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

Alisdair McKay
Federal Reserve Bank of Minneapolis

Christian K. Wolf
MIT & NBER

This version: April 2023
First version: July 2021

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

Keywords: Lucas critique, policy counterfactuals, macroeconomic modeling, business cycles, monetary policy, policy shocks. *JEL codes:* E32, E61.

[†]Email: `alisdair.mckay@mpls.frb.org` and `ckwolf@mit.edu`. We received helpful comments from the editor Chad Jones, three anonymous referees, Isaiah Andrews, Marios Angeletos, Marco Bassetto, Martin Beraja, Anmol Bhandari, Francesco Bianchi, Gabriel Chodorow-Reich, Hal Cole, Peter Ganong, Jordi Galí, Mark Gertler, Yuriy Gorodnichenko, John Grigsby, Ben Moll, Emi Nakamura, Mikkel Plagborg-Møller, Richard Rogerson, Juan Rubio-Ramírez, Jón Steinsson, Robert Ulbricht, Mike Waugh, Iván Werning, Tom Winberry, Tao Zha, and seminar participants at various venues. We also thank Christiane Baumeister, Valerie Ramey and Bent Sørensen for valuable discussions. Jackson Mejia provided superb research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

An important function of macroeconomics is to predict the consequences of changes in policy. In this paper we revisit the role that evidence on policy *shocks*—that is, surprise deviations from a prevailing rule—can play in helping macroeconomists learn about policy *rule* counterfactuals. Existing work mainly uses such policy shocks in two ways. First, in what Christiano et al. (1999) call the "Lucas program", researchers begin by estimating the causal effects of a policy shock in the data, then construct a micro-founded structural model that matches these effects, and finally trust the model as a laboratory for predicting the effects of changes in policy rules. By design, this approach yields counterfactuals that are robust to the Lucas (1976) critique; on the other hand, the researcher needs to commit to a particular parametric model, thus introducing concerns about model misspecification. An alternative approach, proposed by Sims & Zha (1995), instead relies *only* on the estimated policy shock: in their procedure, the economy is subjected to a new policy shock at each date t, with the shocks chosen so that, t-by-t, the counterfactual policy rule holds.[1] This strategy does not require the researcher to commit to a particular model, but it is subject to the Lucas critique: a rule change announced at date 0 will in general have different effects on private-sector decisions than a sequence of surprise policy shocks at $t = 0, 1, \ldots$.

The contribution of this paper is to propose a method that constructs policy counterfactuals using empirical evidence on *multiple distinct* policy shocks, rather than just a single one. Like Sims & Zha, the method does not rely on a particular parametric structural model; at the same time, for a family of models that nests many of those popular in the Lucas program, it yields counterfactuals that are robust to the Lucas critique. At the heart of our methodology lies an identification result. We prove that, for a relatively general family of macro models, the causal effects of contemporaneous as well as news shocks to a given policy rule are sufficient to construct Lucas critique-robust counterfactuals for alternative policy rules. The core intuition is that, by subjecting the economy to multiple distinct policy shocks at date 0 (rather than a new value of a single shock at $t = 0, 1, \ldots$, as done in Sims & Zha), we are able to enforce the contemplated counterfactual policy rule not just *ex post* along the equilibrium path, but also *ex ante* in private-sector expectations. Under our assumptions, doing so is enough to fully sidestep the Lucas critique. While our exact identification result

[1]See for example Ramey (1993), Bernanke et al. (1997), Leeper & Zha (2003), Hamilton & Herrera (2004), Uribe & Yue (2006), Degasperi et al. (2020), Eberly et al. (2020), Brunnermeier et al. (2021), and Antolin-Diaz et al. (2021) for important applications and extensions of this method.

requires knowledge of the causal effects of a very large number of policy shocks, our proposed empirical method can be applied in the empirically relevant case of a researcher with access to only a couple of distinct shocks. We demonstrate the usefulness of the proposed approach with several applications to monetary policy rule counterfactuals.

IDENTIFICATION RESULT. The first part of the paper establishes the identification result. Our analysis builds on a general linear data-generating process, with one key added restriction: policy is allowed to affect private-sector behavior *only* through the current and future expected path of the policy instrument.[2] For example, for monetary policy, the private sector only cares about the expected future path of the nominal rate, and not whether this path is the result of the systematic component of policy—i.e., the policy *rule*—or due to *shocks* to a given rule. We consider an econometrician that lives in this economy and observes data generated under some baseline policy rule, where that rule is subject to shocks. Using standard time-series methods, she can estimate the causal effects of these policy shocks (Ramey, 2016). She then wishes to predict how a certain historical episode would have unfolded or how a particular shock would have propagated under some alternative policy rule.

In this setting we establish the following identification result. Suppose the econometrician is able to estimate how contemporaneous shocks to the prevailing rule as well as news about deviations from that rule *at all future horizons* affect the variables that enter her hypothesized counterfactual rule. Then these estimates contain all the information she needs to construct her desired counterfactual; in particular, she need not know any of the structural equations of the underlying model, including the prevailing policy rule. Key to the proof is our assumption on how policy is allowed to shape private-sector behavior. Since only the expected future path of the policy instrument matters, any given *rule*—characterized by the instrument path that it implies—can equivalently be synthesized by adding *shocks* to the baseline rule. All that is required is that those policy shocks imply the same expected instrument path from date-0 onwards as the counterfactual rule. Finally we show that, given a loss function, our econometrician can leverage the same logic to also characterize *optimal* policy.[3]

How general is the setting of this identification result? Our two key model restrictions

[2]More precisely, the policy rule is allowed to matter only through (a) the expected path of the instrument and (b) equilibrium selection. Our method will construct *one* valid equilibrium corresponding to the hypothesized counterfactual rule; if this rule induces a unique equilibrium, then our method recovers it.

[3]To be clear, our results are silent on the mapping from observables to welfare, and so on the shape of loss functions. Structural models are one way to arrive at such objectives. However, given that objective functions in practice are often derived from a legislated mandate rather than economic theory (e.g., dual mandate), we believe it is useful to have a method of calculating optimal policy for a given objective.

are (i) linearity and (ii) the way that policy is allowed to shape private-sector behavior. We show that (ii) is a feature shared by many business-cycle models, including those with many frictions (e.g., Christiano et al., 2005), shocks (e.g., Smets & Wouters, 2007), and rich micro heterogeneity (e.g., Kaplan et al., 2018). Perhaps the most popular class of structural models violating the restriction is those with an asymmetry of information between the policymaker and private sector (as in Lucas, 1972). In such models, private-sector agents solve a filtering problem, and so the policy rule affects both the dynamics of the policy instrument as well as the information contained in that policy choice; as a result, the policy instrument itself does not afford a full characterization of the policy stance. The linearity assumption (i), on the other hand, is not a conceptual necessity, but rather a practical one. Linearity implies that the effects of policy changes are invariant to their size, their sign, and the state of the economy. Given certainty equivalence, we can thus focus on expected values. As we will see, these simplifications are crucial to connect our theory to empirical evidence. Linearity does, of course, also impose costs: in practice, the methodology that we propose can be used to compare different cyclical stabilization policies (e.g., Taylor rules), but it is less well-suited to study policies that alter the steady state (e.g., changes in the inflation target).

Empirical strategy. The main challenge to operationalizing our identification result is that empirical evidence on the causal effects of policy shocks is limited. Our theory says that we need to select a linear combination of policy shocks at date 0 that perturbs the current and expected future path of the policy instrument just like the contemplated counterfactual rule. This is a daunting informational requirement: in general, to synthesize the effects of any possible expected policy instrument path of length T (with T large in practice), we would need access to T distinct policy shocks that each imply differentially shaped impulse response paths of the policy instrument, thus allowing us to span all of $\mathbb{R}^T$. While existing empirical evidence falls short of this ideal, recent research has however made progress on identifying the effects of at least *some* distinct policy shocks with rather different implications for future expected policy paths.[4] How much can be done with this available evidence?

The idea of our empirical method is to use the available evidence on policy shock trans-

[4] For monetary policy, many of the different popular shock series (e.g., Romer & Romer, 2004; Gertler & Karadi, 2015; Antolín-Díaz & Rubio-Ramírez, 2018; Bauer & Swanson, 2022) are well-known to lead to rather different responses of short-term rates. Other identification strategies explicitly aim to identify shocks at different parts of the yield curve (e.g., Gürkaynak et al., 2005; Antolin-Diaz et al., 2021; Inoue & Rossi, 2021), as required by our theory. For fiscal policy, Ramey (2011) and Ramey & Zubairy (2018) estimate the effects of short-lived as well as more persistent shocks. Mertens & Ravn (2010) and Leeper et al. (2013) are similarly focussed on disentangling shocks with different policy instrument dynamics.

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

Applications. We demonstrate the uses and limitations of our empirical method through several examples. Our object of interest is the propagation of a contractionary investment-specific technology shock under different monetary policy rules. As the inputs to our method, we consider the two most popular monetary policy shock series: those of Romer & Romer (2004) and Gertler & Karadi (2015). Importantly, these two shocks reflect different kinds of monetary news—a relatively transitory innovation for Romer & Romer, and a much more gradual rate change for Gertler & Karadi.

Armed with the causal effects associated with those two distinct nominal interest rate paths, we then apply our empirical method to construct counterfactuals for alternative policy rules that: target the output gap, enforce a Taylor-type rule, peg the nominal rate of interest, target nominal GDP, and minimize a simple dual-mandate loss function. We find that, with the exception of the nominal rate peg, the counterfactual rules can be enforced to quite a high degree of accuracy. The conclusion is that, at least for our investment shock, several rather different monetary policy counterfactuals can already be characterized quite sharply simply by combining existing pieces of empirical evidence on monetary policy shock transmission, without commitment to any particular parametric structural model.

Literature. Our identification result provides a bridge between the micro-founded models of the "Lucas program" (as discussed in Christiano et al., 1999) and the empirical strategy proposed by Sims & Zha (1995). Our results reveal that, in the structural models typically used in the Lucas program, the estimand of the econometric strategy of Sims & Zha is not equal to the true policy rule counterfactual *only* because of expectational effects related to the future conduct of policy. In theory, using multiple distinct policy shocks at date 0 (rather than a single one at each $t \geq 0$) circumvents this problem; in practice, doing so is feasible because a growing literature on the semi-structural identification of policy shocks provides us with a fairly rich body of empirical evidence (see the references in Footnote 4).[5]

Our work also relates to other more recent contributions to counterfactual policy analysis. Beraja (2020) similarly forms policy counterfactuals without relying on particular parametric models. His approach relies on stronger exclusion restrictions in the non-policy block of the economy, but given those restrictions requires less evidence on policy news shocks. Barnichon & Mesters (2021) use policy shock impulse responses to evaluate the optimality of and then improve upon a given policy decision. While their focus is on a single policy choice, we instead study systematic changes in the policy *rule*, requiring additional assumptions on the economic environment—our two assumptions discussed above.[6] More broadly, our work relates to the increasing popularity of a "sufficient statistics" logic for counterfactual analysis (e.g., Chetty, 2009; Arkolakis et al., 2012; Nakamura & Steinsson, 2018). Our identification result reveals that, across a broad class of models, the empirically estimable causal effects of policy shocks are precisely such sufficient statistics.

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

Outline. Section 2 presents our identification result, mapping the effects of policy *shocks* to counterfactuals for policy *rules*. Section 3 introduces our empirical methodology, and Section 4 provides applications to monetary policy rule counterfactuals. Section 5 concludes.

[5] A different route is taken in Leeper & Zha (2003): these authors argue that, if the policy shocks required to implement Sims & Zha are small enough, then it may be credible to *ignore* expectational effects.

[6] Building on our insight of the generality of the policy invariance assumption (ii), Barnichon & Mesters (2023) assume an environment as restrictive as ours as their baseline and then consider the more general case as an extension. Similarly related is Kocherlakota (2019), who presents a dynamic policy game in which the policymaker can select the optimal action via regression analysis. In his setting, the policy action does not cause the private sector to update its beliefs about the future strategy of the policymaker. Policymaker payoffs thus only depend on the current choice and not on the future instrument paths that we emphasize.

2 From policy shocks to policy rule counterfactuals

We begin in Section 2.1 by presenting a stylized version of our identification argument in a simple, illustrative model. We then in Sections 2.2 to 2.5 extend the argument to a general class of infinite-horizon models and discuss its scope and limitations.

The main identification result will be presented for a linearized perfect-foresight economy. Due to certainty equivalence, the equilibrium dynamics of a linear model with uncertainty 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 first-order perturbation techniques.[7] In particular, the perfect-foresight transition paths that we characterize will correspond to *expected* transition paths—or impulse response functions—in the analogous linearized economy with aggregate risk.

2.1 A simple example

This section presents our identification result in the context of the three-equation New Keynesian model (Galí, 2015; Woodford, 2003). Our broader argument, of course, is that the identification results and empirical method presented in the remainder of the paper actually do *not* require knowledge of the underlying structural model; nevertheless, we find it useful to first explain the logic of our results in a familiar setting before then generalizing it.

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

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

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

and a simple Taylor rule as the policy rule,

$$i_t = \phi \pi_t + \underbrace{\nu_{0,t}}_{\text{contemp. shock}} + \underbrace{\nu_{1,t-1}}_{\text{1-period news shock}}. \tag{3}$$

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

In our perfect-foresight set-up, the private-sector equations as well as the policy rule hold for $t = 0, 1, 2, \ldots$. These equations feature two kinds of disturbances. First, ε_t is a cost-push shock; for the illustrative analysis in this section, we find it useful to assume that it induces a first-order moving average wedge in the Phillips curve (2), implying that the effects of the shock will fully die out after two periods. Second, there are the policy shocks $\nu_{\ell,t-\ell}$; here, $\nu_{0,t}$ is a conventional contemporaneous policy shock, while $\nu_{1,t-1}$ denotes a deviation from the policy rule at time t announced at $t-1$ (a 1-period "news" shock). Note that in principle (3) could be generalized to feature a full menu of news shocks $\nu_{\ell,t-\ell}$ for all $\ell > 0$; this extension will be important for our general analysis, but is not needed here as we will only construct policy counterfactuals for the MA(1) shock ε_t. As usual, given a vector of the time-0 shocks $\{\varepsilon_0, \nu_{0,0}, \nu_{1,0}\}$, a perfect-foresight transition path—or impulse response function—consists of the paths $\{y_t, \pi_t, i_t\}$ such that (1) - (3) hold at all $t = 0, 1, 2, \ldots$.

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

Object of interest. Under the baseline policy rule, the impulse response of the economy to a cost-push shock is given as the solution of (1) - (3) for some cost-push shock ε_0 together with $\nu_{\ell,0} = 0$ for $\ell = 0, 1$. We wish to instead characterize the behavior of this economy in response to ε_0 not under the baseline policy rule (3), but instead under some counterfactual policy rule of the form[8]

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

where $\tilde{\phi} \neq \phi$. Note that this thought experiment supposes that the private sector perfectly understands the change in rule: the new relationship between i and π holds for all $t \geq 0$. Our identification result characterizes the information required to construct this counterfactual.

The identification argument. We consider an econometrician living in an economy that satisfies (1) - (3). Using conventional semi-structural time series methods (Ramey, 2016), and with access to suitable identifying assumptions or instruments, that econometri-

[8]For the analysis in this section we will assume that $\tilde{\phi}$ is such that a unique and determinate equilibrium exists. Our general analysis will cover counterfactual equilibrium non-existence and indeterminacy.

cian can in principle estimate how the macroeconomic aggregates $\{y_t, \pi_t, i_t\}$ respond to the cost-push shock ε_t as well as the policy shocks $\{\nu_{\ell,t-\ell}\}_{\ell=0}^{1}$ *under the baseline rule* (3). Our main identification result states that knowledge of these causal effects—and nothing else about the structure of the economy—is sufficient to predict the counterfactual propagation of the shock ε_t under the alternative rule (4). We now describe intuitively why knowledge of these estimable causal effects is sufficient in the simple model (1) - (3), before in Sections 2.2 and 2.3 stating and proving the result for a much more general environment.

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

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

The left-hand side of this equation gives us the impulse response of the interest rate following our combination of three shocks $\{\varepsilon_0, \tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ under the baseline rule (3), while the right-hand side does the same for inflation, just scaled by $\tilde{\phi}$. By our informational assumptions, the econometrician can evaluate the system of equations (5) given ε_0 and for any candidate set of the two policy shocks $\{\tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$. Now suppose a solution $\{\tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ to (5) exists, and then compute the responses of $\{y_t, \pi_t, i_t\}$ to the shock tuple $\{\varepsilon_0, \tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ under the baseline policy rule.[9] The content of our identification result is that those impulse responses are in fact identical to the impulse responses to ε_0 alone under the counterfactual rule (4).

The intuition underlying the identification result is straightforward. Since the private sector's decisions only depend on the expected path of the policy instrument (here just i_0

[9]Our general discussion will address the question of when solutions to equations like (5) actually exist.

and i_1), it follows that it does not matter whether this path comes about due to the systematic conduct of policy or due to policy shocks. Equation (5) leverages this logic, looking for a combination of date-0 policy shocks that results in the counterfactual policy rule (4) holding both at $t = 0$ *and also in expectation* at $t = 1$. In response to these well-chosen shocks, the private sector behaves *as if* the counterfactual rule (4) had been imposed throughout.

Informational requirements & relation to Sims & Zha. Our identification result implies that, to predict policy rule counterfactuals, the econometrician does *not* need to know the structural equations of the economy; rather, all she needs are impulse responses to policy shocks. In particular, she needs the causal effects of the policy shocks on the variables that enter her counterfactual rule (here i_t and π_t) and on any other outcome variables she is interested in (e.g, y_t). With those causal effects in hand, she can map outcomes under the baseline rule—i.e., impulse responses to some non-policy shock of interest—into counterfactual outcomes by computing impulse responses to $\{\tilde{\nu}_{0,0}, \tilde{\nu}_{1,0}\}$ that solve (5).

We emphasize that this argument inherently relies on knowledge of the dynamic causal effects of both the contemporaneous policy shock $\tilde{\nu}_{0,0}$ as well as the policy news shock $\tilde{\nu}_{1,0}$: it is only with those two that we can actually enforce the counterfactual rule along the entire transition path (which here consists of two time periods). The econometric method of Sims & Zha (1995) instead supposes that the econometrician only has access to the causal effects of one policy shock (e.g., $\nu_{0,t}$). With one shock it is generally not possible to enforce the counterfactual rule contemporaneously and in expectation; instead, the proposal of Sims & Zha is to subject the economy to an initial shock $\tilde{\nu}_{0,0}$ to enforce the counterfactual policy rule at $t = 0$ and then another *surprise* contemporaneous policy shock $\tilde{\nu}_{0,1}$ to also enforce it at $t = 1$. The key difference relative to our construction is that the private-sector block did not at $t = 0$ *expect* the counterfactual policy rule to hold at $t = 1$; rather, the rule only holds at $t = 1$ because of yet another surprise. In other words, under the approach of Sims & Zha, the counterfactual policy rule only holds ex post along the equilibrium transition path, but not in ex ante expectation. As a result, as long as policy at $t = 1$ matters for $t = 0$ decisions, the constructed counterfactual will differ from the true counterfactual $\{\boldsymbol{y}_{\tilde{\phi}}(\varepsilon_0), \boldsymbol{\pi}_{\tilde{\phi}}(\varepsilon_0), \boldsymbol{i}_{\tilde{\phi}}(\varepsilon_0)\}$. We will further elaborate on this connection between our identification result and the empirical methodology of Sims & Zha in Section 2.4.

Discussion & outlook. The identification result sketched in this section is special in two respects: first, it is presented within the context of a particular explicit structural model; and second, since impulse responses to ε_0 are non-zero only for two periods, the result required knowledge of the effects of two policy shocks ($\nu_{0,0}$ and $\nu_{1,0}$). The remainder of this section will state and prove our main identification result in the context of a general class of infinite-horizon models. In terms of our informational requirements, the key change will be that the econometrician now needs to know the causal effects of *all* policy shocks $\{\nu_{\ell,0}\}_{\ell=0}^{\infty}$, rather than just the first two. The economic intuition on the other hand will be exactly the same: the argument will work as the long as the private-sector block of the model depends on the policy rule only through the path of the policy instrument, as was the case here.

2.2 General environment & objects of interest

We consider a linear, perfect-foresight, infinite-horizon economy. Throughout, we will use boldface to denote time paths for $t = 0, 1, 2, \dots$, and all variables are expressed in deviations from the deterministic steady state. The economy is summarized by the system

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

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

w_t and x_t are n_w- and n_x-dimensional vectors of endogenous variables, z_t is a n_z-dimensional vector of policy instruments, ε_t is a n_ε-dimensional vector of exogenous structural shocks, and ν_t is an n_z-dimensional vector of policy shocks.[10] The distinction between w and x is that the variables in x are observable while those in w are not; in particular, x contains the outcomes of interest for our econometrician as well as the arguments of the counterfactual policy rule that she contemplates.[11] The linear maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ summarize the non-policy block of the economy, yielding $n_w + n_x$ restrictions for each t. Our key assumption—echoing the model of Section 2.1—is that the maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ do not depend on the coefficients of the policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$; instead, policy only matters through the path of the instrument z, with the rule (7) giving n_z restrictions on the policy instruments for each t. As in our simple example, entries of the shock vectors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}$ for $t > 0$ should again be interpreted as news shocks. In particular, the policy shock vector $\boldsymbol{\nu}$ collects the full menu of contemporaneous and news shocks to the prevailing policy rule at all horizons, thus generalizing the two-shock

[10] The boldface vectors $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\varepsilon}, \boldsymbol{\nu}\}$ stack the time paths for all variables (e.g., $\boldsymbol{x} = (\boldsymbol{x}_1', \dots, \boldsymbol{x}_{n_x}')'$). The linear maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_\varepsilon\}$ and $\{\mathcal{A}_x, \mathcal{A}_z\}$ are conformable and are all assumed to map bounded sequences into bounded sequences.

[11] For expositional simplicity, we do not include $\boldsymbol{w}$ as an argument of the *baseline* policy rule (7), though doing so would not pose a problem. The key restriction is that the *counterfactual* policy rule only features variables observable to the econometrician.

set-up that we considered in the simple three-equation model.

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

Assumption 1. *The policy rule in* (7) *induces a unique equilibrium.*

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

SCOPE. Our identification results in Section 2.3 and the empirical analysis in Section 3 will apply to *any* structural model that can be written in the general form (6) - (7). As emphasized before, in addition to linearity, the key property of this environment for our purposes is that policy enters the non-policy block *only* through the path $\boldsymbol{z}$ of policy variables; equivalently, in the linearized economy with aggregate risk, policy affects private-sector decisions only through current and expected future 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 standard Lucas-program approach fit into our framework (6) - (7). Such models are routinely linearized, and their linear representation features the separation between policy rule and non-policy block that our theoretical results require. For example, the analysis in Section 2.1 has already illustrated that one particular canonical model environment—the textbook three-equation New Keynesian model—fits into our framework.[12] By the exact same line of reasoning, even workhorse estimated business-cycle models (e.g., Christiano et al., 2005; Smets & Wouters, 2007) as well as recent quantitative HANK or heterogeneous-firm models (e.g., Auclert et al., 2020; McKay & Wieland, 2021; Ottonello & Winberry, 2020) fit into our structure. For example, in HANK-type models, the Euler equation of the representative household is simply replaced by a more general "aggregate consumption function" (e.g., Auclert et al., 2018; Wolf, 2021):

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

where c is consumption, y is income, π is inflation, i is the nominal rate, ε^d is a demand shock, and $\mathcal{C}_\bullet$ are matrices of derivatives of the consumption function. Such models continue to fit

[12] For reference, we in Appendix A.1 write down the model (1) - (3) in the form (6) - (7).

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

While thus clearly relatively general, our framework also has some important limitations. Recall that our two key restrictions on the model are (i) linearity and (ii) the way the policy instrument is allowed to shape private-sector behavior. The separation between policy and non-policy block embedded in (ii) is violated in some structural models. Important examples are environments that feature an asymmetry of information between the policymaker and the private sector (e.g. Lucas, 1972). In such models, private-sector agents solve a filtering problem, and in general the coefficients of the policy rule will matter for this filtering problem both through the induced movements of the policy instrument *and* through the information contained in those movements. The separation between the private-sector and policy blocks of the model at the heart of our results will thus break down—that is, the coefficients in $\mathcal{H}_x$ depend directly on the policy rule (see Appendix A.2 for a formal derivation).

As we discuss in Appendix A.8, the linearity restriction (i) on the other hand is not conceptual, but practical. By linearity, the effects of policy are sign-, size-, and state-invariant. Given certainty equivalence, we can focus on expected policy instrument paths, thus substantially reducing the informational requirements of our identification results and facilitating their empirical application.[13] The costs of linearity are twofold. First, our identification results will generally not yield *globally* valid policy counterfactuals. Second, we will be able to construct counterfactuals for alternative policy rules that change the policymaker's response to aggregate perturbations (e.g., different Taylor rules), but our results are unlikely to apply to policies that change the model's steady state (e.g., changes in the inflation target).

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

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

[13] To be clear, what we are requiring is linearity of the non-policy block (6). Non-linearity of the policy (e.g., due to a binding zero lower bound), on the other hand, poses no particular challenge. This point is discussed further in Appendix A.9.

This alternative policy rule is also assumed to induce a unique equilibrium. We will discuss further in Section 2.3 what happens if this assumption is violated.

Assumption 2. *The policy rule in* (8) *induces a unique equilibrium.*

We emphasize that the arguments of the counterfactual policy rule are macroeconomic *observables* x and z; naturally, our empirical identification result will not allow evaluation of counterfactual rules that directly involve unobservable objects.[14] Given this alternative rule $\tilde{\mathcal{A}}$, we ask: what are the dynamic response paths $\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ to some given exogenous non-policy shock path $\boldsymbol{\varepsilon}$?

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

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^{n_x} \lambda_i \boldsymbol{x}_i' W \boldsymbol{x}_i \tag{9}$$

where i indexes the n_x distinct (again observable) macro aggregates collected in x, λ_i denotes policy weights, and $W = \text{diag}(1, \beta, \beta^2, \cdots)$ allows for discounting.[15] As for our general counterfactual rule, we assume that the optimal policy problem has a unique solution.

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

We are then interested in two questions. First, what rule is optimal for a policymaker with preferences as in (9)? Second, given that optimal rule $(\mathcal{A}_x^*, \mathcal{A}_z^*)$, what are the corresponding dynamic response paths $\boldsymbol{x}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\mathcal{A}^*}(\boldsymbol{\varepsilon})$ for a given non-policy shock path $\boldsymbol{\varepsilon}$?

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

[14] For example, the counterfactual rule cannot depend on the natural rate of interest, though it could of course depend on an estimate of the natural rate based on observables.

[15] We emphasize that our results are completely silent on the *shape* of the loss function, with structural modeling still the most natural way of obtaining a mapping from observables to welfare. We instead take as given the loss function and ask what kind of empirical evidence would be most useful to figure out how to minimize the loss. We furthermore note that our focus on a separable quadratic loss functions is in line with standard optimal policy analysis, but not essential. As shown in Appendix A.3, our results extend to the non-separable quadratic case, where the loss is now given by $\frac{1}{2}\boldsymbol{x}'Q\boldsymbol{x}$ for a weighting matrix Q.

we would like to predict how the change in policy rule would affect the unconditional second-moment properties of the observed macroeconomic aggregates x.

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

2.3 Identification results

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

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

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

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

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

These informational requirements are the natural generalization of those for the simple model in Section 2.1. First, since we are now considering an infinite-horizon economy, any given shock generates entire *paths* of impulse responses, corresponding to the rows of the Θ's.

Second, rather than two policy shocks, we now need to know causal effects corresponding to the full menu of possible contemporaneous and news shocks $\boldsymbol{\nu}$—all columns of the Θ_ν's.

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

Proposition 1. *Under Assumptions 1 and 2, we can recover the policy counterfactuals $\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ for a counterfactual rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$ as*

$$\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) = \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) \equiv \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{x,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}} \tag{11}$$

$$\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon}) = \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}) \equiv \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon}) + \Theta_{z,\nu,\mathcal{A}} \times \tilde{\boldsymbol{\nu}} \tag{12}$$

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

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

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

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

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

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

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

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

Proposition 1 implies that we can recover the desired policy counterfactual as a function of observables alone—our econometrician needs to know the policy shock causal effect matrices $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ and the effects of the shock $\boldsymbol{\varepsilon}$ under the baseline rule, $\{\boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon}), \boldsymbol{z}_{\mathcal{A}}(\boldsymbol{\varepsilon})\}$, but she need *not* know the structural equations of the underlying model. The key equation (13) in Proposition 1 is the infinite-horizon analogue of the bivariate system (5) from our two-period example in Section 2.1. The intuition is as before: since we know how all possible perturbations $\boldsymbol{\nu}$ to the baseline rule affect the variables x and z entering the counterfactual rule, we can always construct a date-0 shock vector $\tilde{\boldsymbol{\nu}}$ that mimics the equilibrium path of z under the new rule. But since the first model block (6) depends on the policy rule *only* via the expected instrument path, the equilibrium allocations under the new counterfactual rule and the perturbed prevailing rule are the same. The only difference relative to the simple two-period model is that, because we now consider an infinite-horizon setting, we in general require evidence on contemporaneous and all possible future news shocks to the baseline rule in order to be able to mimic an arbitrary alternative policy rule.[16]

What happens if Assumption 2—which maintains that the counterfactual rule delivers a unique equilibrium—is violated? We can distinguish two cases. First, if no equilibrium exists under the contemplated counterfactual policy rule, then the system (13) will simply not have a solution. Second, if multiple equilibria exist, then impulse responses to any $\{\boldsymbol{\varepsilon}, \tilde{\boldsymbol{\nu}}\}$ where $\tilde{\boldsymbol{\nu}}$ solves (13) will be *a* valid equilibrium for the counterfactual rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$. For example, in the simple New Keynesian model of Section 2.1, applying our identification results for the counterfactual rule $\tilde{\phi} = 0$—i.e., a nominal interest rate peg—would deliver the economy's fundamental (minimum state variable, or MSV) equilibrium.

Optimal policy. A very similar argument applies for optimal policy analysis.

Proposition 2. *Consider a policymaker with loss function* (9)*. Under Assumptions 1 and 3, for any $\boldsymbol{\varepsilon}$, the solution to the optimal policy problem is implemented by the rule $\{\mathcal{A}_x^*, \mathcal{A}_z^*\}$ with*

$$\mathcal{A}_x^* = \left(\lambda_1 \Theta'_{x_1,\nu,\mathcal{A}} W, \lambda_2 \Theta'_{x_2,\nu,\mathcal{A}} W, \ldots, \lambda_{n_x} \Theta'_{x_{n_x},\nu,\mathcal{A}} W\right), \tag{16}$$

$$\mathcal{A}_z^* = \mathbf{0}. \tag{17}$$

[16] While Proposition 1 applies to a particular shock path $\boldsymbol{\varepsilon}$, it is immediate that the exact same argument also applies to a particular *historical scenario* (Antolin-Diaz et al., 2021): a historical scenario is simply a set of forecast paths $\boldsymbol{x}_{\mathcal{A}}$ and $\boldsymbol{z}_{\mathcal{A}}$ at a given point in time, and we can use the logic of Proposition 1 to recover the analogous counterfactual historical scenario $\boldsymbol{x}_{\tilde{\mathcal{A}}}$ and $\boldsymbol{z}_{\tilde{\mathcal{A}}}$.

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

Proof. The solution to the policy problem is characterized by the following conditions:

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

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

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

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

where φ_z is the multiplier on (7). It follows from (24) that $\boldsymbol{\varphi}_z = \mathbf{0}$. Then (21) - (23) together with (6) determine the same unique solution for $\{\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}\}$ as before, and $\boldsymbol{\nu}^*$ adjusts residually to satisfy (7). The original problem and the alternative problem are thus equivalent. Next note that, by Assumption 1, we can re-write the alternative problem's constraint set as

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

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

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

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

Proposition 2 reveals that, in conjunction with a given policymaker loss function, the information required to construct valid counterfactuals for arbitrary policy rules also suffices to characterize *optimal* policy rules.[17] The intuition is exactly as before: since we know the causal effects of every possible policy perturbation $\boldsymbol{\nu}$ on the policymaker targets $\boldsymbol{x}$, we in particular know the space of those targets that is implementable through policy actions. At an optimum, we must be at the point of this space that minimizes the policymaker loss. As before, it does not matter whether this optimum is attained through some systematic policy rule or through shocks to an alternative rule.

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

In practice, however, researchers may not be able to isolate all distinct aggregate non-policy shocks. Our third identification result states that, in some cases, it is nevertheless possible to recover the desired counterfactual second-moment properties. Since the result requires some investment in additional notation, we only state the main idea here and relegate further details to Appendix A.5. The key assumption allowing us to make progress is "invertibility": we need to assume that the structural vector moving average representation of the observable data x and z under the baseline policy rule is invertible with respect to the structural shocks driving the economy. This assumption, while restrictive (Plagborg-Møller & Wolf, 2022), is routinely imposed in conventional structural vector autoregression analysis (Fernández-Villaverde et al., 2007). Under this assumption, researchers need not be able to separately observe all of the individual structural shocks; instead, it suffices to simply apply

[17] By certainty equivalence, the results from our perfect-foresight analysis readily extend to *stochastic* linear-quadratic control problems. We can in that case re-write the derived optimal policy rule as a forecasting targeting rule (Svensson, 1997):

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

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

our counterfactual prediction results in Propositions 1 and 2 to the Wold innovations and then collect the results in the form of a counterfactual vector moving average. Appendix A.5 also discusses why this argument fails in the non-invertible case.

Role of the baseline rule $\{\mathcal{A}_x, \mathcal{A}_z\}$. All identification results in this section were stated using the causal effects of policy shocks $\boldsymbol{\nu}$ *relative to the baseline policy rule* $\{\mathcal{A}_x, \mathcal{A}_z\}$. We would like to emphasize, however, that this baseline rule only plays a limited role in our analysis, and that it in particular does not need to be known by the econometrician.[18]

In our proofs of Propositions 1 and 2, the baseline policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$ functions as a reference point: we find the sequence of policy shocks relative to that rule that implements the desired counterfactual rule. This choice of reference point, however, is ultimately immaterial: since private-sector behavior in (6) is shaped by policy only through the instrument path $\boldsymbol{z}$, all that matters for our results is knowledge of how macroeconomic outcomes $\boldsymbol{x}$ are related to paths of policy instruments $\boldsymbol{z}$. For example, under the natural assumption that $\Theta_{z,\nu,\mathcal{A}}$ is invertible—i.e., the policymaker can implement any sequence of the policy instrument—we could post-multiply all causal effect matrices by $\Theta_{z,\nu,\mathcal{A}}^{-1}$, thus writing policy causal effects not in terms of shocks $\boldsymbol{\nu}$ relative to a given rule, but instead directly in terms of instrument paths $\boldsymbol{z}$. This change in reference point leaves our identification results completely unchanged, but will prove useful when later connecting those theoretical identification results with empirical evidence on policy shock propagation in Section 3.

Discussion. The theoretical identification results in Propositions 1 and 2 offer a bridge between the "Lucas program" (e.g., see Christiano et al., 1999)—a strategy that relies on micro-founded structural models to form policy counterfactuals—and the purely empirical approach of Sims & Zha (1995). The propositions reveal that, under our assumptions, impulse responses to policy *shocks*—objects that are estimable using semi-structural empirical techniques—suffice to predict the effects of changes in systematic policy *rules*. Key to our argument is the use of *multiple distinct* policy shocks. By using many such shocks (and all realized at date 0), counterfactual rules can be imposed not just ex post but also in ex ante expectation, and this turns out to be enough to circumvent the Lucas critique. We further elaborate on the connection between our results and the approach of Sims & Zha—which

[18] Moreover, our results will continue to hold if the baseline policy rule underwent changes during the sample period. For our purposes, the key requirement is that the private-sector behavioral relationships (6) have remained stable over the observed sample period. We provide further details in Appendix A.4.

uses one policy shock, set to a new level at each date t—in Section 2.4.

Our results can be interpreted as part of the recent effort to bring insights from the "sufficient statistics" approach popular in public finance to macroeconomics (Chetty, 2009; Nakamura & Steinsson, 2018). For a large family of structural models and policy counterfactuals, policy shock impulse responses are sufficient statistics in the sense that we can directly use them to compute the desired counterfactuals, without actually requiring knowledge of the structural equations of the model. To leverage Propositions 1 and 2, an econometrician does not need to make detailed assumptions on the private-sector block, nor does she need to know the policy rule that generated the observed data.

2.4 Illustration & relation to Sims & Zha (1995)

This section provides a visual illustration of our identification results and their relationship to the approach of Sims & Zha (1995). As our laboratory we use a HANK model as in Wolf (2021), with details of the parameterization relegated to Appendix A.1. In this environment we will compute policy counterfactuals in multiple ways: first by using the actual structural equations of the model to simply solve the model with a counterfactual policy rule; and then by using model-implied impulse responses to policy shocks to implement either the approach of Sims & Zha or our identification result in Proposition 1.

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

$$i_t = \phi_\pi \pi_t + \sum_{\ell=0}^{\infty} \nu_{\ell,t-\ell} \tag{28}$$

for $\phi_\pi = 1.5$. In particular, we recover a) the impulse responses $\{\boldsymbol{x}_\mathcal{A}(\boldsymbol{\varepsilon}), \boldsymbol{z}_\mathcal{A}(\boldsymbol{\varepsilon})\}$ to a contractionary cost-push shock ε_t under (28) and b) the causal effects of contemporaneous and news policy shocks $\boldsymbol{\nu}$ to (28), $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$. We emphasize that those causal effects would be estimable by an econometrician living in this economy and with access to valid instruments for the cost-push shock ε_t as well as the policy shocks $\{\nu_{0,t}, \nu_{1,t}, \dots\}$.

We entertain the following counterfactual policy rule:

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

for $\phi_i = 0.9$, $\phi_\pi = 2$, $\phi_y = 0.5$. The dotted and solid lines in all three panels of Figure 1 show the true model-implied impulse responses of output and inflation to a cost-push shock ε_t under the baseline rule (28) (dotted) and the counterfactual rule (29) (solid), where both

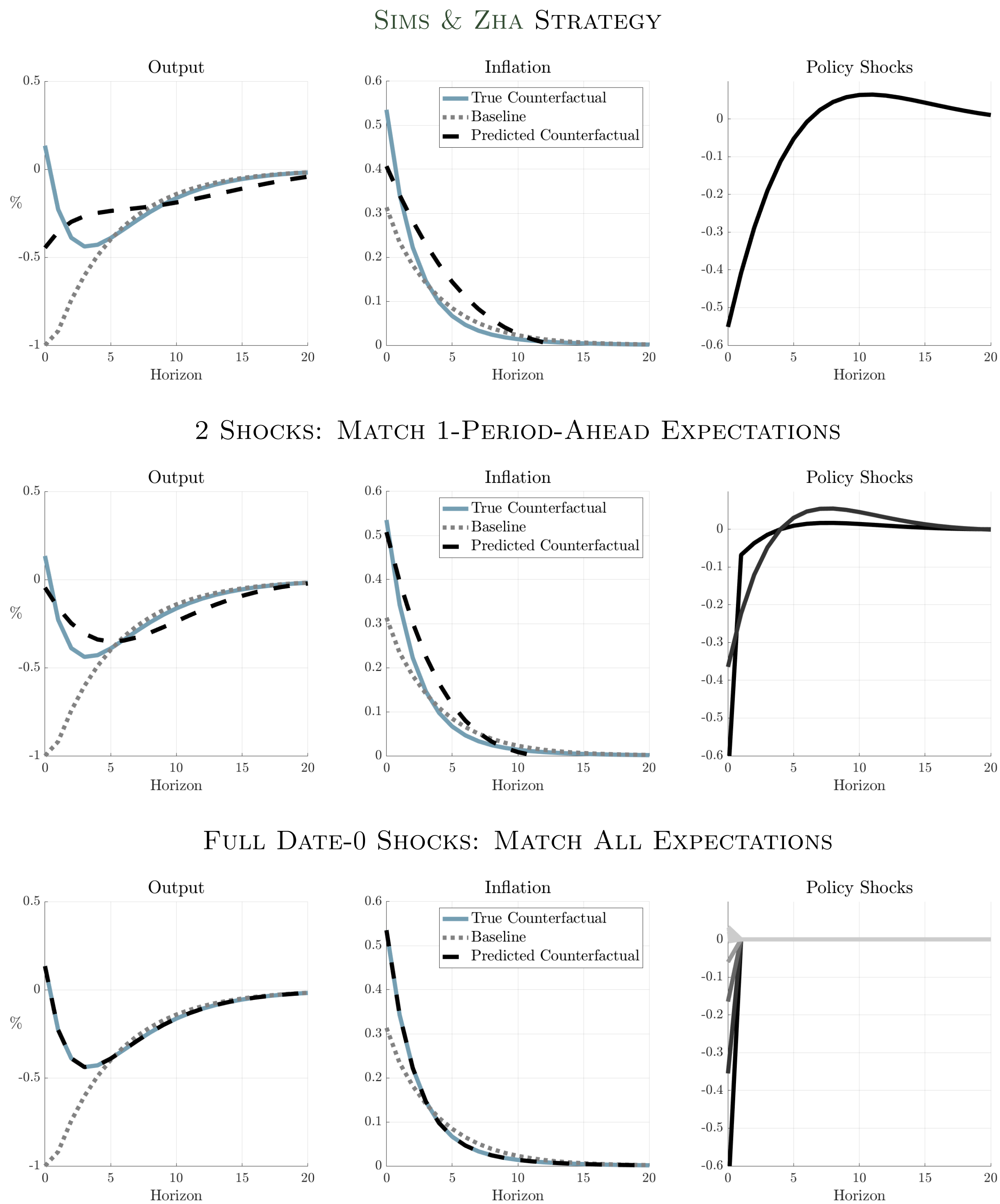


Figure 1: The dotted and solid lines show output and inflation responses to the cost-push shock ε_t under the policy rules (28) and (29) in the HANK model. The dashed lines give counterfactuals constructed through the policy shocks on the right. The top panel uses repeated realizations of a single policy shock to enforce (29) ex post, as in Sims & Zha. The middle panel uses repeated realizations of two policy shocks to enforce (29) ex post and in one-period-ahead expectation. The lower panel shows our method, which uses a single realization of many policy shocks to enforce (29) along the entire expected path. Lighter shades correspond to news about policy at longer horizons.

of these lines are computed from the structural equations of the model.

We now seek to recover the desired counterfactual (solid) *only* through knowledge of the dynamic causal effects of policy shocks, and *without* actually relying on any of the structural equations of the model. The panels of Figure 1 show results for three possible strategies to predict the counterfactual propagation of the cost-push shock.

Estimand of Sims & Zha. The top panel begins with the empirical strategy of Sims & Zha (1995). Here the econometrician was only able to estimate the dynamic causal effects of the first entry of $\boldsymbol{\nu}$ (i.e., the contemporaneous shock $\nu_{0,t}$), and then uses a sequence of such policy shocks—one at each $t = 0, 1, 2, \ldots$—to enforce the counterfactual rule (29) ex post along the equilibrium transition path. The right panel shows the sequence of policy shocks that implements this strategy, and the dashed lines in the left and middle panels give the responses of output and inflation to the original cost-push shock *plus* the derived sequence of monetary policy shocks. The main takeaway is that those dashed lines are not equal to the true counterfactual (solid). Intuitively, the issue is that the contemplated counterfactual rule is only imposed ex post, but not in ex ante expectation. Since expectations about the future affect the present, enforcing the rule through ex post surprises is not the same as switching and committing to a different rule from time $t = 0$ onwards. Visually, the importance of ex post surprises is evident in the right panel: to map the baseline rule into the counterfactual rule, the econometrician requires a *sequence* of expansionary policy shocks $\nu_{0,t}$, with those shocks remaining large throughout the entire first year after the shock.

Towards our identification result. The middle and bottom panels now illustrate the logic of our identification result—with multiple policy shocks, the econometrician has enough degrees of freedom to impose the counterfactual rule not just ex post, but also in expectation. As a warmup, the middle panel considers a case in which the econometrician is able to estimate the causal effects of the first two entries of $\boldsymbol{\nu}$ (i.e., a contemporaneous and a one-period forward guidance policy shock). Such access to multiple shocks suggests a natural generalization of Sims & Zha: use the two policy shocks at each $t \geq 0$ to enforce the desired counterfactual rule not only ex post (as Sims & Zha do with one shock), but also in ex ante expectation for the next period.[19] Since the counterfactual policy rule is now imposed both ex post and in ex ante expectation for one period, the predicted counterfactuals (dashed) are closer to the truth (solid); correspondingly, the policy shock sequences in the right panel

[19]We present implementation details for this approach in Appendix A.7.

feature smaller ex post surprises dated $t = 1, 2, \ldots$. The bottom panel—which corresponds to our identification result—simply continues this logic. With access to the causal effects of the full vector of policy shocks $\boldsymbol{\nu}$, the econometrician can rely purely on date-0 shocks (right panel) to enforce the counterfactual rule not just ex post but also in ex ante expectation. Under our assumptions, doing so suffices to circumvent the Lucas critique and recover the correct counterfactual (left and center panels).

To summarize, the top- and bottom-right panels illustrate the core difference between the empirical method of Sims & Zha and our identification result. In the former, the researcher has access to a single policy shock, and uses a sequence of realizations of that shock to enforce the counterfactual rule. In our approach, the researcher has access to many shocks and only uses shocks at date-0 to enforce the counterfactual rule. Our identification result thus clearly has substantially higher informational requirements, but this increase in information brings with it the similarly substantial benefit of robustness to Lucas critique concerns.

2.5 Discussion

The central takeaway from the analysis in this section is that—under our maintained structural assumptions—systematic policy rule counterfactuals can, at least in principle, be constructed purely through empirical measurement, and in a way that is robust to Lucas critique concerns. In the remainder of the paper we discuss how to operationalize our insights. The main challenge is that the informational requirements underlying our identification results are quite high: the researcher needs evidence on the causal effects of a full menu of policy shocks that shift expectations of policy *at all possible horizons*. Section 3 presents an empirical strategy for the relevant case of researchers with access to only a few distinct identified policy shocks. We will then in Section 4 illustrate this empirical strategy through several applications to systematic monetary policy rule counterfactuals.

3 Empirical method

This section presents our empirical method for constructing policy rule counterfactuals with evidence on multiple, but a limited number of, distinct policy shocks. Section 3.1 illustrates the basic logic of our method with an illustrative example based on the oil shock application of Bernanke et al. (1997). Section 3.2 then introduces the general methodology.

Throughout, the discussion in this section will leverage the following connection between our theoretical identification results in Section 2.3 and empirical evidence on policy shock propagation. For our theoretical analysis, we found it convenient to think of contemporaneous and news shocks $\boldsymbol{\nu}$ that perturb some fixed prevailing policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$ horizon by horizon. For connecting to data, however, this perspective is less useful—empirical evidence on policy shock causal effects just gives impulse responses and is generally silent on the underlying policy rule. Instead, a more instructive way forward is to realize that the informational requirements underlying our identification results could equivalently be phrased in terms of *policy instrument paths*, as already discussed in Section 2.3: to implement our results, the econometrician needs to know the causal effects associated with all possible time paths of the policy instrument $\boldsymbol{z}$. Empirical work that studies a given identified policy shock simply gives us the dynamic causal effects associated with a *particular* path of the policy instrument, without any reference to the underlying policy rule, to whether a policy shock is contemporaneous or "news", and in fact without even requiring stability of that underlying rule. The basic idea of our empirical method is to combine those instrument time paths to mimic the effects of a switch to a counterfactual policy rule.

3.1 Illustrative example

To illustrate the basic logic of our proposed empirical method as transparently as possible we begin with a stylized example that emulates the monetary policy counterfactual analysis of Bernanke et al. (1997). Like those authors, we consider an econometrician that wishes to predict the (counterfactual) propagation of oil price shocks in the absence of a monetary policy reaction—i.e., the canonical "zeroing-out" policy counterfactual.[20]

Revisiting Bernanke et al. (1997). Figure 2 provides a stylized representation of how the econometrician could use our identification result to construct her desired oil shock counterfactual. We emphasize that the impulse responses in this figure are purely illustrative; they do not come from any empirical analysis or structural model.

As a first step, the econometrician begins by estimating the effects of an oil price shock under the prevailing monetary reaction function, exactly as in Bernanke et al. (1997). In the stylized example here, the oil shock leads to an increase in prices (top-left panel); the

[20]In notation of Section 2, such "zeroing-out" corresponds to a counterfactual policy rule that sets $\boldsymbol{z} = \mathbf{0}$. It is of course well-known that rules of this sort—for example a nominal interest rate peg—often lead to equilibrium indeterminacy, violating Assumption 2 (Sargent & Wallace, 1981). As discussed in Section 2.3, the counterfactuals presented here should thus be interpreted as corresponding to one particular equilibrium associated with this policy rule.

Stylized representation of the empirical method

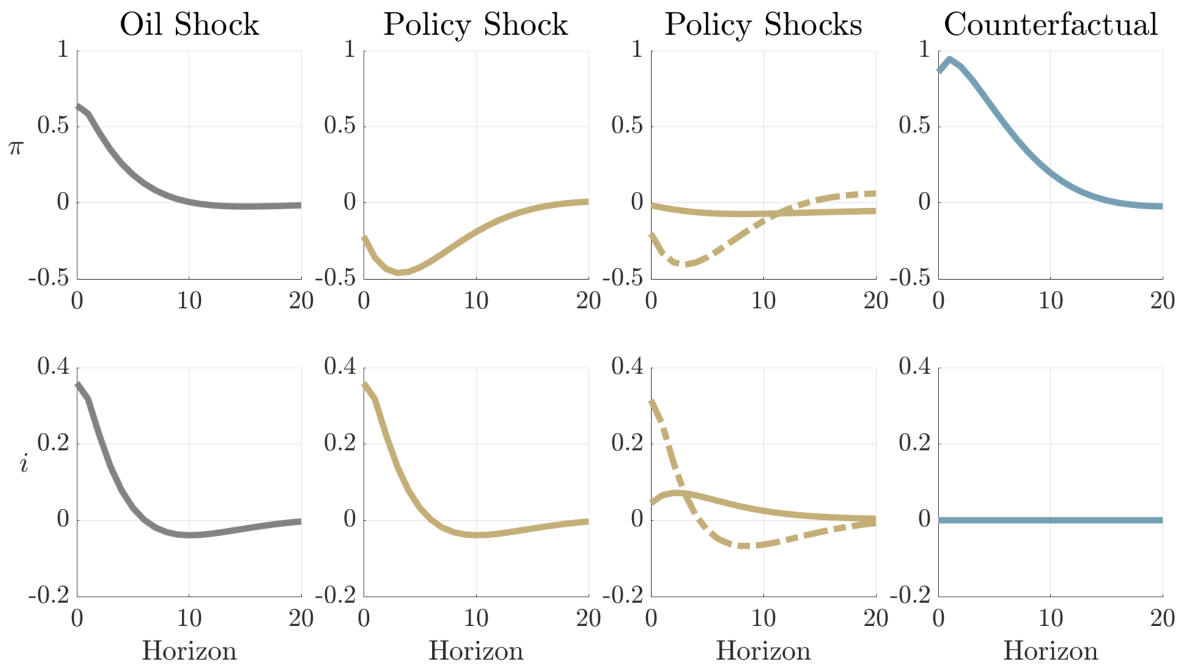


Figure 2: Inflation (π) and nominal interest rate (i) impulse responses to: oil shock under the baseline rule (left panel, solid); monetary policy shocks to the baseline rule (two middle panels, solid and dashed); and oil shock under the counterfactual rule (right panel, solid). All impulse responses are purely illustrative; they do not come from any empirical exercise or structural model.

monetary authority furthermore leans against this inflationary pressure through an increase in nominal interest rates (bottom-left panel). By our identification result, she next needs to estimate the effects of a monetary policy shock—or a linear combination of such policy shocks—that moves nominal interest rates *from date*-0 *onwards* exactly like the observed endogenous interest rate response to the oil shock. The two middle panels show two possible scenarios. In the left one, the econometrician was able to identify a single monetary policy shock that induces the exact same path of nominal interest rates as the oil shock. In the right one, she estimated two separate policy shocks (one solid, one dashed), with the sum of the two replicating the interest rate path after the oil shock. In both cases, the identified policy shocks decrease inflation (top panels). Given either of these estimates, the econometrician can apply our identification result: she simply needs to subtract the impulse responses shown in the second or third column from those in the first column. The results are then shown in the fourth column: interest rates are now by construction unresponsive, and inflation increases by more than under the baseline policy response. It follows from Proposition 1 that any structural model consistent with (i) our general model framework (6) - (7), (ii) the original propagation of the oil shock (first column) and (iii) either one of the two middle columns on monetary policy shock propagation will necessarily agree with this "zeroing-out" counterfactual displayed in the right panel.

Discussion. We emphasize that the illustrative example displayed in Figure 2 is stylized in two ways. First, using either of the estimated monetary policy shocks, the econometrician was able to *perfectly* enforce the desired policy counterfactual using only date-0 shocks. In actual applications this will not be possible in general, so approximations will be needed. Second, the counterfactual rule that we considered was particularly simple, taking the form of an exogenous interest rate path rather than a more complicated relationship between endogenous equilibrium outcomes (like, e.g., a Taylor rule). Our empirical method, presented in the next section, is the natural generalization of the stylized example: the researcher considers an arbitrary counterfactual rule of our general form (8), and then enforces it *as well as possible* using the available policy shock evidence.

3.2 Counterfactuals with a limited number of policy shocks

We consider a researcher that has access to estimates of n_s distinct policy shocks associated with n_s distinct response paths of the policy instrument z.[21] We denote the causal effects of these shocks by $\{\Omega_{x,\mathcal{A}}, \Omega_{z,\mathcal{A}}\}$, where each of the n_s columns of the Ω's gives the impulse response to a distinct identified policy shock. Given these lower-dimensional causal effect maps, and given a non-policy shock $\boldsymbol{\varepsilon}$ and a counterfactual rule $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$, the proof strategy of Proposition 1 will fail in general. We would now need to set

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

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

Our proposal is to simply select the weights $\boldsymbol{s}$ on the n_s date-0 shocks to enforce the

[21] In saying that a researcher has access to policy shocks that induce different instrument paths, we are implicitly assuming that these differences in instrument paths reflect different identification strategies capturing different linear combinations of policy shocks rather than statistical noise or violations of the identifying assumptions. We justify this interpretation in our empirical application in Section 4.

desired counterfactual rule *as well as possible.* In practice, this means solving the problem

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

The output of the simple problem (31) is the best approximation to the desired policy counterfactual within the space of empirically identified policy shock paths. By our identification results in Section 2 and because all shocks are dated $t = 0$ (i.e., no ex post surprises), this approach is robust to the Lucas critique. In the illustrative example of Figure 2, the available evidence on policy shocks in the middle panels was sufficient to set the argument of (31) exactly to zero. In actual applications, on the other hand, we will not perfectly enforce the desired policy counterfactual; rather, we will approximate it as closely as possible. The richer the menu of policy shocks we have access to, the better the approximation will become, eventually converging to the truth (as $n_s \to \infty$). The important limitation of our approach is thus that, for small n_s, it will not always be possible to construct an accurate approximation of the desired counterfactual rule—sometimes we will be able to set the implementation error in (31) close to zero, other times it will be large. The practical usefulness of our proposed method is thus an inherently application-dependent question.

By Proposition 2, our identification results also allow researchers to learn about *optimal* counterfactual policy rules, given some exogenously specified loss function. Appendix B.2 shows how to apply our Lucas critique-robust method to such questions of optimal policy design. Very briefly, the idea is to use date-0 policy shocks to reduce the policymaker loss as much as possible. Our approach thus minimizes the loss function by perturbing the baseline policy response in directions spanned by the set of empirically identified policy shocks.[22] Finally, for both rule counterfactuals and for optimal policy, we in Appendix B also describe how to leverage our results to construct counterfactual average business-cycle statistics.

4 Application to monetary policy counterfactuals

This section applies our empirical method to construct monetary policy rule counterfactuals. We proceed in two steps. First, in Section 4.1, we provide a brief review of existing evidence

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

on monetary policy shock transmission—the key input to our empirical method. Second, in Section 4.2, we apply our method to study the propagation of investment-specific technology shocks under various counterfactual monetary rules.

4.1 A review of monetary policy shock evidence

In order to implement our empirical method, we require evidence on multiple distinct monetary policy shocks that induce different time paths for nominal interest rates. The empirical literature has devised many different strategies to isolate quasi-random variation in the conduct of monetary policy (see Ramey, 2016, as well as the discussion below). Since monetary authorities affect current and future expected interest rates, monetary policy is inherently multi-dimensional, and so it is not surprising that distinct identified policy shocks capture different dimensions of policy: some identification schemes will capture transitory impulses, while others reflect more persistent deviations from the policy rule.[23] The empirical evidence that we leverage is consistent with this observation.

Our applications in Section 4.2 will use two of the most canonical monetary policy shock series: those of Romer & Romer (2004) and Gertler & Karadi (2015). Importantly, those two monetary shock series are likely to be informative about very different monetary experiments. While the Romer & Romer shock is rather short-lived (i.e., mostly reflecting contemporaneous shocks $\nu_{0,t}$), the Gertler & Karadi shock is well-known to move longer-term nominal interest rates and is thus more likely to have a larger forward guidance component (i.e., in greater part reflecting $\nu_{\ell,t}$ for $\ell > 0$). Our applications in the next section reveal that even this relatively modest amount of evidence is in fact enough to tightly characterize several important monetary policy rule counterfactuals.

While we have chosen to focus on the most well-known and well-understood policy shock series for our main applications, we emphasize that similar arguments about interest rate time profiles apply just as well to several other popular monetary policy shock series. First, as we discuss in detail in Appendix C.4, the monetary shock series of Miranda-Agrippino & Ricco (2021) and Aruoba & Drechsel (2022)—shock measures that seek to improve on the original series of Romer & Romer and Gertler & Karadi in various ways—induce similar dynamics, with one shock more transitory and the other more persistent. Second, some prior work has explicitly split monetary policy shock series by their effects on different points of

[23] A related argument was made by Sims (1998): there is no need for different identification strategies to yield correlated measures of policy shocks, simply because the identified shocks may capture different sources of variation in policy. We thank our discussant Valerie Ramey for pointing out that connection.

the yield curve, leveraging the intuitive idea that no two monetary policy surprises are likely to shift the overall yield curve in exactly the same way. Estimates of this type are for example presented in Gürkaynak et al. (2005), Antolin-Diaz et al. (2021), and Inoue & Rossi (2021), and would offer natural alternatives as an input to our empirical method.[24]

4.2 Counterfactual policy rule exercises

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

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

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

The second input are the causal effects of a menu of different monetary policy shocks. For this we consider the shock series of Romer & Romer (2004) and Gertler & Karadi (2015), as already discussed in Section 4.1. To correctly account for joint uncertainty in the estimation of the effects of the two policy shocks, we study their propagation through a single VAR. For robustness, we also repeat all of our policy counterfactual applications with the shock series of Miranda-Agrippino & Ricco and Aruoba & Drechsel—two less well-known but arguably somewhat more robust shock series—and find similar results. All results for these alternative shock measures are reported in Appendix C.4.

[24]We note that this discussion also extends to fiscal shocks. For government spending, Ramey (2011) explicitly distinguishes between shocks reflecting gradual military build-ups and more transitory upticks in purchases. For taxes, Mertens & Ravn (2014) separate unanticipated (transitory) and anticipated (gradual) tax shocks. We leave applications of our methodology to fiscal policy counterfactuals to future work.

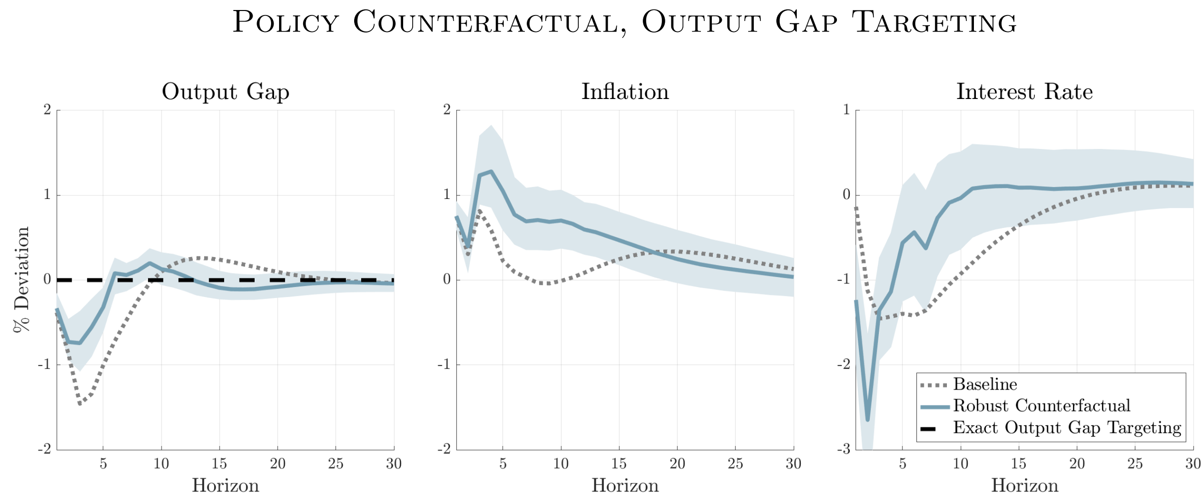


Figure 3: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted) and the best feasible approximation to output gap targeting (solid), computed following (31). The shaded areas correspond to 16th and 84th percentile confidence bands. Perfect output gap targeting (i.e., $\widehat{y}_t = 0$ for all t) is displayed as the black dashed line.

Counterfactual policy results. We use our methodology to construct counterfactuals for several different alternative monetary policy rules: output gap targeting; a standard Taylor (1999) rule; a nominal rate peg; nominal GDP targeting; and the optimal policy rule corresponding to a loss function with equal weight on the output gap and a weighted average of current and lagged inflation (i.e., average inflation targeting).

First, Figure 3 shows our counterfactual results for output gap stabilization. The identified investment technology shock has both a cost-push as well as a negative demand component, consistent with theory (e.g., see Justiniano et al., 2010). Under the baseline policy rule (dotted), nominal interest rates are cut relatively aggressively, though not by enough to stabilize the output gap; furthermore inflation stays moderately above target.[25] Under our approximation to output gap targeting, nominal interest rates are cut much more aggressively, essentially stabilizing the output gap from around a couple of quarters after the shock, at the cost of persistently higher inflation. Given the well-documented lags in monetary policy transmission, it seems unlikely that *any* nominal interest rate path could actually stabilize the output gap in the immediate aftermath of the investment shock; we thus believe

[25] To the extent that our sample saw changes in the systematic conduct of monetary policy, the displayed impulse responses will average over the in-sample observed monetary reactions to the investment shock. See the discussion in Appendix A.4 for further details.

Policy Counterfactual, Taylor Rule

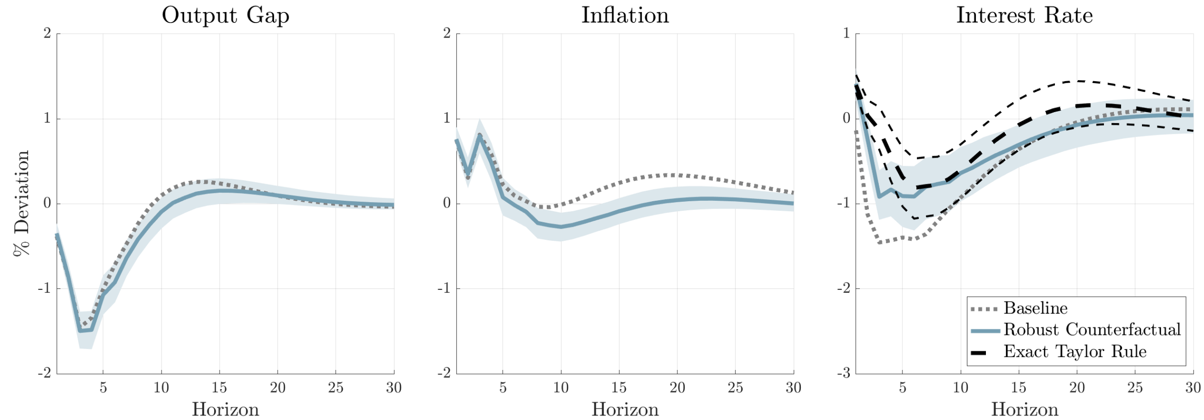


Figure 4: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing rule (dotted) and the best feasible approximation to a simple Taylor-type rule $\widehat{i}_t = 0.5\widehat{i}_{t-1} + 0.5 \times (1.5\widehat{\pi}_t + \widehat{y}_t)$ (solid), computed following (31). The shaded areas correspond to 16th and 84th percentile confidence bands. The distance between dashed and solid lines in the right panel is the implementation error (i.e., the argument of (31)).

that our empirical analysis yields an accurate approximation to what a strict output gap targeting policy can actually achieve in practice.[26]

Second, Figure 4 shows the results for a Taylor-type rule with strong responses to inflation and the output gap as well as moderate nominal interest rate smoothing. Due to the observed increase in inflation, this policy rule actually dictates a much less aggressive rate cut, resulting in somewhat lower output and inflation at medium horizons. In the right panel, the distance between the dashed and solid lines indicates whether or not our method is able to accurately implement the counterfactual rule. While the solid lines show our counterfactual path of nominal interest rates, the dashed lines instead use the counterfactual Taylor rule to map the output gap and inflation paths shown in the left and middle panels into paths of nominal rates. The distance between the solid and dashed lines is the argument of (31)—i.e., the policy rule implementation error. We see that the contemplated counterfactual Taylor rule is imposed relatively well throughout, except at a couple of quarters after the initial shock (where interest rates are still cut by too much relative to the Taylor rule prescription).

Third, we proceed in the spirit of the recent change in the Federal Reserve's policy framework and consider a policymaker with preferences over output and *average* inflation

[26] In the notation of Section 2, these statements correspond to the idea that perfect output gap targeting—i.e., the rule $\boldsymbol{y} = 0$, with y denoting the output gap—is not implementable (i.e., Assumption 2 is violated).

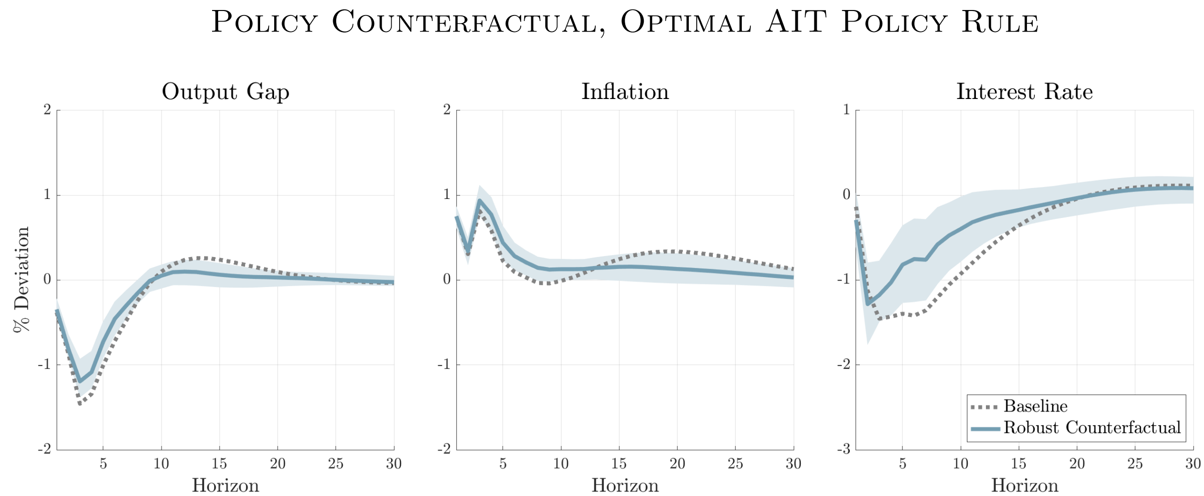


Figure 5: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted) and the best feasible approximation to an optimal average inflation targeting monetary policy rule (solid), computed as discussed in Appendix B.2. The shaded areas correspond to 16th and 84th percentile confidence bands.

$\bar{\pi}_t$, where $\bar{\pi}_t = \sum_{\ell=0}^{K} \omega_\ell \pi_{t-\ell}$.[27] We then represent the loss function of a dual mandate policymaker with preferences over average inflation as

$$\mathcal{L} = \lambda_\pi \bar{\boldsymbol{\pi}}' W \bar{\boldsymbol{\pi}} + \lambda_y \boldsymbol{y}' W \boldsymbol{y} \tag{32}$$

with $\lambda_\pi = \lambda_y = 1$, $W = \text{diag}(1, \beta, \beta^2, \cdots)$ and $\beta = 1/1.01$. Results for our optimal policy counterfactual are displayed in Figure 5. The key takeaway here is that this optimal policy counterfactual differs very little from actually observed outcomes. In other words, there is little room to improve upon the observed allocation by changing policy within the space of policy instrument paths spanned by our two identified policy shocks.

Appendices C.3 and C.4 present several further applications. First, we consider the two remaining policy counterfactuals: nominal GDP targeting and a nominal interest rate peg. We find that nominal GDP targeting can be implemented very accurately; interestingly, this counterfactual looks quite similar to our estimated outcomes under the baseline rule, with interest rates cut only slightly less aggressively. Matters look different for a nominal interest rate peg, however. Here, nominal rates in our best Lucas critique-robust counterfactual still

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

fall by quite a bit too much, in particular at short horizons. Our method thus in this case does not allow an accurate characterization of the desired counterfactual. Second, we repeat our analysis with the alternative shock series of Miranda-Agrippino & Ricco and Aruoba & Drechsel. Those two shocks give similar impulse responses to our baseline shock measures, and so our systematic policy rule counterfactuals are not affected much.

Counterfactual second-moment properties. While our analysis in this section has focused on policy counterfactuals *conditional* on some given non-policy shock, we have also used our identification results to construct counterfactual *unconditional* business-cycle statistics. Specifically, our object of interest are counterfactual aggregate business-cycle statistics under optimal policy for a policymaker with preferences as in (32). As discussed in Sections 2.3 and 3.2, recovering this counterfactual requires us to apply our policy counterfactual mapping separately to the impulse responses for each reduced-form Wold innovation of the observed macroeconomic aggregates, then stacking the resulting impulse responses into a new counterfactual Wold representation, and finally using this Wold representation to derive counterfactual second moments.

Results for this application are presented in Appendix C.5. Consistent with our "conditional shock" results in Figure 5, we find that—at least within the space of identified policy shock causal effects—only moderate policy improvements would have been feasible, with our constructed counterfactual volatilities of the output gap and inflation only somewhat below the actually observed level.

Discussion. The results from our applications in this section reveal that existing empirical evidence on policy shocks is already sufficient to tightly restrict policy rule counterfactuals for several prominent alternative monetary policy strategies. At the same time, we emphasize that our empirical method is clearly not *always* applicable: for some non-policy shocks and some counterfactual rules, it will not be possible to enforce the counterfactual rule accurately. In particular, the counterfactuals that we constructed for the investment shock application were relatively accurate precisely because the investment shock is rather transitory, thus only requiring knowledge of the effects of similarly transitory interest rate changes, along the lines of those implied by the Romer & Romer and Gertler & Karadi monetary policy shocks (see Appendix C.2 for the exact paths). More persistent non-policy shocks $\boldsymbol{\varepsilon}$ necessarily induce more persistent policy instrument movements and thus would correspondingly require empirical evidence on highly persistent policy shocks (e.g., far-ahead forward guidance).

5 Conclusions

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

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

In closing, we would like to re-iterate two important considerations for researchers who contemplate using our approach. First, our method is silent on issues of equilibrium uniqueness. It will construct *one* valid equilibrium for the counterfactual policy rule, but nothing guarantees uniqueness; for that, additional theoretical arguments are needed. Second, our empirical method relies on linearity and thus should only be used when this assumption is appropriate. Structural modelers often use linearization as a means of computing equilibria; in such structural contexts, the uses and limitations of linear methods are well-understood. Those same principles apply to the use of our method.

References

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

Antolín-Díaz, J. & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for svars. *American Economic Review*, *108*(10), 2802–29.

Arkolakis, C., Costinot, A., & Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, *102*(1), 94–130.

Aruoba, S. B. & Drechsel, T. (2022). Identifying monetary policy shocks: A natural language approach. CEPR Discussion Paper No. DP17133.

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.

Barnichon, R. & Mesters, G. (2023). A Sufficient Statistics Approach for Macro Policy Evaluation. Working Paper.

Bauer, M. D. & Swanson, E. T. (2022). A reassessment of monetary policy surprises and high-frequency identification. Technical report, National Bureau of Economic Research.

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

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

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

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

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

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

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

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

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

Degasperi, R., Hong, S., & Ricco, G. (2020). The global transmission of us monetary policy. CEPR Discussion Paper No. DP14533.

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

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., & Watson, M. W. (2007). ABCs (and Ds) of understanding VARs. *American Economic Review*, *97*(3), 1021–1026.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Mankiw, N. G. & Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. *The Quarterly Journal of Economics*, *117*(4), 1295–1328.

McKay, A. & Wieland, J. F. (2021). Lumpy durable consumption demand and the limited ammunition of monetary policy. *Econometrica*, *89*(6), 2717–2749.

McKay, A. & Wolf, C. (2022). Optimal policy rules in HANK. Technical report, Working Paper, FRB Minneapolis.

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

Mertens, K. & Ravn, M. O. (2014). A reconciliation of SVAR and narrative estimates of tax multipliers. *Journal of Monetary Economics*, *68*, S1–S19.

Miranda-Agrippino, S. & Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, *13*(3), 74–107.

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.

Plagborg-Møller, M. & Wolf, C. K. (2022). Instrumental variable identification of dynamic variance decompositions. *Journal of Political Economy*, *130*(8), 2164–2202.

Ramey, V. (1993). How important is the credit channel in the transmission of monetary policy? In *Carnegie-Rochester Conference Series on Public Policy*, volume 39, (pp. 1–45). Elsevier.

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

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

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

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

Sargent, T. J. & Wallace, N. (1981). Some unpleasant monetarist arithmetic. *Federal Reserve Bank of Minneapolis Quarterly Review*, *5*(3), 1–17.

Sims, C. A. (1998). Comment on Glenn Rudebusch's "Do Measures of Monetary Policy in a Var Make Sense?". *International Economic Review*, *39*(4), 933–941.

Sims, C. A. & Zha, T. (1995). Does monetary policy generate recessions? Working Paper.

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.

Taylor, J. B. (1999). A historical analysis of monetary policy rules. In *Monetary policy rules* (pp. 319–348). University of Chicago Press.

Uribe, M. & Yue, V. Z. (2006). Country spreads and emerging countries: Who drives whom? *Journal of International Economics*, *69*(1), 6–36.

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. (2003). *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, (ii) implementation details for our empirical methodology in Section 3, and (iii) several supplementary findings and alternative experiments complementing our empirical applications in 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 **43**

A.1 Examples of nested models . . . 43
A.2 Filtering problems . . . 46
A.3 More general loss functions . . . 47
A.4 Changing policy rules . . . 47
A.5 Counterfactual second-moment properties . . . 49
A.6 Optimal policy counterfactual in HANK . . . 51
A.7 Counterfactuals with a limited number of shocks . . . 53
A.8 Global identification argument . . . 53
A.9 Non-linear policy rules . . . 56

B Details for empirical method **57**

B.1 Policy rule counterfactuals . . . 57
B.2 Optimal policy counterfactuals . . . 57

C Supplementary details for monetary applications **59**

C.1 Data . . . 59
C.2 Shock & policy dynamic causal effects . . . 60
C.3 Results for omitted monetary policy counterfactuals . . . 61
C.4 Counterfactuals with alternative shock measures . . . 63
C.5 Counterfactual second-moment properties . . . 65

A Supplementary theoretical results

This appendix provides several results complementing our theoretical identification analysis of Section 2. Appendix A.1 discusses examples of macro models that are nested by our results, Appendix A.2 gives an example of a model that is not, Appendix A.3 extends our optimal policy arguments to more general loss functions, Appendix A.4 works out what happens when the policy rule itself is changing, Appendix A.5 provides the details for unconditional second-moment counterfactuals, Appendix A.6 studies optimal monetary policy in our illustrative HANK model, Appendix A.7 shows how we construct counterfactuals with a limited number of policy shocks (as displayed in Figure 1), Appendix A.8 provides a global identification analysis with even higher informational requirements, and finally Appendix A.9 extends our results to the case where only the policy rule is non-linear.

A.1 Examples of nested models

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

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

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

while the Euler equation can be written as

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

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

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

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

Behavioral model. Our general framework (6) - (7) is rich enough to nest popular behavioral models such as the cognitive discounting framework of Gabaix (2020) or the sticky information set-up of Mankiw & Reis (2002). We here provide a sketch of the argument for a particular example—the consumption-savings decision of behavioral consumers. Our discussion leverages sequence-space arguments as in Auclert et al. (2021).

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

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

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

QUANTITATIVE HANK MODEL. The HANK model used for our quantitative illustration in Section 2.4 is a simplified version of that in Wolf (2021). Since the model is standard our discussion here will be relatively brief.

- **Demand block.** The economy is populated by a unit continuum of households that can save in a nominally risk-free, liquid asset. We set the steady-state quarterly real return on the asset to $\bar{r} = 0.01$, and its supply as a share of quarterly output to 1.04 (Kaplan et al., 2018). Households have time-separable log preferences over consumption, with discount factor β. Their total non-asset income is $(1 - \tau_y)e_{it}y_t + \tau_t$. Here y_t is aggregate income and e_{it} is idiosyncratic household productivity, with $\int_0^1 e_{it}di = 1$ at all t and where e_{it} follows the income process of Kaplan et al. (2018), ported to discrete time. Households also receive lump-sum transfers τ_t from the government; we set the steady-state level of transfers as a share of quarterly output to 0.05. We recover the time-invariant income tax rate τ_y to balance the government budget.

  The consumption block of the model is solved in two steps. First, we iterate over the household discount factor β to clear the liquid asset market in steady state. Second, we differentiate the aggregate consumption function around the deterministic steady state, giving the linearized relation

  $$\boldsymbol{c} = \mathcal{C}_y \boldsymbol{y} + \mathcal{C}_i \boldsymbol{i} + \mathcal{C}_\pi \boldsymbol{\pi} + \mathcal{C}_\tau \boldsymbol{\tau}$$

  where c is consumption, y is aggregate income, i is the nominal rate on the liquid asset, π is inflation, and τ denotes the uniform lump-sum transfer. Without uninsurable income risk this demand block would collapse to the familiar Euler equation.

- **Supply block.** The supply relation of our economy is a standard NKPC:[28]

  $$\pi_t = \kappa y_t + \frac{1}{1+\bar{r}}\pi_{t+1} + \varepsilon_t$$

  where ε_t is a contractionary cost-push shock and we set $\kappa = 0.1$. We assume that ε_t follows a standard AR(1) process with persistence 0.8.

[28] See McKay & Wolf (2022) for a discussion of the assumptions on primitives necessary to derive such a supply relation from a combination of nominal rigidities and consumer labor supply in a HANK model.

- **Policy.** The fiscal authority fixes the total amount of outstanding government debt at its steady state level and adjusts lump-sum transfers to balance the budget. The monetary authority follows the policy rules described in Section 2.4.

A.2 Filtering problems

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

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

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

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

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

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

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

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

and so

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

But since

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

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

A.3 More general loss functions

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

$$\mathcal{L} = \frac{1}{2}\boldsymbol{x}'Q\boldsymbol{x} \tag{A.1}$$

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

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

so we can recover the optimal policy rule as

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

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

A.4 Changing policy rules

Our results apply without change to economies in which the policy rule is changing over time; instead, for our purposes, the key requirement is that the non-policy block (6) remains stable. To see why, consider an econometrician that observes data generated from an economy described first by the pair of equations

$$\mathcal{H}_w\boldsymbol{w} + \mathcal{H}_x\boldsymbol{x} + \mathcal{H}_z\boldsymbol{z} + \mathcal{H}_\varepsilon\boldsymbol{\varepsilon} = \mathbf{0} \tag{A.2}$$

$$\mathcal{A}_{x,1}\boldsymbol{x} + \mathcal{A}_{z,1}\boldsymbol{z} + \boldsymbol{\nu}_1 = \mathbf{0} \tag{A.3}$$

before then changing to

$$\mathcal{H}_w\boldsymbol{w} + \mathcal{H}_x\boldsymbol{x} + \mathcal{H}_z\boldsymbol{z} + \mathcal{H}_\varepsilon\boldsymbol{\varepsilon} = \mathbf{0} \tag{A.4}$$

$$\mathcal{A}_{x,2}\boldsymbol{x} + \mathcal{A}_{z,2}\boldsymbol{z} + \boldsymbol{\nu}_2 = \mathbf{0} \tag{A.5}$$

That is, while the private-sector block remains stable throughout, the policy rule has changed over time. In keeping with Assumption 1 we assume that both policy rules induce a unique bounded equilibrium for any bounded sequence of $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}_1\}$ or $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}_2\}$.

Now consider an econometrician that separately studies the propagation of policy shocks on the two subsamples. In particular, suppose she has successfully estimated the causal effects of some bounded policy shock vector $\boldsymbol{\nu}_1$ under the first rule. Then, by the exact same logic as in the proof of Proposition 1, we know that the similarly bounded shock sequence

$$\boldsymbol{\nu}_2 = (\mathcal{A}_{x,1} - \mathcal{A}_{x,2})\boldsymbol{x}_{\mathcal{A}_1}(\boldsymbol{\nu}_1) + (\mathcal{A}_{z,1} - \mathcal{A}_{z,2})\boldsymbol{z}_{\mathcal{A}_1}(\boldsymbol{\nu}_1) + \boldsymbol{\nu}_1$$

under the new policy rule will induce the exact same impulse responses—that is, we have

$$\boldsymbol{x}_{\mathcal{A}_1}(\boldsymbol{\nu}_1) = \boldsymbol{x}_{\mathcal{A}_2}(\boldsymbol{\nu}_2), \quad \boldsymbol{z}_{\mathcal{A}_1}(\boldsymbol{\nu}_1) = \boldsymbol{z}_{\mathcal{A}_2}(\boldsymbol{\nu}_2)$$

The same argument also works in reverse, mapping any bounded $\boldsymbol{\nu}_2$ into a similarly bounded $\boldsymbol{\nu}_1$ with the exact same causal effects. This in particular implies that our econometrician could use her evidence on policy shock propagation from either subsample to implement our identification results in Section 2.3. Intuitively, since the private-sector block is unchanged, policy shocks in both cases identify the exact same space of dynamic causal effects, just with respect to two different reference points—the rule $\{\mathcal{A}_{x,1}, \mathcal{A}_{z,1}\}$ in the first subsample, and the rule $\{\mathcal{A}_{x,2}, \mathcal{A}_{z,2}\}$ in the second subsample.

As the final step in the argument suppose now that the econometrician would estimate policy shock causal effects across the two (large) subsamples; in particular, suppose that the relative sizes of her two subsamples are $(\omega, 1-\omega)$. In that case, direct projection on an instrumental variable that correlates with a given shock sequence $\boldsymbol{\nu}$ would asymptotically recover a weighted average of sub-sample causal effects, e.g., $\omega\Theta_{x,\nu,\mathcal{A}_1} \times \boldsymbol{\nu} + (1-\omega)\Theta_{x,\nu,\mathcal{A}_2} \times \boldsymbol{\nu}$. The overall estimated dynamic causal effect matrices would thus just be

$$\begin{aligned} \Theta_{x,\nu,\mathcal{A}} &\equiv \omega\Theta_{x,\nu,\mathcal{A}_1} + (1-\omega)\Theta_{x,\nu,\mathcal{A}_2} \\ \Theta_{z,\nu,\mathcal{A}} &\equiv \omega\Theta_{z,\nu,\mathcal{A}_1} + (1-\omega)\Theta_{z,\nu,\mathcal{A}_2}. \end{aligned}$$

Importantly, post-multiplying those dynamic causal effects matrices by *any* policy shock vector $\boldsymbol{\nu}$ yields sequences of private-sector outcomes x and policy instruments z that are consistent with the common (sample-invariant) private-sector block (A.2) = (A.4).

Now consider using those causal effect matrices to implement Proposition 1; that is, given baseline non-policy shock impulse responses $\{\boldsymbol{x}(\boldsymbol{\varepsilon}), \boldsymbol{z}(\boldsymbol{\varepsilon})\}$ (which could come from either subsample, or also from a weighted average), find $\tilde{\boldsymbol{\nu}}$ such that

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

and then compute the impulse responses

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

Since the counterfactual policy rule holds by construction of $\tilde{\boldsymbol{\nu}}$, and since the averaged policy shock causal effects $\{\Theta_{x,\nu,\mathcal{A}}, \Theta_{z,\nu,\mathcal{A}}\}$ embed the common (sample-invariant) private-sector block, it follows by the same arguments as in the proof of Proposition 1 that those impulse responses equal $\boldsymbol{x}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$ and $\boldsymbol{z}_{\tilde{\mathcal{A}}}(\boldsymbol{\varepsilon})$, respectively.

A.5 Counterfactual second-moment properties

Our analysis is largely focussed on constructing counterfactuals conditional on particular non-policy shock paths $\boldsymbol{\varepsilon}$. This is in keeping with much of the empirical policy counterfactual literature that followed the lead of Sims & Zha (1995) (e.g., Bernanke et al., 1997; Eberly et al., 2020; Antolin-Diaz et al., 2021). However, under some additional assumptions, our results can also be used to construct *unconditional* counterfactual second-moment properties—that is, predict how variances and covariances of macroeconomic aggregates would change under a counterfactual rule. This section provides the detailed argument.

Setting. We consider a researcher that observes and is interested in the counterfactual properties of some vector of aggregates $y = (x, z)$—the endogenous outcomes and policy instruments of our main analysis. We assume that, under the prevailing policy rule, this vector of macro aggregates follows a standard structural vector moving average representation:[29]

$$y_t = \sum_{\ell=0}^{\infty} \Theta_\ell \varepsilon_{t-\ell} = \Theta(L)\varepsilon_t, \quad \varepsilon_t \sim N(0, I) \tag{A.6}$$

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

We would like to predict the second-moment properties of the macroeconomic aggregates y_t under some counterfactual policy rule (8).

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

PROCEDURE. Our proposed procedure has three steps. First, the researcher estimates the Wold representation of the observables y_t. Second, using Proposition 1, she maps the impulse responses to the Wold errors into new impulse responses corresponding to the counterfactual policy rule. Third, she stacks those new impulse responses to arrive at a new vector moving average representation, and from this representation constructs a new set of second-moment properties. Our identification result states that, if the vector moving representation (A.6) under the baseline rule is invertible, then this procedure correctly recovers the desired counterfactual second moments.

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

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

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

$$\Gamma_y(\ell) = \sum_{m=0}^{\infty} \tilde{\Theta}_m \tilde{\Theta}'_{m+\ell}. \tag{A.7}$$

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

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

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

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

Now consider the two lag polynomials $\tilde{\Theta}(L)$ and $\tilde{\Psi}(L)$. Since $\varepsilon_t^* = P\varepsilon_t$, applying our counterfactual mapping to the jth Wold innovation $\varepsilon_{j,t}^*$ gives causal effects $\tilde{\Psi}_j(L) = \tilde{\Theta}(L) \times p_j'$, where p_j is the jth row of P.[30] Thus overall we have

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

But then the second-moment properties of y_t implied by our proposed procedure are given as

$$\sum_{m=0}^{\infty} \tilde{\Psi}_m \tilde{\Psi}_{m+\ell}' = \sum_{m=0}^{\infty} \tilde{\Theta}_m P'P\tilde{\Theta}_{m+\ell}' = \sum_{m=0}^{\infty} \tilde{\Theta}_m \tilde{\Theta}_{m+\ell}' = \Gamma_y(\ell) \tag{A.8}$$

which is exactly equal to (A.7), completing the argument.

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

A.6 Optimal policy counterfactual in HANK

Section 2.4 used a HANK model to illustrate the logic of Proposition 1—the general counterfactual rule identification result. We here do the same for the analogous optimal policy identification result in Proposition 2.

We consider a policymaker with a standard dual mandate loss function

$$\mathcal{L} = \lambda_\pi \boldsymbol{\pi}' W \boldsymbol{\pi} + \lambda_y \boldsymbol{y}' W \boldsymbol{y} \tag{A.9}$$

[30] To see this, let $\tilde{\boldsymbol{\nu}}_i$ denote the sequence of policy shocks that maps the true shock ε_i into its counterfactual causal effects. Then the sequence of policy shocks $\tilde{\boldsymbol{\nu}}_j^*$ that implements our counterfactual mapping for the reduced-form shock ε_j^* is given as $\tilde{\boldsymbol{\nu}}_j^* = \sum_i \tilde{\boldsymbol{\nu}}_i p_{ij}$, and thus we have $\tilde{\Psi}_j(L) = \sum_i \tilde{\Theta}_i(L) p_{ij}$, as claimed.

Optimal Policy, HANK Model

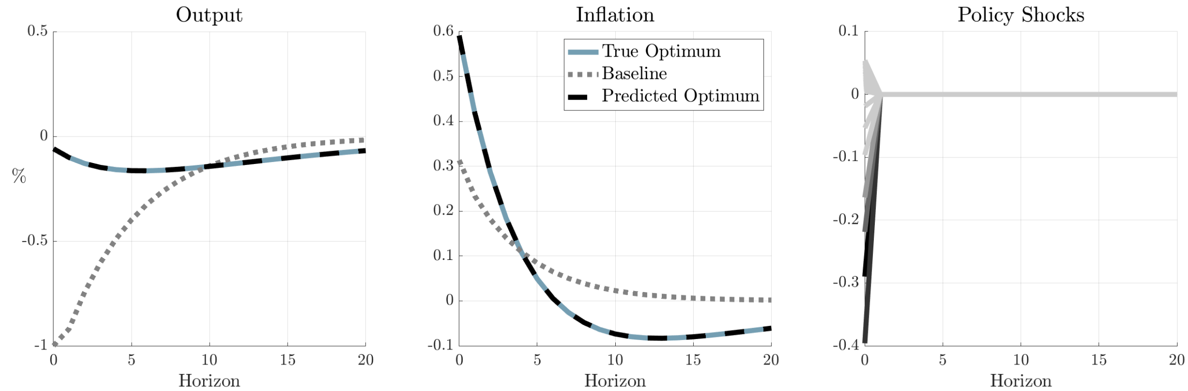


Figure A.1: The dotted and solid lines in the left and middle panels show output and inflation responses to the cost-push shock ε_t for the HANK model with policy rule (28) and the optimal rule for the loss function (A.9). The dashed lines give output and inflation counterfactuals constructed through the policy shocks on the right, set in line with Proposition 2. Lighter shades correspond to news about policy at longer horizons.

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

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

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

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

The optimal rule accommodates the inflationary cost-push shock more than the baseline rule (28), so the required policy "shock" is persistently negative (i.e., expansionary). Consistent with our discussion in Figure 1, we choose to display those shocks $\tilde{\boldsymbol{\nu}}$ in a way that emphasizes that the optimum is achieved through a sequence of *date-0* policy shocks.

A.7 Counterfactuals with a limited number of shocks

In Figure 1 we constructed counterfactuals using a limited number n_s of policy shocks. We here provide the computational details for this construction. We discuss the general case of a researcher with access to n_s shocks (which converges to our identification result for $n_s \to \infty$), with the original proposal of Sims & Zha (1995) nested as the $n_s = 1$ special case.

The approach of Sims & Zha leverages the idea that evidence on one policy shock—i.e., any single fixed path $\boldsymbol{\nu}$—is sufficient to enforce any given counterfactual *ex post*. With n_s distinct shocks, the counterfactual rule can be implemented *ex post* as well as in *ex ante* expectation for the next $n_s - 1$ time periods. To compute the counterfactuals corresponding to this multi-shock case we proceed as follows. First, at $t = 0$, we solve for the n_s-dimensional vector of policy shocks $\boldsymbol{\nu}^0_{1:n_s} \equiv (\nu^0_0, \dots, \nu^0_{n_s-1})'$ such that, in response to $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}^0_{1:n_s}$ the counterfactual rule holds at $t = 0$ and is expected to hold for $t = 1, \dots, n_s - 1$. Output and inflation at $t = 0$ are simply given as the thus-derived impulse responses to $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}^0_{1:n_s}$. Second, at $t = 1$, we solve for the n_s-dimensional vector of shocks $\boldsymbol{\nu}^1_{1:n_s} \equiv (\nu^1_0, \dots, \nu^1_{n_s-1})'$ such that, in response to the time-0 shocks $\{\boldsymbol{\varepsilon}, \boldsymbol{\nu}^0_{1:n_s}\}$ and the time-1 shocks $\boldsymbol{\nu}^1_{1:n_s}$, the counterfactual policy rule holds at $t = 1$ and in expectation for $t = 2, \dots, n_s$. These impulse responses then give us output and inflation at $t = 1$. Continuing iteratively, we then obtain the entire output and inflation impulse responses, as plotted in the left and middle panels of Figure 1. The corresponding policy shock paths are shown in the right panel.

A.8 Global identification argument

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

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

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

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{z}) = \mathbf{0}. \tag{A.10}$$

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

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

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.10) - (A.11) has a unique solution. We write this solution as

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

We want to construct counterfactuals under the alternative policy rule

$$\tilde{\mathcal{A}}(\boldsymbol{x}, \boldsymbol{z}) = \mathbf{0} \tag{A.12}$$

replacing (A.11). We again assume that the system (A.10) and (A.12) has a unique solution, now written as $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}})$. If we are interested in the counterfactual following a particular path of exogenous events, then we are interested in selections from these vectors.

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

$$x(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{x}}, \quad z(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{z}} \tag{A.13}$$

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

$$\tilde{\mathcal{A}}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) = \mathbf{0}. \tag{A.14}$$

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

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

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

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

$$\begin{aligned} \mathcal{H}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) &= \mathbf{0} && \text{(A.15)} \\ \mathcal{A}(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}})) + \tilde{\mathcal{A}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}) - \mathcal{A}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}) &= \mathbf{0} && \text{(A.16)} \end{aligned}$$

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

$$\begin{aligned} \mathcal{H}(\tilde{x}(\mathbf{0}), \tilde{z}(\mathbf{0})) &= \mathcal{H}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}) = \mathbf{0} && \text{(A.17)} \\ \tilde{\mathcal{A}}(\tilde{x}(\mathbf{0}), \tilde{z}(\mathbf{0})) &= \tilde{\mathcal{A}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}) = \mathbf{0} && \text{(A.18)} \end{aligned}$$

Since the system (A.15) - (A.16) by assumption has a unique solution for any $\tilde{\boldsymbol{\nu}}$, it thus follows that we must have $\{x(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{x}}, z(\tilde{\boldsymbol{\nu}}) = \tilde{\boldsymbol{z}}\}$.

We now show that any solution to (A.14) must generate $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}})$. Proceeding by contradiction, consider any other $\tilde{\boldsymbol{\nu}}$ that solves (A.14) and suppose that either $x(\tilde{\boldsymbol{\nu}}) \neq \tilde{\boldsymbol{x}}$ and/or $z(\tilde{\boldsymbol{\nu}}) \neq \tilde{\boldsymbol{z}}$. By definition of the functions $x(\bullet)$ and $z(\bullet)$ together with the property (A.14) we know that

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

and so $(x(\tilde{\boldsymbol{\nu}}), z(\tilde{\boldsymbol{\nu}}))$ is a solution of (A.10) and (A.12) that is distinct from $(\tilde{x}, \tilde{z})$. But by assumption only one such solution exists, so we have a contradiction.

□

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

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

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

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

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

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

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

A.9 Non-linear policy rules

We emphasize that the simplicity of our baseline identification results in Section 2.3 relative to the much more involved discussion in Appendix A.8 hinges on linearity of the non-policy block (6). Non-linear policy blocks, on the other hand, are straightforward to handle with the same informational requirements as in our baseline analysis. Specifically, suppose the object of interest is the counterfactual propagation of a shock sequence $\boldsymbol{\varepsilon}$ under the following example of a non-linear policy rule:

$$\boldsymbol{z} = \max\left\{\tilde{\mathcal{A}}_x \boldsymbol{x}, \underline{z}\right\} \tag{A.19}$$

Compared to (8), this policy rule is simplified to set $\tilde{\mathcal{A}}_z = -I$, but then enriched to allow for a canonical form of non-linearity—a kink in the policy rule, e.g., a zero lower bound constraint on nominal interest rates. Proceeding perfectly analogously to the proof of Proposition 1, we see that we can construct counterfactuals corresponding to (A.19) by solving for a sequence of policy shocks $\tilde{\boldsymbol{\nu}}$ such that (A.19) holds for all $t = 0, 1, \ldots$. Intuitively, by our unchanged assumptions on the non-policy block (6), it is still just the time path of the policy instrument that matters for private-sector behavior, irrespective of whether this time path is generated by a linear rule like (7) or a non-linear rule like (A.19).

B Details for empirical method

This appendix provides econometric implementation details for our empirical method. Appendix B.1 begins with counterfactuals for a given alternative policy rule, while Appendix B.2 discusses optimal policy counterfactuals.

B.1 Policy rule counterfactuals

The solution to the problem (31) is given as

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

The final step is simply to compute impulse responses to the combination of (i) the original non-policy shock $\boldsymbol{\varepsilon}$ and (ii) the derived policy shocks $\boldsymbol{s}$. For counterfactual second-moment properties (as discussed in Appendix A.5) the only change is that these steps are applied separately for each innovation in the Wold representation of observed macro aggregates.

B.2 Optimal policy counterfactuals

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

$$\min_{\boldsymbol{s}} \frac{1}{2}\sum_{i=1}^{n_x} \lambda_i \boldsymbol{x}_i' W \boldsymbol{x}_i \tag{B.1}$$

such that

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

This gives the optimality conditions:

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

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots)$.

Solving this system gives our optimal policy counterfactual. The solution is given as

$$\boldsymbol{s} = -\left[\Omega'_{x,\mathcal{A}} \left(\Lambda \otimes W\right) \Omega_{x,\mathcal{A}}\right]^{-1} \times \left[\Omega'_{x,\mathcal{A}} \left(\Lambda \otimes W\right) \boldsymbol{x}_{\mathcal{A}}(\boldsymbol{\varepsilon})\right].$$

As before, for counterfactual second-moment properties, the analysis is repeated for impulse responses to all Wold innovations separately.

C Supplementary details for monetary applications

This appendix provides further results supplementing the discussion in Section 4 on our systematic monetary policy rule counterfactual applications. Appendices C.1 and C.2 begin by describing the data and our baseline monetary policy shock causal effect estimates. Results for the policy counterfactuals omitted in the main text are presented in Appendix C.3, and we investigate the robustness of our results to the use of other monetary polishock measures in Appendix C.4. Finally, we in Appendix C.5 illustrate how to use our monetary shock estimates to construct counterfactual second-moment properties.

C.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, we use the series `ygap_hp` of Barnichon & Mesters (2020).[31] For inflation, we compute annual changes in the GDP deflator (using the series `pgdp` from the replication files of Ramey (2016)). Finally, we consider the federal funds rate as our measure of the policy rate, obtained from the St. Louis Federal Reserve FRED database. In keeping with much prior work, we also additionally control for commodity prices, with our measure obtained from the replication files of Ramey (2016) (`lpcom`). All series are quarterly.

Shocks & identification. We take the investment-specific technology shock series from Ben Zeev & Khan (2015) (`bzk_ist_news` in the replication files of Ramey (2016)), the Romer & Romer (2004) shock series from the replication and extension of Wieland & Yang (2020) (`rr_3`), and the high-frequency monetary policy surprise series from Gertler & Karadi (2015) (`mp1_tc` in the replication files of Ramey (2016)).[32] When applicable, the shock series are aggregated to quarterly frequency through simple averaging.

[31] All results are essentially unchanged if we use a measure of log real GDP instead (`rgdp` scaled by `pop`, taken from the replication files of Ramey (2016)).

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

In Appendix C.4 we examine the robustness of our conclusions to other policy shock series—those of Aruoba & Drechsel (2022) and Miranda-Agrippino & Ricco (2021). For the former, we obtain the shock series directly from their replication files (`shock`). For the latter, we use the publicly available replication files to construct the SVAR-IV shock series for the full sample (from 1979:M1 onwards), with the shocks constructed at the posterior mode of the estimated reduced-form VAR (the specification for their Figure 3).

C.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, 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.[33] For identification, we order the Gertler & Karadi shock first (again consistent with the results in Plagborg-Møller & Wolf (2021)) and the Romer & Romer shock second-to-last, before the federal funds rate (the additional "exogeneity insurance" as in Romer & Romer, 2004).

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

Results. The OLS point estimates for the technology shocks of Ben Zeev & Khan (2015) are reported as the dotted lines in Figure 3. For monetary policy, the estimated causal effects for our two outcomes of interest as well as the policy instrument are displayed in Figure C.1. The results are in line with prior work: both policy shocks induce the expected signs of the output gap and inflation responses, though the response shapes are quite distinct, consistent with the differences in the induced interest rate paths. We also note that the magnitudes of

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

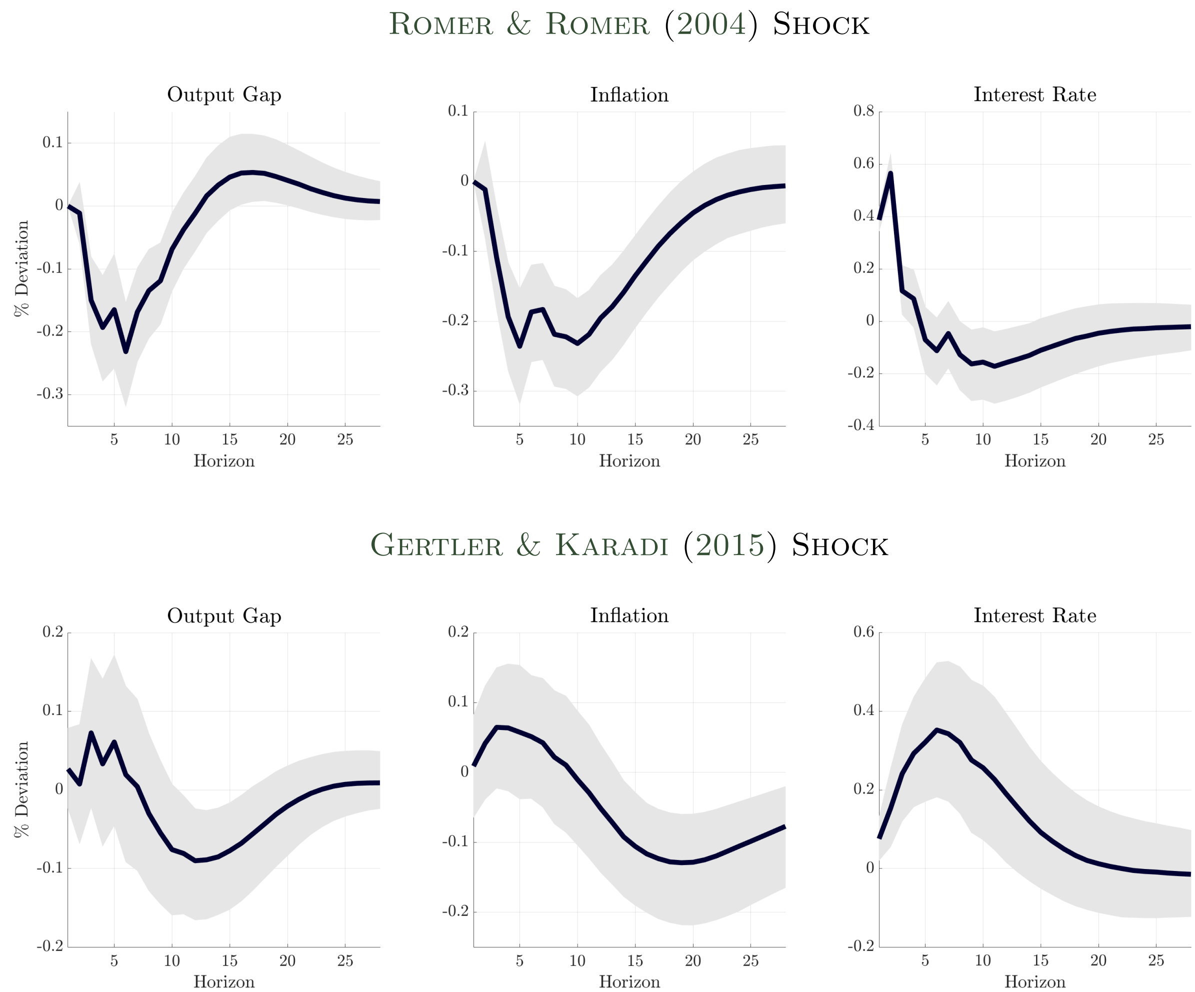


Figure C.1: Impulse responses after the Romer & Romer monetary policy shock (top panel) and the Gertler & Karadi monetary policy shock (bottom panel). The grey areas correspond to 16th and 84th percentile confidence bands, constructed using 10,000 draws from the posterior distribution of the reduced-form VAR parameters.

the estimated responses are at the lower end of empirical estimates (c.f. Table 2 and Figures 1-2 in Ramey, 2016).

C.3 Results for omitted monetary policy counterfactuals

In Section 4 we presented detailed results for only three of our policy rule counterfactuals—strict output gap targeting, the Taylor rule, and optimal average inflation targeting policy. We here provide the remaining results.

Policy Counterfactual, Interest Rate Peg

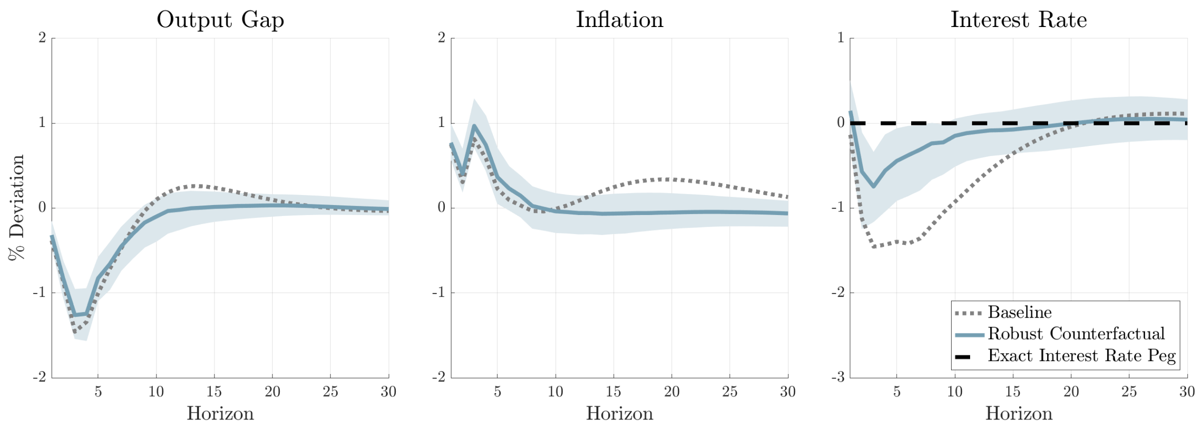


Figure C.2: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted) and the best feasible approximation to a nominal rate peg (solid), computed following (31). The shaded areas correspond to 16th and 84th percentile confidence bands. Perfect nominal rate peg (i.e., $\widehat{i}_t = 0$ for all t) is displayed as the black dashed line.

Nominal interest rate peg. Results for the nominal interest rate peg are presented in Figure C.2. We see that the desired counterfactual policy is implemented well from a couple of quarters out onwards, but that nominal rates are still cut by quite a bit too much immediately after the shock. Since rates are cut by less than in the baseline, the output gap and inflation remain marginally lower for a longer period of time. Compared with the policy counterfactuals discussed in Section 4, we see that a nominal rate peg is a counterfactual policy that is not spanned particularly well by our available monetary shock evidence.

Nominal GDP targeting. Results for nominal GDP targeting are presented in Figure C.3. The counterfactual policy is implicitly defined by the targeting rule

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

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

Policy Counterfactual, Nominal GDP Targeting

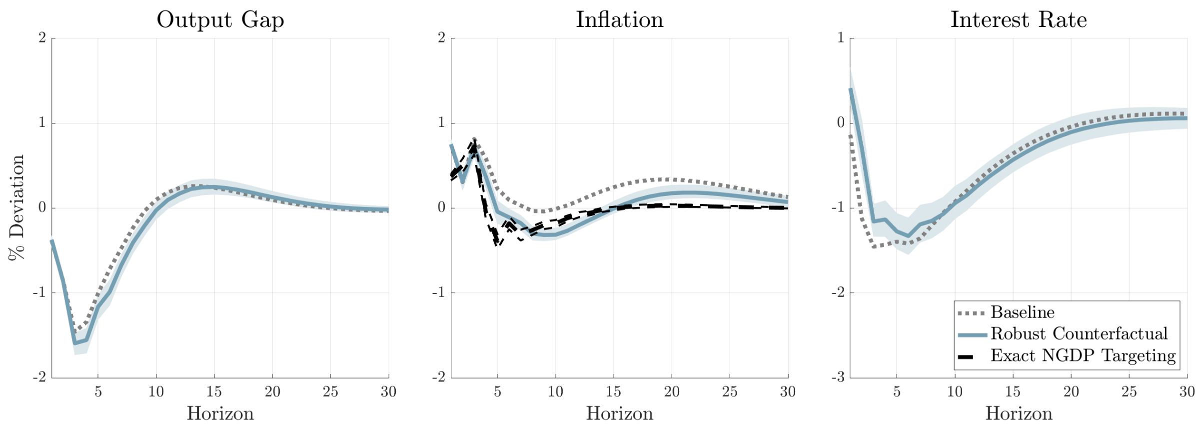


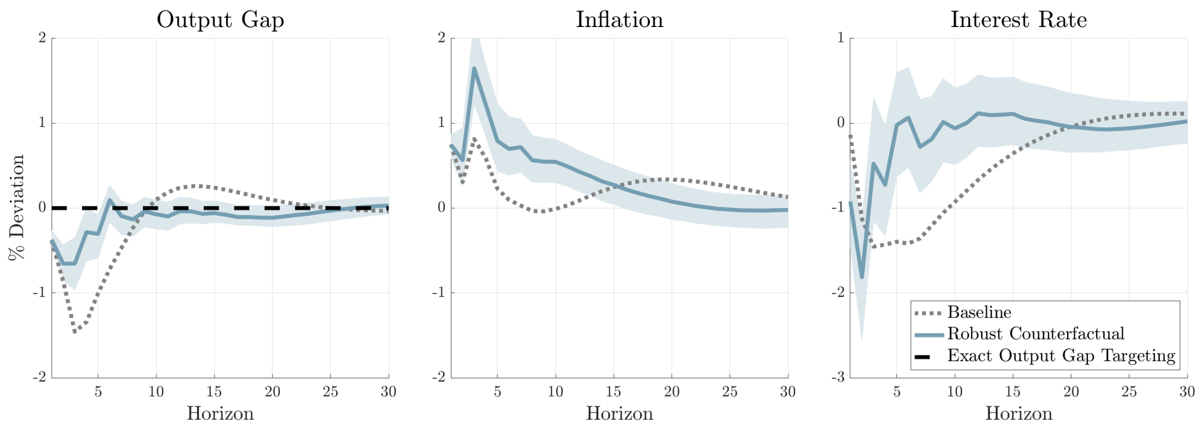
Figure C.3: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted) and the best feasible approximation to nominal GDP targeting (solid), computed following (31). The shaded areas correspond to 16th and 84th percentile confidence bands. Given each counterfactual draw for the output gap, the sequence of inflation corresponding to perfect nominal GDP targeting (i.e., so that $\widehat{\pi}_t + (\widehat{y}_t - \widehat{y}_{t-1}) = 0$ for all t) is displayed as the black dashed line.

C.4 Counterfactuals with alternative shock measures

Some recent work has questioned the validity of the canonical monetary policy shocks of Romer & Romer and Gertler & Karadi (see Ramey, 2016; Nakamura & Steinsson, 2018, and the references therein). To examine the robustness of our conclusions to the use of alternative measures of monetary policy shocks, we now use the policy shock series of Miranda-Agrippino & Ricco (2021) and Aruoba & Drechsel (2022). These shock series are constructed using methods similar to those of Gertler & Karadi and Romer & Romer, but use a richer set of controls for the state of the economy as perceived by the Federal Reserve.

We study the propagation of these shocks in a single integrated VAR, exactly as in our baseline analysis. We find that the two shocks differ in the implied interest rate movements, with the shock of Miranda-Agrippino & Ricco (2021) mirroring the transitory rate movement of Romer & Romer (2004), and the shock of Aruoba & Drechsel (2022) similar to the gradual interest rate movement of Gertler & Karadi (2015). We then leverage these shock estimates to construct monetary policy rule counterfactuals, proceeding exactly as in Section 4. Results for our two main systematic policy rule counterfactuals—output gap targeting and the Taylor rule—are displayed in Figure C.4. The main takeaway is that the systematic monetary policy rule counterfactuals are very similar to our headline results. The underlying reason is simply

Policy Counterfactual, Output Gap Targeting, Alternative Shocks



Policy Counterfactual, Taylor Rule, Alternative Shocks

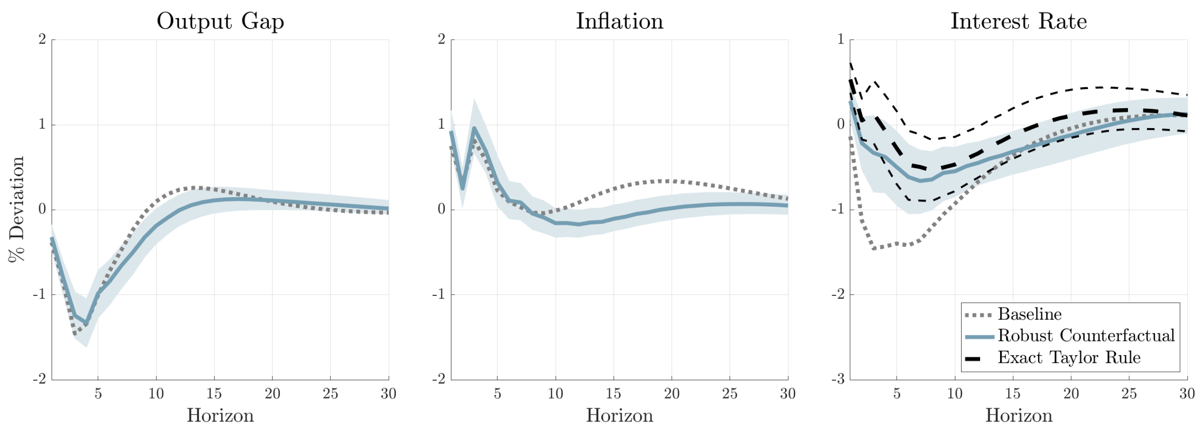


Figure C.4: Output gap, inflation and interest rate impulse responses to a contractionary investment-specific technology shock under the prevailing baseline rule (dotted grey) and the best feasible approximation to output gap targeting (solid, top panel) and a simple Taylor-type rule $\widehat{i}_t = 0.5\widehat{i}_{t-1} + 0.5 \times (1.5\widehat{\pi}_t + \widehat{y}_t)$ (solid, bottom panel) computed following (31) and using the monetary shocks of Miranda-Agrippino & Ricco (2021) and Aruoba & Drechsel (2022). The shaded areas correspond to 16th and 84th percentile confidence bands.

that the impulse responses to the Miranda-Agrippino & Ricco and Aruoba & Drechsel shocks are quite similar to those displayed in Figure C.1 for Romer & Romer and Gertler & Karadi. The perhaps most notable difference is that the shocks of Miranda-Agrippino & Ricco and Aruoba & Drechsel have somewhat *larger* effects on output and inflation (for a given peak interest rate response), so the interest rate cut for the output gap targeting counterfactual is somewhat less steep, and the inflation spike is somewhat more pronounced.

C.5 Counterfactual second-moment properties

In this section we illustrate how estimates of monetary policy shock causal effects can also be used to construct counterfactual *average* business-cycle statistics. Specifically, we construct optimal policy counterfactuals for the the average inflation targeting loss function (32).

Our procedure follows the steps outlined in Appendices A.5 and B.2. First, we estimate the Wold representation for our three macroeconomic observables (output gap, inflation, policy rate), giving us impulse responses to the three reduced-form Wold innovations. Then, for each of these three reduced-form shocks, we find the linear combination of date-0 monetary policy shocks that minimizes the policymaker loss function. We then stack these three sets of impulse responses in a new, counterfactual Wold representation, and finally use it to construct counterfactual second-moment properties. We do so for 10,000 draws of monetary policy shock causal effects from our reduced-form VAR.

	Standard Deviation (per cent)	**Correlation with Output Gap**
Baseline		
Output Gap	1.52	—
Inflation	1.69	0.14
Nominal Rate	2.68	0.27
Counterfactual		
Output Gap	1.27	—
	(1.06, 1.44)	
Inflation	1.38	-0.02
	(1.27, 1.53)	(-0.18, 0.11)
Nominal Rate	2.12	0.39
	(1.73, 2.62)	(0.24, 0.53)

Table C.1: Baseline and counterfactual business-cycle statistics for the best Lucas critique-robust approximation to an optimal average inflation targeting monetary policy rule, computed as discussed in Appendix B.2 applied to each of the three reduced-form innovations in the Wold representation of output gap, inflation, and policy rate. The values in brackets correspond to 16th and 84th percentile confidence bands.

Results are reported in Table C.1, where the top panel shows business-cycle statistics under observed policy conduct while the bottom panel presents our optimal policy counterfactual. We see that the empirically available subspace of two identified monetary shock paths suffices to somewhat lower the standard deviation of the output gap and inflation. However, the reported gains in aggregate volatility are not substantial, suggesting that rather little policy improvement was feasible within our identified space of policy shock causal effects. These conclusions echo our *conditional* shock conclusions in Section 4.2.