano et al., 2005).

In this section we ask how “typical” structural models connect the dynamic causal effects of policy shocks across different horizons. Are the causal effects of different policy (news) shocks largely independent objects? Or do standard models imply tight restrictions across those responses? Economic intuition suggests the latter: for example, we may expect impulse responses to forward guidance about interest rates at a horizon of three years to be highly informative about the response to news about interest rates at a horizon of two years or four years. In this section we provide two perspectives on this question. The takeaway from both is that, at least through the lens of canonical macro models, individual policy shock impulse responses are powerful “identified moments” (in the language of Nakamura & Steinsson, 2018) for the *universe* of possible systematic policy rule counterfactuals.

4.1 Extrapolation with a partial model

This section presents our first insight: for a special but very important family of systematic policy rule counterfactuals, and for a large family of structural models, we can prove that the entire map of dynamic causal effects of policy shocks needed for our identification results lives in an extremely tightly restricted space. We use this observation to revisit our application in Section 3.2.

COUNTERFACTUALS OF INTEREST. We consider a researcher interested in the behavior of the aggregate output gap and inflation under counterfactual policy rules of the form

$$\tilde{\mathcal{A}}_\pi \pi + \tilde{\mathcal{A}}_y \mathbf{y} = \mathbf{0} \tag{43}$$

Note that (43) nests contemporaneous as well as average inflation targeting, nominal GDP targeting, as well as strict output gap and inflation stabilization. Furthermore, the optimal policy for a policymaker with a “dual mandate” loss function (38) yields a policy rule of the form (43). Counterfactual rules of the sort (43) are thus of substantial interest, and in particular contain as special cases the policy counterfactual studied in Section 3.2.

By our results in Section 2, knowledge of the two dynamic causal effect maps $\Theta_{\pi,\nu,\mathcal{A}}$ and $\Theta_{y,\nu,\mathcal{A}}$ is sufficient to construct counterfactuals for alternative policy rules like (43). More precisely, we in fact only require knowledge of *relative* policy shock impulses: if $\Theta_{\pi,\nu,\mathcal{A}}$ is invertible, then the proofs of our two identification results in Section 2 apply without change using only knowledge of $\Theta_{y,\nu,\mathcal{A}} \times \Theta_{\pi,\nu,\mathcal{A}}^{-1}$. Intuitively, for counterfactual rules of the form (43), we can effectively treat inflation as the policy instrument and then use the relative (or normalized) causal effects $\Theta_{y,\nu,\mathcal{A}} \times \Theta_{\pi,\nu,\mathcal{A}}^{-1}$ to determine the output path associated with a given inflation path.¹⁸ Our key insight is that, for any structural model that features a Phillips curve relationship between y and π , that Phillips curve fully pins down the required relative causal effects independently of the rest of the model.

EXTRAPOLATING SHOCK IMPULSE RESPONSES VIA DYNAMIC PHILLIPS CURVES. Consider a structural model that features a Phillips curve relationship—that is, a structural link between inflation and leads and lags of the output gap, with policy shocks moving the

¹⁸The assumption that the policymaker can implement any desired path of inflation is generally satisfied in standard business-cycle models. For example, in the simple model of Section 2.1, it is straightforward to verify that $\Theta_{\pi,\nu,\mathcal{A}}$ is an upper-triangular, invertible matrix. We provide further details in Appendix C.1.

economy along this function. Using our perfect-foresight notation of Section 2, we can write a general Phillips curve relationship as

$$\boldsymbol{\pi} = \Pi_y \times \mathbf{y} + \Pi_\varepsilon \times \boldsymbol{\varepsilon}. \quad (44)$$

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

$$\Pi_y = \begin{pmatrix} \kappa & \kappa\beta & \kappa\beta^2 & \dots \\ 0 & \kappa & \kappa\beta & \dots \\ 0 & 0 & \kappa & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (45)$$

The crucial implication of (44) is that, conditional on policy shocks $\boldsymbol{\nu}$, the co-movements of output and inflation are fully characterized by the map Π_y :

$$\Theta_{\pi,\nu,\mathcal{A}} = \Pi_y \times \Theta_{y,\nu,\mathcal{A}}. \quad (46)$$

In words, we can map output gap impulse responses into inflation impulse responses (and of course vice-versa) using only the matrix Π_y . Knowledge of Π_y is thus sufficient to construct counterfactuals for alternative policy rules of the general form (43), exactly as required. We can thus conclude that typical structural models force the rich dynamic causal effects that matter for many important policy counterfactual experiments to actually live in a very low-dimensional space—the relative output-inflation impulse response space induced only by the model’s Phillips curve. This observation suggests a refinement of the conventional impulse response matching approach to counterfactual policy evaluation: it suffices to use identified policy shocks to estimate one partial model block. The resulting counterfactuals will be independent of all further details of the model—including preferences, technology, the nature of household expectation formation, and so on.

APPLICATION. We provide an illustration of these insights by returning to our policy counterfactual experiments from Section 3.2. Differently from that analysis, our assumptions on Π_y will allow us to construct exact counterfactuals in line with Propositions 1 and 2, rather than being limited to the identified subspaces of causal effects.

POLICY COUNTERFACTUALS FOR INVESTMENT SHOCKS, PC EXTRAPOLATION

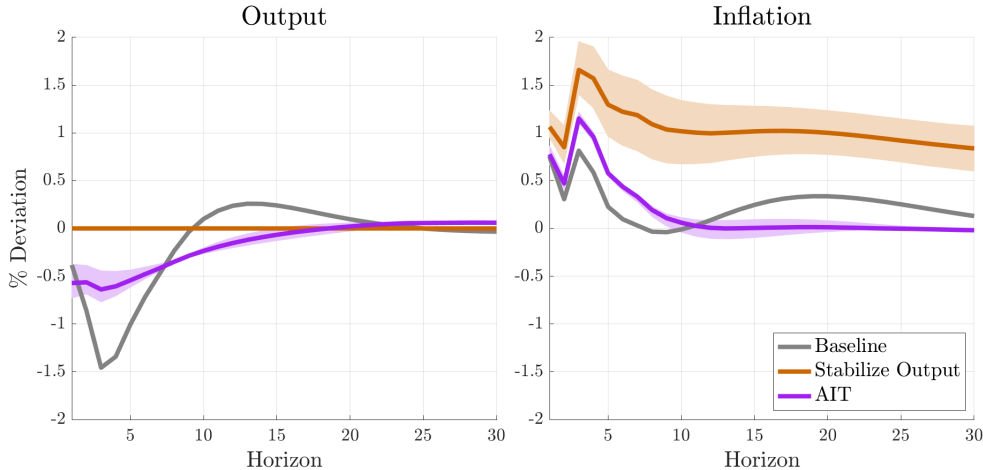


Figure 6: Output gap and inflation impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (grey), a counterfactual rule that perfectly stabilizes output (grey), and the optimal policy rule for a dual-mandate policymaker with preferences over current output and averaged past inflation (see Appendix B.3) (purple). The shaded areas correspond to 16th and 84th percentile confidence bands.

We assume that Π_y is derived from a hybrid Phillips curve relationship:

$$\pi_t = \gamma_b \pi_{t-1}^4 + \gamma_f \mathbb{E}_t [\pi_{t+4}^4] + \kappa y_t + \varepsilon_t \quad (47)$$

where $\pi_{t-1}^4 = \frac{1}{4} \times (\pi_{t-1} + \pi_{t-2} + \pi_{t-3} + \pi_{t-4})$. Appendix C.1 shows the linear map Π_y corresponding to this Phillips curve specification. We then estimate the parameters $\{\gamma_b, \gamma_f, \kappa\}$ (and so all of Π_y) using evidence on identified monetary policy shocks. The econometric challenge is that the estimated policy causal effects $\{\Omega_{\pi, \mathcal{A}}, \Omega_{y, \mathcal{A}}\}$ will not perfectly align with the parametric structure imposed by (47); thus, following Barnichon & Mesters (2020), we simply find the best possible fit. Our estimation uses the identified monetary policy shocks of Romer & Romer (2004), already discussed in Section 3.

Given an estimate of Π_y , we can construct the two desired counterfactuals: output gap and inflation impulse responses to investment-specific technology shocks under counterfactual policy rules that a) perfectly stabilize output and b) are optimal for a dual mandate policymaker that assigns equal weights on aggregate output and an average of current and lagged inflation. The results are reported in Figure 6. First, perfect output gap stabilization implies persistently elevated inflation relative to the baseline rule outcome, even moreso

than in the identified subspace estimation of Section 3.¹⁹ While the inflation counterfactual is similar on impact, inflation at longer horizons remains more elevated, largely due to the strong backward-looking component in our estimated parametric Phillips curve.²⁰ 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 (47), we can by Proposition 2 in fact explicitly characterize the full optimal policy rule as

$$\lambda_\pi \bar{\Pi}' \bar{\boldsymbol{\pi}} + \lambda_y (\Pi'_y)^{-1} \mathbf{y} = \mathbf{0} \quad (48)$$

where $\bar{\boldsymbol{\pi}}$ 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 B.3), and Π_y is displayed in Appendix C.1. We note that (48) takes the form of an implicit targeting rule (Svensson, 1997): it imposes a set of restrictions that current, lagged and expected future values of inflation and the output gap must satisfy at all times when policy is set optimally.

DISCUSSION & LIMITATIONS. 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 (47) and (iii) is consistent with the empirical monetary policy shock estimates of Romer & Romer (2004) will invariably produce the same counterfactuals as in Figure 6, and in particular yield the optimal policy rule (48).

The limitations of our analysis in this section are twofold. First, even for counterfactual policy rules that fall into our category (43), our approach only yields inflation and output counterfactuals, but is silent on the corresponding path for the policy instrument (e.g., the nominal rate of interest). Second, the approach does not apply to counterfactuals outside of that category; for example, if the policymaker contemplates an alternative rule that explicitly specifies the policy instrument as a function of observables, then the response map for the instrument itself ($\Theta_{z,\nu,\mathcal{A}}$) also matters. The next section—which develops a general measure of the informativeness of a given policy shock causal effect for all other policy shocks—speaks to these more challenging counterfactuals.

¹⁹Note that the assumed invertibility of $\Theta_{\pi,\nu,\mathcal{A}}$ together with the invertible Π_y implies invertibility of $\Theta_{y,\nu,\mathcal{A}}$ via (46), so perfect output gap stabilization is implementable.

²⁰In fact, it can be shown that the backward-looking component is large enough to imply that, for perfect output stabilization, inflation dynamics are non-stationary.

4.2 Extrapolation with a full model

If the counterfactual of interest is neither contained in the empirically identified subspace of policy shocks (as in Section 3) nor covered by the particular set of counterfactuals studied in Section 4.1, then researchers will need to specify a full structural model—thus bringing us back to the standard Lucas program. In this section we argue that the results of the present paper can still be useful even in this case.

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

LOCAL INFORMATIVENESS IN A GENERAL STRUCTURAL MODEL. We consider a researcher that entertains a particular structural model $\zeta \in Z$, where ζ denotes a vector of the model’s structural parameters. As a result of model estimation (or simply through some kind of prior information), the researcher entertains a distribution over that parameter vector:

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

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

Our formalization of the notion of informativeness is inspired by—though conceptually distinct from—Andrews et al. (2020).²¹ In a neighborhood of ζ_0 , the covariance matrix of

²¹We compute the exact same measure of informativeness as Andrews et al., (49). The interpretation, however, is rather different: Andrews et al. jointly estimate a model as well as descriptive statistics (their γ), while we study the informativeness of certain features of the model (our γ) for others (our c) conditional on the particular estimated model.

(c, γ) is given as

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

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

$$\Delta_i \equiv \frac{\Sigma_{c_i\gamma} \Sigma_\gamma^{-1} \Sigma_{\gamma c_i}}{\Sigma_{c_i}} \in [0, 1] \quad (49)$$

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

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

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

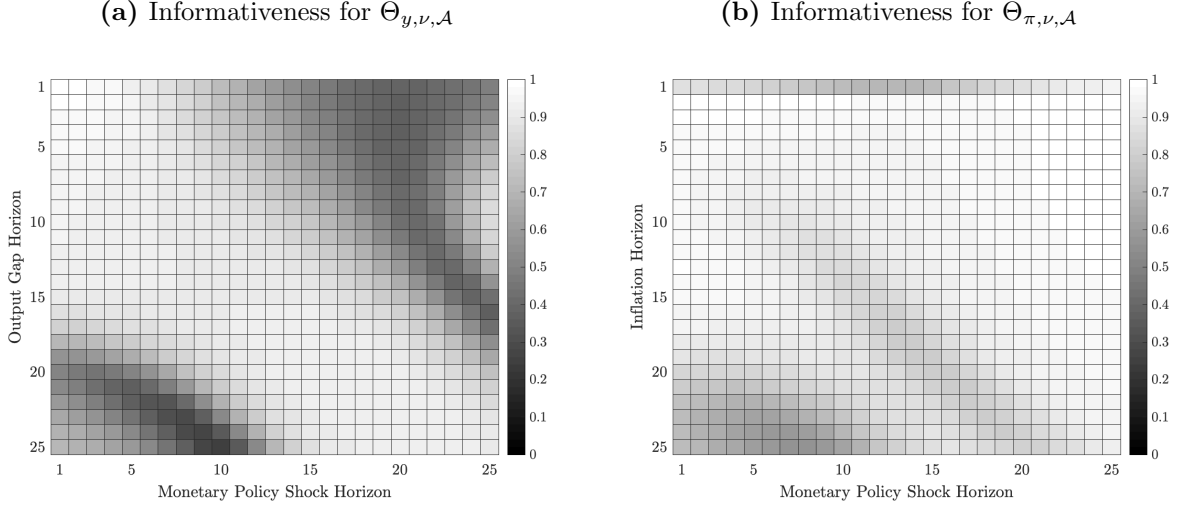


Figure 7: Output gap and inflation informativeness for monetary policy shocks in the structural model of Smets & Wouters (2007), computed using (49) and for the observables γ defined in (50).

in the matrices $\{\Omega_{y,\mathcal{A}}, \Omega_{\pi,\mathcal{A}}\}$. We then proceed as follows: for the output causal effect map $\Theta_{y,\nu,\mathcal{A}}$ as the counterfactual c , we select as our observables γ the short- and medium-run average responses of output to our two identified policy instrument paths, i.e.,

$$\gamma = \left(\frac{1}{4} \sum_{h=1}^4 \Omega_{y,\mathcal{A}}(h, \bullet), \frac{1}{12} \sum_{h=5}^{16} \Omega_{y,\mathcal{A}}(h, \bullet) \right) \quad (50)$$

We proceed similarly for inflation. We thus in both cases ask the question: how much does knowledge of *only* the average short- and medium-run causal effects of the observed instrument paths onto the outcome of interest—i.e., four numbers—restrict the remainder of the (high-dimensional) policy shock causal effect maps? Conclusions for alternative sets of observables as well as details for our computations are provided in Appendix C.2.

The left panel of Figure 7 shows the informativeness measure Δ_i of our four output gap impulse response moments for the rest of $\Theta_{y,\nu,\mathcal{A}}$, while the right panel does the same for inflation. The heatmaps reveal that informativeness is reasonably high throughout, with averages of 0.75 for the output gap and 0.91 for inflation. Informativeness is particularly high for short-term shocks—corresponding to our two identified instrument paths—and relatively short horizons—corresponding to the averaged impulse responses in our γ 's—and decreases away from the main diagonal. Adding a third long-run average to our observables γ , the

INFORMATIVENESS Δ FOR RULE (37) IN SMETS & WOUTERS

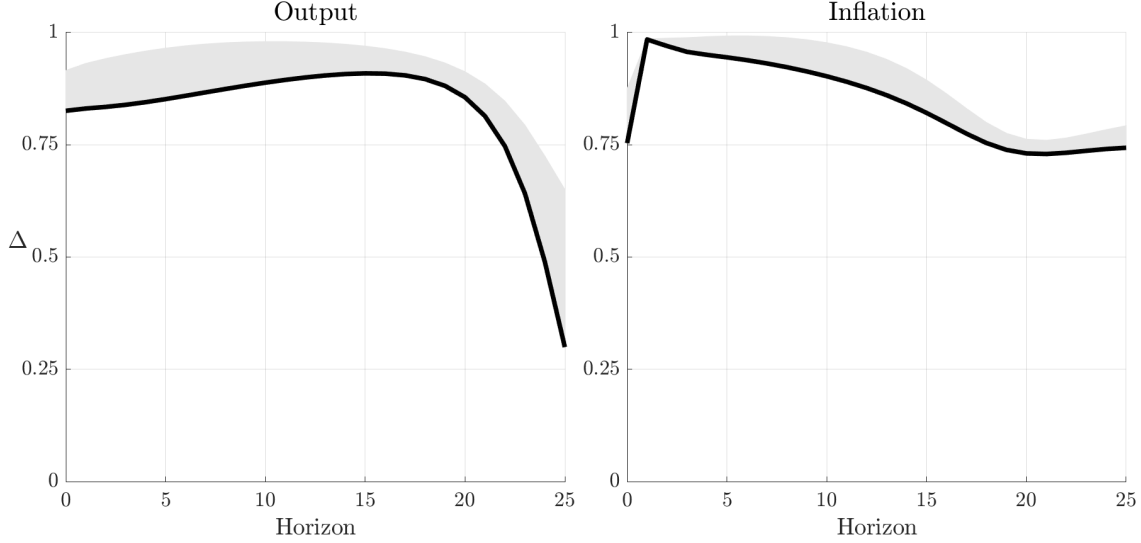


Figure 8: Time paths of the output gap and inflation informativeness in the structural model of Smets & Wouters (2007) for the counterfactual rule (37), to be implemented following the model’s estimated investment-specific technology shock, for γ defined as in (50) (solid line) and adding a third, long-run impulse response observable (shaded, see Footnote 22).

average Δ ’s increase to 0.94 and 0.95, respectively.²²

While Figure 7 depicts our measure of informativeness for the *entire* causal effect maps, in practice counterfactuals for typical business-cycle fluctuations are likely to depend mainly on impulse responses to contemporaneous and a couple of short-run policy news shocks, as for example suggested by our illustrative analysis in Section 2.4. Given this observation, we would expect informativeness for such *particular* policy counterfactuals to be even higher than the averages across all of the shock causal effects reported above. We illustrate this conjecture by computing our informativeness measure for a particular shock—the investment-specific technology shock of Smets & Wouters—and a particular counterfactual rule—the rule (37) previously considered in Section 2.4. We pick this rule because it implies substantial interest rate inertia and thus lies outside of the purview of the partial model analysis from Section 4.1. Figure 8 presents the results, plotting horizon-by-horizon informativeness for the desired counterfactual, in solid for our baseline γ (i.e., four observables) and shaded if we add long-run response for each shock. Exactly as expected, the informativeness measures are higher than the averages reported before at short horizons, before falling at longer horizons.

²²To be precise, we set $\gamma = \left(\frac{1}{4} \sum_{h=1}^4 \Omega_{y,\mathcal{A}}(h, \bullet), \frac{1}{8} \sum_{h=5}^{12} \Omega_{y,\mathcal{A}}(h, \bullet), \frac{1}{8} \sum_{h=13}^{20} \Omega_{y,\mathcal{A}}(h, \bullet) \right)$.

DISCUSSION. Closely building on Andrews et al. (2020), the analysis in this section has introduced a tool that allows researchers to communicate—given their maintained parametric structural model—which moments of the data “drive” their reported policy counterfactuals. As expected in light of our theoretical identification results, we find that impulse responses to identified policy shocks can be highly informative “identified moments” (Nakamura & Steinsson, 2018) for structural policy rule counterfactuals.

5 Conclusions

The standard approach to counterfactual analysis for changes in policy rules relies on fully-specified general equilibrium models. Our identification results instead point in a different direction: 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.

An important challenge in implementing this strategy is that its informational requirements are high. In Section 3 we demonstrated by example that existing evidence is already rich enough to allow us to study certain interesting policy counterfactuals. For counterfactuals outside of the space already covered by existing empirical evidence, a natural reaction is to impose more economic 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 4.1) as well as further general results on how impulse responses to different policy shocks are tied together (as in Section 4.2) would be particularly welcome.

The key message of this paper for future empirical work is to emphasize the value of empirical strategies that recover the dynamic causal effects associated with different *time paths* of policy instruments. Every additional piece of empirical evidence on a different policy instrument path will 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.

References

- Andrews, I., Gentzkow, M., & Shapiro, J. M. (2020). On the informativeness of descriptive statistics for structural estimates. *Econometrica*, 88(6), 2231–2258.
- Antolin-Diaz, J., Petrella, I., & Rubio-Ramírez, J. F. (2021). Structural scenario analysis with SVARs. *Journal of Monetary Economics*, 117, 798–815.
- Auclert, A., Bardóczy, B., Rognlie, M., & Straub, L. (2021). Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models. Working Paper.
- Auclert, A., Rognlie, M., & Straub, L. (2018). The Intertemporal Keynesian Cross. Technical report, National Bureau of Economic Research.
- Auclert, A., Rognlie, M., & Straub, L. (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model. Technical report, National Bureau of Economic Research.
- Barnichon, R. & Mesters, G. (2020). Identifying modern macro equations with old shocks. *The Quarterly Journal of Economics*, 135(4), 2255–2298.
- Barnichon, R. & Mesters, G. (2021). Testing Macroeconomic Policies with Sufficient Statistics. Working Paper.
- 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. (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., & Ristinieni, A. (2021). A Toolkit for Computing Constrained Optimal Policy Projections (COPPs). ECB Working Paper.
- Eberly, J. C., Stock, J. H., & Wright, J. H. (2020). The federal reserve’s current framework for monetary policy: A review and assessment. *International Journal of Central Banking*, *16*(1), 5–71.
- Fernández-Villaverde, J., Rubio-Ramírez, J. F., & 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.
- Hamilton, J. D. & Herrera, A. M. (2004). Oil shocks and aggregate macroeconomic behavior: The role of monetary policy. *Journal of Money, Credit and Banking*, *36*(2), 265–286.
- 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.

- Justiniano, A., Primiceri, G. E., & Tambalotti, A. (2010). Investment shocks and business cycles. *Journal of Monetary Economics*, *57*(2), 132–145.
- Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, *111*(4), 1092–1125.
- Kaplan, G., Moll, B., & Violante, G. (2018). Monetary Policy according to HANK. *American Economic Review*, *108*(3), 697–743.
- Kilian, L. & Lewis, L. T. (2011). Does the Fed Respond to Oil Price Shocks? *The Economic Journal*, *121*(555), 1047–1072.
- 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*.
- Nakamura, E. & Steinsson, J. (2018). Identification in macroeconomics. *Journal of Economic Perspectives*, *32*(3), 59–86.
- Ottonello, P. & Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, *88*(6), 2473–2502.
- Plagborg-Møller, M. & Wolf, C. K. (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. (2011). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics*, 126(1), 1–50.
- Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of macroeconomics*, 2, 71–162.
- 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.
- 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*, 16(1), 133–193.
- Wieland, J. F. & Yang, M.-J. (2020). Financial dampening. *Journal of Money, Credit and Banking*, 52(1), 79–113.

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

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

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

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

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

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

Contents

A	Supplementary theoretical results	47
A.1	Examples of nested models	47
A.2	Filtering problems	49
A.3	Global identification argument	50
A.4	More general loss functions	53
A.5	Counterfactuals with finitely many shocks	53
B	Details for identified subspace analysis	56
B.1	Data	56
B.2	Shock & policy dynamic causal effects	56
B.3	Average inflation targeting loss function	58
B.4	Computing best linear approximations	59
C	Details for “identified moment” results	61
C.1	NKPC theory & estimation	61
C.2	Informativeness in the Smets & Wouters model	62

A Supplementary theoretical results

This appendix provides several results complementing our theoretical identification analysis of Section 2. Appendix A.1 discusses examples of models that are nested by our results, Appendix A.2 gives an example of a popular model that is not, Appendix A.3 provides a global identification analysis with even higher informational requirements, and Appendix A.4 extends our optimal policy arguments to more general loss functions. Finally in Appendix A.5, we compare our approach with that of the Sims (1980, 1982, 1987) program.

A.1 Examples of nested models

We provide further details on four sets of models: the simple example of Section 2.1, the three-equation New Keynesian model referenced in Section 2.2, a general class of behavioral models, and the quantitative HANK model of Section 2.4.

SIMPLE EXAMPLE. Solving the system (1) - (4) gives

$$\begin{aligned} x_0(\boldsymbol{\varepsilon}, \boldsymbol{\nu}) &= \frac{1}{1 + \beta\phi} \left[\varepsilon_0 - \beta\nu_0 + \frac{\gamma}{1 + \beta\phi} \{\varepsilon_1 - \beta\nu_1\} \right] \\ x_1(\boldsymbol{\varepsilon}, \boldsymbol{\nu}) &= \frac{1}{1 + \beta\phi} [\varepsilon_1 - \beta\nu_1] \\ z_0(\boldsymbol{\varepsilon}, \boldsymbol{\nu}) &= \frac{\phi}{1 + \beta\phi} \left[\varepsilon_0 - \beta\nu_0 + \frac{\gamma}{1 + \beta\phi} \{\varepsilon_1 - \beta\nu_1\} \right] + \nu_0 \\ z_1(\boldsymbol{\varepsilon}, \boldsymbol{\nu}) &= \frac{\phi}{1 + \beta\phi} [\varepsilon_1 - \beta\nu_1] + \nu_1 \end{aligned}$$

Solving (10) for $\tilde{\boldsymbol{\nu}}$ gives

$$\begin{aligned} \tilde{\nu}_0 &= \frac{\tilde{\phi} - \phi}{(1 + \beta\phi) + \beta(\tilde{\phi} - \phi)} \left[\varepsilon_0 + \frac{\gamma}{1 + \beta\phi} \left(1 - \beta \frac{\tilde{\phi} - \phi}{(1 + \beta\phi) + \beta(\tilde{\phi} - \phi)} \right) \varepsilon_1 \right] \\ \tilde{\nu}_1 &= \frac{\tilde{\phi} - \phi}{(1 + \beta\phi) + \beta(\tilde{\phi} - \phi)} \varepsilon_1 \end{aligned}$$

Plugging these expressions for $\tilde{\boldsymbol{\nu}}$ into the impulse responses of the system and simplifying, we recover (11) - (12).

THREE-EQUATION NK MODEL. We begin with the non-policy block. The Phillips curve can be written as

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

while the Euler equation can be written as

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

Letting $\mathbf{x} \equiv (\boldsymbol{\pi}', \mathbf{y}')$, we can stack these linear maps into the form (14). Finally the policy rule can be written as

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

which fits into the form of (15).

BEHAVIORAL MODEL. 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 θ would correspond to

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

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

summarized as

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

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

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

Behavioral frictions thus merely affect the matrices that enter our general non-policy block (14), but do not affect the separation of policy- and non-policy blocks at the heart of our identification result.

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

A.2 Filtering problems

To illustrate how an asymmetry in information between the private sector and the policy authority can break our separation of the policy and non-policy blocks in (14) - (15) even for a linear model, we consider a standard Lucas (1972) island model with a slightly generalized policy rule. The policy authority sets nominal demand x_t according to the rule

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

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

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

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

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

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

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

and so

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

But since

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

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

A.3 Global identification argument

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

SETTING. We consider an economy that runs for T periods overall. As in our main analysis, the economy consists of a private block and a policy block. Differently from our main analysis, there is no exogenous non-policy shock sequence $\boldsymbol{\varepsilon}$; rather, there is a stochastic event ω_t each period, with stochastic events drawn from a finite (n_ω -dimensional) set. Let $x_t(\omega^t)$ be the value of the endogenous variables after history $\omega^t \equiv \{\omega_0, \omega_1, \dots, \omega_t\}$ and let $z_t(\omega^t)$ be the realization of the policy instruments after history ω^t . Let \boldsymbol{x} and \boldsymbol{z} be the full contingent plans for for all $t \in \{0, 1, \dots, T\}$ and all histories. \boldsymbol{x} and \boldsymbol{z} are vectors in $\mathbb{R}^{n_x \times N}$ and $\mathbb{R}^{n_z \times N}$ respectively, where $N = n_\omega + n_\omega^2 + \dots + n_\omega^{T+1}$.

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

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{z}) = \mathbf{0}. \tag{A.1}$$

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

$$\bar{\mathcal{A}}(\mathbf{x}, \mathbf{z}) + \boldsymbol{\nu} = \mathbf{0} \quad (\text{A.2})$$

where the vector of policy shocks $\boldsymbol{\nu}$ is now $n_z \times N$ dimensional. We assume that, for any $\boldsymbol{\nu} \in \mathbb{R}^{n_z \times N}$, the system (A.1) - (A.2) has a unique solution. We write this solution as

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

We want to construct counterfactuals under the alternative policy rule

$$\tilde{\mathcal{A}}(\mathbf{x}, \mathbf{z}) = \mathbf{0} \quad (\text{A.3})$$

replacing (A.2). We again assume that the system (A.1) and (A.3) has a unique solution, now written as $(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$. If we are interested in the counterfactual following a particular path of exogenous events, then we are interested in selections from these vectors.

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

$$\bar{x}(\tilde{\boldsymbol{\nu}}) = \tilde{\mathbf{x}}, \quad \bar{z}(\tilde{\boldsymbol{\nu}}) = \tilde{\mathbf{z}} \quad (\text{A.4})$$

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

$$\tilde{\mathcal{A}}(\bar{x}(\tilde{\boldsymbol{\nu}}), \bar{z}(\tilde{\boldsymbol{\nu}})) = \mathbf{0}. \quad (\text{A.5})$$

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

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

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

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

$$\mathcal{H}(\bar{x}(\tilde{\boldsymbol{\nu}}), \bar{z}(\tilde{\boldsymbol{\nu}})) = \mathbf{0} \quad (\text{A.6})$$

$$\bar{\mathcal{A}}(\bar{x}(\tilde{\boldsymbol{\nu}}), \bar{z}(\tilde{\boldsymbol{\nu}})) + \tilde{\mathcal{A}}(\tilde{\mathbf{x}}, \tilde{\mathbf{z}}) - \bar{\mathcal{A}}(\tilde{\mathbf{x}}, \tilde{\mathbf{z}}) = \mathbf{0} \quad (\text{A.7})$$

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

$$\mathcal{H}(\tilde{x}, \tilde{z}) = \mathbf{0} \quad (\text{A.8})$$

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

It follows that $\{\bar{x}(\tilde{\nu}) = \tilde{x}, \bar{z}(\tilde{\nu}) = \tilde{z}\}$ is a solution of the system (A.6) - (A.7). By assumption this system has a unique solution, so it must be that $\tilde{\nu}$ satisfies $\{\bar{x}(\tilde{\nu}) = \tilde{x}, \bar{z}(\tilde{\nu}) = \tilde{z}\}$.

We now show that any solution to (A.5) must generate (\tilde{x}, \tilde{z}) . Proceeding by contradiction, consider any other $\tilde{\nu}$ that solves (A.5) and suppose that either $\bar{x}(\tilde{\nu}) \neq \tilde{x}$ and/or $\bar{z}(\tilde{\nu}) \neq \tilde{z}$. By definition of the functions $\bar{x}(\bullet)$ and $\bar{z}(\bullet)$ together with the property (A.5) we know that

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

and so $(\bar{x}(\tilde{\nu}), \bar{z}(\tilde{\nu}))$ is a solution of (A.1) and (A.3) that is distinct from (\tilde{x}, \tilde{z}) . But by assumption only one such solution exists, so we have a contradiction. □

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

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

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

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

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

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

A.4 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

$$\mathcal{L} = \mathbf{x}'Q\mathbf{x} \tag{A.10}$$

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.10) subject to (33). The first-order conditions of this problem are

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

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 \mathbf{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.5 Counterfactuals with finitely many shocks

We in this section provide further details regarding the conventional Sims (1980, 1982, 1987) program approach to constructing policy counterfactuals. We begin with a discussion of how to construct the Sims program counterfactual. We then discuss the extension to a finite

ALTERNATIVE POLICY RULE, HANK MODEL (SIMS PROGRAM)

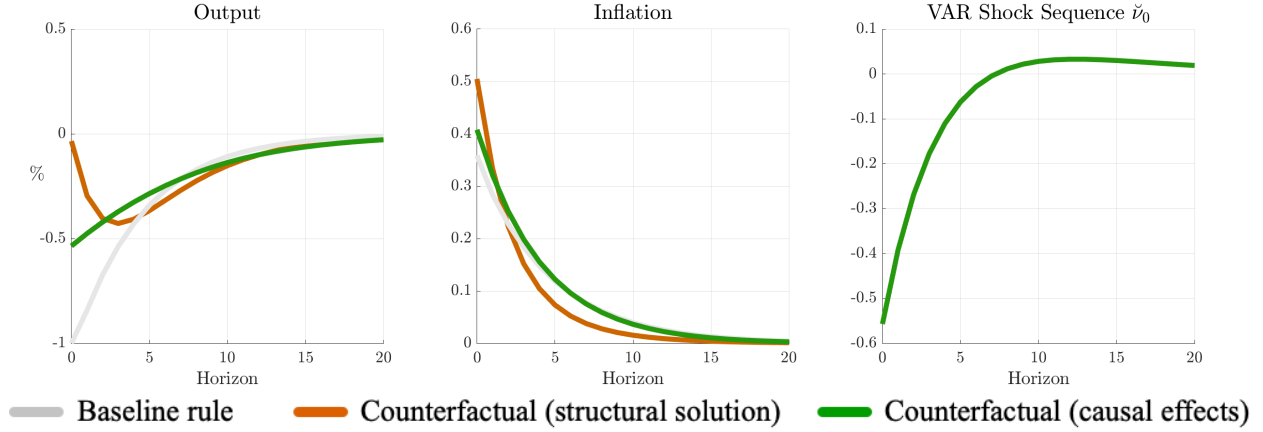


Figure A.1: Output and inflation impulse responses for the HANK model with policy rules (36) and (37) vs. impulse responses to a sequence of contemporaneous shocks $\check{\nu}_0$ that enforce (37) along the equilibrium transition path. The impact output contraction under the prevailing baseline rule is normalized to -1% .

number of observed policy shocks.

STANDARD SIMS PROGRAM. This approach builds policy counterfactuals using empirical estimates of the dynamic causal effects of a single (contemporaneous) policy shock; that is, the researcher knows the first column of the maps in $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$. To predict the behavior of the economy under an alternative path of the policy instrument, the economy is then subjected to a *sequence* of contemporaneous policy shocks $\check{\nu}_0$ 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. To understand how these counterfactuals relate to the Lucas program counterfactuals we note that they may be interpreted as the response to a systematic change in policy under auxiliary structural assumptions—assumptions that put further structure on Θ . 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:

$$\Theta_{q,\nu,\mathcal{A}} = \begin{pmatrix} \Theta_{q,\nu,\mathcal{A}}(1,1) & 0 & 0 & \dots \\ \Theta_{q,\nu,\mathcal{A}}(2,1) & \Theta_{q,\nu,\mathcal{A}}(1,1) & 0 & \dots \\ \Theta_{q,\nu,\mathcal{A}}(3,1) & \Theta_{q,\nu,\mathcal{A}}(2,1) & \Theta_{q,\nu,\mathcal{A}}(1,1) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad q \in \{x, z\} \quad (\text{A.11})$$

where $\Theta_{\bullet}(i, j)$ denotes the (i, j) th entry of a map Θ_{\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 (A.11) is inconsistent with typical (rational-expectations) macroeconomic models with forward-looking agents. In such environments, using the structure in (A.11) to predict the effects of changes in systematic policy rules will run afoul of the Lucas critique. Figure A.1—which implements the Sims program approach in our HANK environment—provides a quantitative illustration.

B Details for identified subspace analysis

This appendix provides further details for the applications of our identified subspace analysis in Section 3.2. Appendices B.1 and B.2 present the data and estimation strategy used for our identification of non-policy and policy shock causal effects in aggregate time series data. Appendix B.3 then provides further details on the dual mandate loss function used for our optimal policy counterfactuals, and Appendix B.4 explains how to in practice compute the best linear approximations for (40) and (41).

B.1 Data

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

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

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

B.2 Shock & policy dynamic causal effects

For maximal consistency, we try to estimate all impulse responses within a common empirical specification. For the investment-specific technology shocks, we order the shock measure first in a recursive VAR containing our outcomes of interest (following Plagborg-Møller & Wolf,

POLICY SHOCK IMPULSE RESPONSES, ROMER & ROMER (2004)

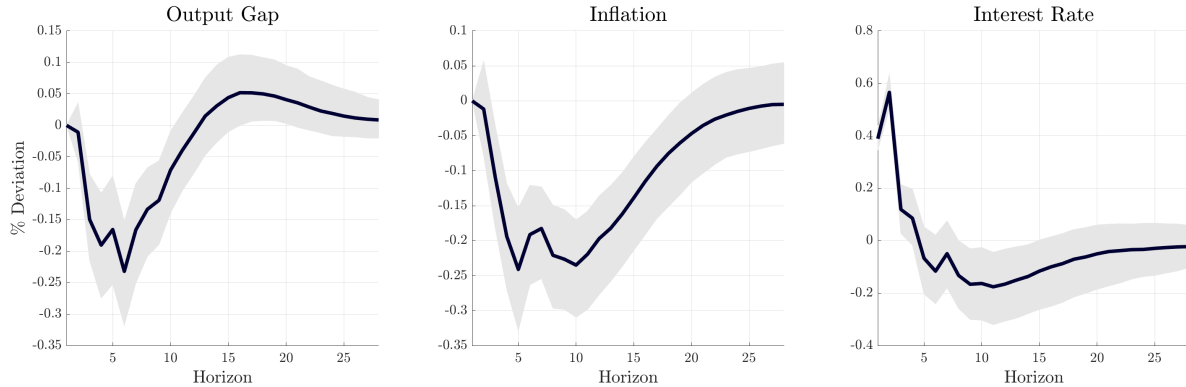


Figure B.1: Impulse responses after the Romer & Romer shock. The grey areas correspond to 16th and 84th percentile confidence bands, constructed using 10,000 draws from the posterior distribution of the reduced-form VAR parameters.

2021), estimated on a sample from 1969:Q1–2007:Q4. For our two monetary policy shocks, we estimate a single VAR in the two shock series, our three outcomes of interest, as well as commodity prices, also estimated from 1969:Q1–2007:Q4.²³ For identification, we order the Gertler & Karadi shock first (again consistent with the results in Plagborg-Møller & Wolf (2021)) and the Romer & Romer last (exactly as in Romer & Romer (2004)).

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. Since the transmission of both shocks is estimated within a single VAR, we can simply draw from the model posterior and then compute the counterfactuals for each draw, thus taking into account joint estimation uncertainty.

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

²³The Gertler & Karadi shock series is only available from 1988 onwards. We thus follow prior work in the macro IV literature (e.g. Känzig, 2021) and set the missing values to zero.

POLICY SHOCK IMPULSE RESPONSES, GERTLER & KARADI (2015)

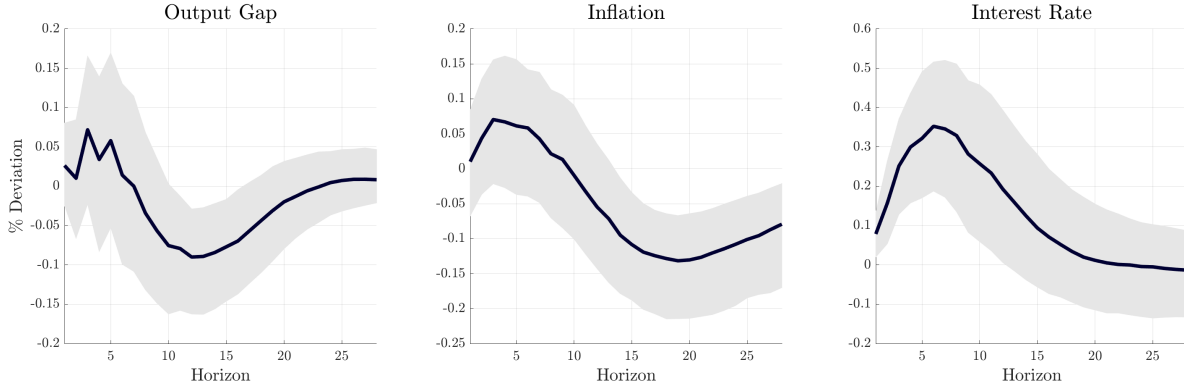


Figure B.2: Impulse responses after the Gertler & Karadi shock. The grey areas correspond to 16th and 84th percentile confidence bands, constructed using 10,000 draws from the posterior distribution of the reduced-form VAR parameters.

with the differences in the induced interest rate paths.

FURTHER SYSTEMATIC RULE COUNTERFACTUALS. In Section 3.2 we constructed the best possible approximation to our desired counterfactual policy rule—perfect output stabilization—given the estimated non-policy shock impulse response paths $\{\mathbf{x}(\boldsymbol{\varepsilon}), \mathbf{z}(\boldsymbol{\varepsilon})\}$. Figure B.3 instead reports the best approximation to those shock impulse responses subject to the constraint that perfect output stabilization is possible. By construction the interest rate and inflation paths are unchanged, while the output contraction for the first couple of quarters is smaller, exactly mirroring the results in Figure 4.²⁴

B.3 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

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

²⁴Note that the uncertainty bands for the initial output contraction mirror the uncertainty regarding our estimated causal effects of monetary policy shocks on output.

POLICY COUNTERFACTUAL FOR INVESTMENT SHOCKS, IDENTIFIED SUBSPACE

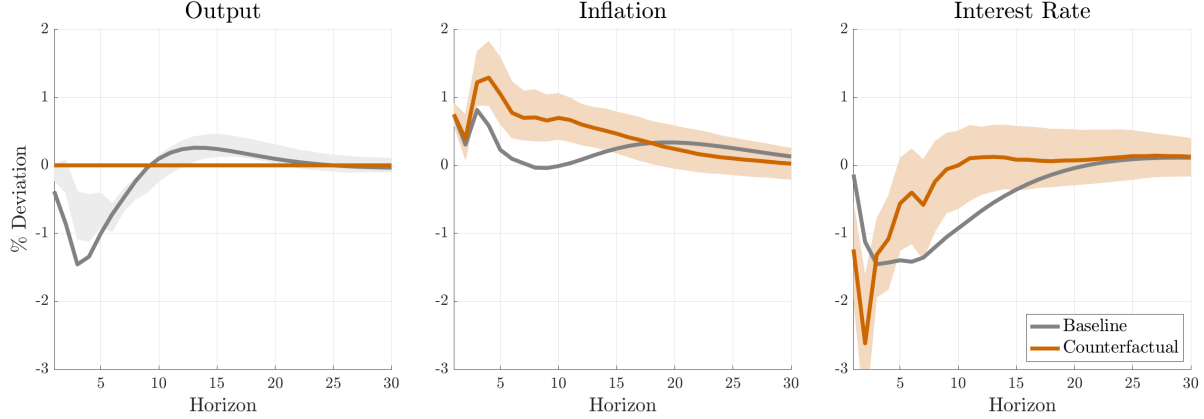


Figure B.3: The dark grey lines show impulse responses of the output gap, inflation and interest rate to the estimated contractionary investment-specific technology shock. The light shaded grey area shows the best approximation to those paths such that perfect output stabilization is feasible using our identified shocks, constructed following (41) (16th and 84th percentile confidence bands). The orange lines show impulse responses under the counterfactual rule, with the shaded areas corresponding to 16th and 84th percentile confidence bands.

Here H denotes the maximal (lagged) horizon that enters the inflation averaging, and ω_ℓ denotes the weight on the ℓ 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 $\bar{\pi} = \bar{\Pi} \times \pi$.

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

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

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

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

Using the definition of Π_y and simplifying, we recover (48).

B.4 Computing best linear approximations

The solution to our first optimization problem (40) is given as

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

For the second optimization problem (41), we get the first-order conditions

$$\begin{aligned} (\mathbf{x}^\dagger - \mathbf{x}_{\mathcal{A}}(\boldsymbol{\varepsilon})) + \tilde{\mathcal{A}}'_x \boldsymbol{\lambda} &= \mathbf{0} \\ (\mathbf{z}^\dagger - \mathbf{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})) + \tilde{\mathcal{A}}'_z \boldsymbol{\lambda} &= \mathbf{0} \\ \Omega'_{x,\mathcal{A}} \tilde{\mathcal{A}}'_x \boldsymbol{\lambda} + \Omega'_{z,\mathcal{A}} \tilde{\mathcal{A}}'_z \boldsymbol{\lambda} &= \mathbf{0} \\ \tilde{\mathcal{A}}_x (\mathbf{x}^\dagger + \Omega_{x,\mathcal{A}} \times \mathbf{s}) + \tilde{\mathcal{A}}_z (\mathbf{z}^\dagger + \Omega_{z,\mathcal{A}} \times \mathbf{s}) &= \mathbf{0} \end{aligned}$$

where $\boldsymbol{\lambda}$ denotes the multiplier on the constraint forcing the new rule to hold exactly. Solving, we obtain the paths $(\mathbf{x}^\dagger, \mathbf{z}^\dagger)$.

Finally, for the third problem, our computational implementation is already discussed in detail in Appendix A.5.

C Details for “identified moment” results

Appendix C.1 elaborates on the dual mandate counterfactuals and partial model block estimation of Section 4.1, while Appendix C.2 does the same for our quantitative informativeness results of Section 4.2

C.1 NKPC theory & estimation

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

RECOVERING POLICY COUNTERFACTUALS FROM Π_y . Knowledge of Π_y —together with the assumption that $\Theta_{\pi,\nu,\mathcal{A}}$ is invertible, i.e., any path of inflation is in principle implementable—is sufficient to construct output and inflation counterfactuals corresponding to alternative rules of the general form (43). Formally, we can recover the desired counterfactual outcomes by solving the system

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

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

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

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

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

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

SPECIAL CASE: AUGMENTED PHILLIPS CURVE. Consider the augmented Phillips curve (47). Along a perfect foresight transition path, we can write this relationship as

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

We thus have

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

ESTIMATION DETAILS. Barnichon & Mesters (2020) show how to use estimates of monetary policy impulse responses to identify a Phillips curve relationship of the form (47). For our empirical analysis in Figure 6 we closely follow their estimation strategy; since we use almost the same data (see Appendix B.1), our estimation results are very similar to theirs. In particular, for our headline results in Figure 6 we also impose the constraint that $\gamma_f + \gamma_b = 1$, so our confidence sets are almost identical to those reported in panel (B) of Figure II in the original Barnichon & Mesters (2020).

C.2 Informativeness in the Smets & Wouters model

We here provide further details for our analysis in Section 4.2. We first discuss several computational details and then present further robustness results.

COMPUTATIONAL DETAILS. Our estimation of the structural model of Smets & Wouters (2007) uses replication codes kindly provided by Johannes Pfeifer.²⁵ The estimation yields the posterior mode ζ_0 and the variance-covariance matrix Σ_ζ .

We compute the monetary policy shock causal effect maps $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ by solving the model using sequence-space methods, and then sequentially adding all different contemporaneous and news shocks to the policy rule. We then compute $\frac{\partial \Theta_{x,\nu,\mathcal{A}}}{\partial \zeta}$ and $\frac{\partial \Theta_{z,\nu,\mathcal{A}}}{\partial \zeta}$ using

²⁵The code is available at <https://sites.google.com/site/pfeiferecon/dynare>.

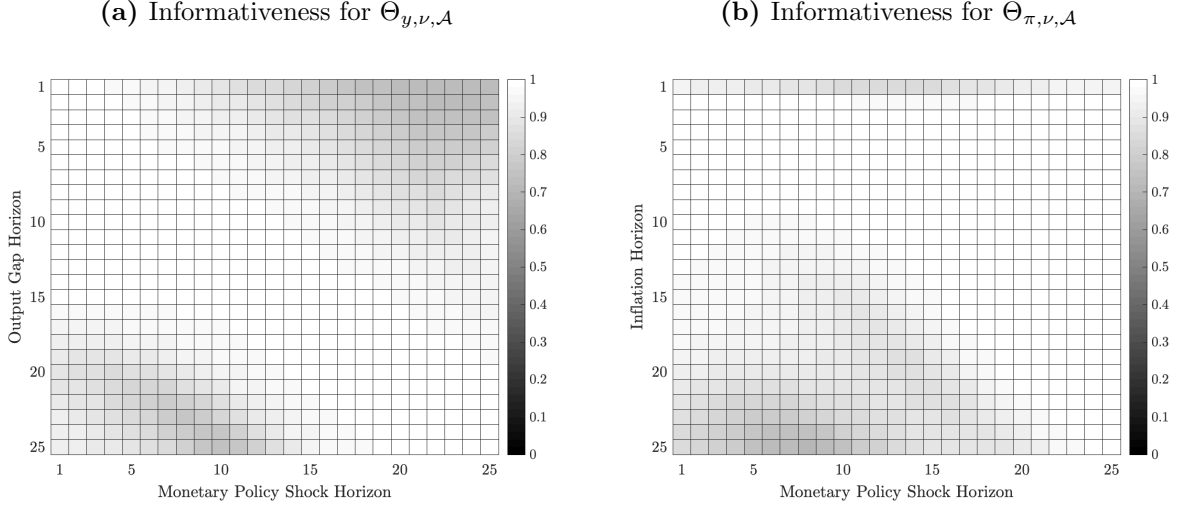


Figure C.1: Output gap and inflation informativeness for monetary policy shocks in the structural model of Smets & Wouters (2007), computed using (49) and for our extended set of observables γ (see Footnote 22).

finite-difference methods. Given that all counterfactuals c and observables γ are functions of $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$, we can use these derivative matrices to construct the variance-covariance matrix Σ for \hat{c} and $\hat{\gamma}$.

Our observables γ are chosen using the estimated policy instrument paths from our empirical analysis, plotted in Figure B.1 and Figure B.2. We take the point estimates of the interest rate path, and then at the estimated mode ζ_0 construct the sequence of monetary policy shocks ν_{rr} and ν_{gk} that would correspond to the two identified shocks. Our observables γ are then computed from the model-implied output and inflation impulse responses to those two shocks ν_{rr} and ν_{gk} . The informativeness of the impulse responses to these particular shocks for impulse responses to all other possible shocks is reported in Figure 7.

For the particular counterfactual studied in Figure 8, we consider the investment-specific technology shock estimated by Smets & Wouters. We then, at the model’s mode ζ_0 , compute the particular monetary policy shock paths $\tilde{\nu}$ that would map the investment-specific technology shock under the baseline to its counterfactual propagation under our alternative rule (37). Figure 8 shows the informativeness of our selected observed impulse responses (i.e., entries of the causal effects of ν_{rr} and ν_{gk}) for the responses to this $\tilde{\nu}$.

ROBUSTNESS. Our results are largely unchanged for alternative observables γ , as long as the dimensionality is kept as in our baseline analysis. For example, if we instead condition on impulse responses at horizons 3 and 8, then the average Δ ’s for the analogous versions of the

heatmaps in Figure 7 are 0.76 and 0.89, respectively. Instead projecting impulse responses on a simple quadratic spline (as in Barnichon & Mesters, 2020) gives 0.86 and 0.87.

For completeness we also report the informativeness heatmaps for the three observed impulse responses, as also shown in Figure 8. Results are reported in Figure C.1. As expected, the informativeness Δ is even higher than in Figure 7.