Evaluating Policy Counterfactuals: A "VAR-Plus" Approach†

Tomás E. Caravello — MIT
Alisdair McKay — FRB Minneapolis
Christian K. Wolf — MIT & NBER

February 7, 2024

Abstract: In a general family of linearized structural macroeconomic models, the counterfactual evolution of the economy under alternative policy rules is fully pinned down by: (i) reduced-form projections with respect to a large information set; and (ii) the causal effects of policy on macroeconomic aggregates. This identification result motivates a three-step approach. First, the researcher constructs (i) directly from time-series data. Second, she partially estimates (ii) using semi-structural methods for policy shock analysis. Third, if needed, she uses structural models of policy transmission to extrapolate beyond those observed policy experiments, giving the missing parts of (ii). We document attractive robustness properties of this procedure, and use it to study U.S. business-cycle fluctuations under alternative assumptions on the conduct of monetary policy.

Keywords: Lucas critique, policy counterfactuals, monetary policy, policy shocks, impulse response matching, invertibility. *JEL codes:* E32, E61.

---

†Email: tomasec@mit.edu, alisdair.mckay@mpls.frb.org and ckwolf@mit.edu. We received helpful comments from Marco Bassetto, Anmol Bhandari, Martin Beraja, Ricardo Caballero, Gabriel Chodorow-Reich, Martin Eichenbaum, Ellen McGrattan, Ben Schumann, Iván Werning, and seminar participants at various venues. Wolf acknowledges that this material is based upon work supported by the NSF under Grant #2314736. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

1 Introduction

How would the economy have evolved if policy had been set differently? For example, how would a different systematic monetary policy reaction function have shaped the average business-cycle? And how would it have changed particular historical episodes?

The dominant methodological approach to answering such policy counterfactual questions is the "quantitative DSGE" paradigm (e.g., Smets and Wouters, 2007). Here, the researcher first builds a model that can account for the entire history of macroeconomic fluctuations, then changes policy, and finally re-solves the model.[1] An obvious concern with this approach is model mis-specification; in particular, it is widely argued that the primitive shocks added to the model to generate cyclical fluctuations are "dubiously structural" and thus not plausibly policy-invariant (e.g., see Chari et al., 2009).[2]

In this paper we instead contribute to a recent literature that tries to evaluate policy counterfactuals with less reliance on explicit model structure (e.g., Barnichon and Mesters, 2023; McKay and Wolf, 2023). We propose a new methodological approach, formally justify it with an identification result, and then apply it to re-evaluate U.S. business-cycle fluctuations under alternative assumptions on the systematic conduct of monetary policy. Relative to the DSGE paradigm, our approach relies on strictly weaker structural assumptions; in particular, it does not require the researcher to provide a structural account of the primitive shocks that drive the business cycle. Relative to McKay and Wolf (2023), the proposed approach can be applied to a larger set of policy counterfactual questions.

Identification result. The first part of the paper establishes the identification result that underlies our method. We are interested in the counterfactual evolution of the macroeconomy under alternative policy rules, both unconditional—i.e., how the "average" business cycle would unfold—and conditional on particular historical episodes. We show that, across a large family of linearized structural macroeconomic models, these counterfactuals are pinned down by two "sufficient statistics."

(i) *Reduced-form projections.* The first statistic is a set of reduced-form projections. For unconditional average business-cycle counterfactuals, those projections are impulse re-

[1] This approach is particularly prominent in central banks. A notable recent example is Crump et al. (2023), who re-evaluate U.S. monetary policy during the post-covid inflationary episode.

[2] Commonly discussed examples of such "dubiously structural" shocks are price and wage mark-up shocks as well as innovations to household discount factors as consumer demand shocks.

sponses of macro aggregates to reduced-form ("Wold") innovations. For counterfactuals conditional on particular episodes, the projections are forecasts, from each date in the episode of interest. Importantly, we require that these projections are relative to an information set that spans the (unknown) shocks buffeting the economy—i.e., we are maintaining the assumption of "invertibility" (Fernández-Villaverde et al., 2007).

(ii) *Policy causal effects.* The second statistic is the set of dynamic causal effects of changes in policy on current and future macroeconomic aggregates—i.e., the space of macroeconomic outcomes that is achievable through manipulation of the policy instrument(s) now and in the future. For example, for monetary policy, the researcher needs to know the causal effects of changes in the expected policy rate at all future horizons.

The identification result reveals that any two structural models in the class we consider—no matter their detailed parametric structure, and in particular completely independently of the structural shocks they feature—that agree on these two sufficient statistics will also agree on the policy counterfactuals that they imply. The intuition has two parts. First, knowledge of policy causal effects ensures that we can correctly predict how any given reduced-form projection would be altered by a hypothetical change in policy (by McKay and Wolf, 2023). Second, given the assumption of invertibility, correctly predicting how reduced-form projections change is actually equivalent to correctly predicting the counterfactual propagation of the economy's true (though unknown) structural shocks. The key step in the argument is that, by invertibility, the unknown true structural shocks are in fact a one-to-one function of reduced-form forecast innovations. But since predictions based on a one-to-one function of the true shocks are equal to predictions based on those shocks themselves, there is actually no need to separately identify the true primitive shocks.

Methodology. We propose to operationalize this identification result using a three-step "VAR-Plus" methodology. The approach first constructs the required sufficient statistics, and then translates them into the counterfactuals using the identification result.

(i) A convenient way of estimating the autocovariance function of the data—and from here constructing the Wold representation and thus the required reduced-form projections—is through a reduced-form Vector Autoregression (VAR). This VAR should be specified with the invertibility requirement in mind; in practice, this suggests including a large set of macroeconomic observables.

(ii) The researcher uses semi-structural empirical methods to learn about the causal effects of policy changes on macroeconomic outcomes where possible, and leverages additional economic structure to extrapolate beyond that evidence where necessary.

a) Using the standard semi-structural time series toolkit, the researcher estimates the causal effects of identified shocks to the policy instrument under consideration. In practice, since empirical evidence on policy shocks is limited, this step will only *partially* pin down the required full space of policy causal effects. VARs are again a convenient technique for estimating such causal effects in practice.

To answer a given policy counterfactual question, these "VAR" steps may or may not suffice; intuitively, the challenge is that the empirical policy shock evidence only pins down the effects of *some*—but not *all possible*—changes in policy. If the evidence does not suffice, then our methodology relies on additional structural assumptions—the "Plus" step.

b) The researcher uses a model of policy transmission—or multiple such models—to extrapolate beyond the empirical evidence on policy causal effects. Specifically, she considers a list of candidate models, and then jointly estimates them by requiring consistency with the available policy shock evidence—i.e., model estimation via impulse-response matching (Christiano et al., 2005, 2010). This step yields a distribution over models of policy transmission and thus over the causal effects of *any possible* change in policy; importantly, that distribution is by design consistent with the available policy shock evidence, and then extrapolates beyond it using the structure embedded in the contemplated list of models.[3]

With those sufficient statistics in hand, the researcher can evaluate the policy counterfactual.

Discussion. The principal appeal of our method relative to the standard DSGE approach is its reliance on weaker structural assumptions, affording it additional robustness. First, for some counterfactuals, the VAR steps already suffice, obviating the need to rely on model structure for further policy causal effect extrapolation. In such cases, our method is semi-structural, and thus robust to *arbitrary* forms of model mis-specification, at least within the class of models covered by the identification result. Second, to the extent that structure is

[3]Implicitly, the discussion presupposes that at least one of the contemplated models is sufficiently rich to match the empirical policy causal effect evidence.

needed, it is "only" to model the effects of changes in policy on macro aggregates. A model of policy transmission is a *partial* model in the precise sense that it requires no statements on either the shock processes driving the economy or on the policy rule. Our approach is thus robust to mis-specification in these model blocks; in particular, it sidesteps all concerns about mis-specification in "dubiously structural" shocks (Chari et al., 2009).

Applications. We showcase our method with applications to monetary policy counterfactuals. We begin by constructing the required sufficient statistics.

(i) We estimate a large-dimensional reduced-form VAR, with the specification closely following that of Angeletos et al. (2020). We collect the VAR-implied Wold innovation impulse responses as well as forecasts at each in-sample date t.

(ii) a) We take the monetary shock series of Aruoba and Drechsel (2022), and use standard VAR methods to estimate its dynamic causal effects on macroeconomic aggregates. This evidence pins down the propagation of a *transitory* change in interest rates.

b) To extrapolate from transitory to more *persistent* rate changes, we consider a variety of quantitative business-cycle models, notably canonical RANK and HANK models (Christiano et al., 2005; Kaplan et al., 2018); we then also study extended behavioral versions of these models with cognitive discounting (Gabaix, 2020). We estimate this suite of models through a Bayesian impulse response matching procedure, targeting the transitory interest rate change evidence from a). Our first result is that all of those models can match the targeted monetary shock evidence quite closely. The baseline RANK and HANK models furthermore largely agree on the extrapolation beyond the targeted policy experiment; thus, by our identification results, they will agree on *any* possible monetary policy counterfactual. The "behavioral" variants of those models, on the other hand, extrapolate quite differently; in particular, they imply much weaker effects of future policy on current outcomes.

With the inputs in hand, we then study three monetary policy counterfactuals.

1. We ask whether U.S. monetary policy could have reduced the volatility of the aggregate output gap as well as inflation over a post-war sample period. Our analysis suggests that substantial volatility reductions would have been feasible, in particular for output.

2. We study how the Great Recession would have evolved in the absence of a binding lower bound on nominal rates. We find that a standard "dual mandate" monetary authority

would have reduced nominal interest rates substantially into negative territory (around -4 per cent), delivering output stabilization at the cost of somewhat higher inflation.

In these first two applications, the policy counterfactuals are already largely pinned down by the causal effects of transitory changes in nominal rates—a policy experiment on which we have good empirical evidence. The "VAR" steps thus essentially suffice, with no further need for the model extrapolation ("Plus") step.

3. We evaluate monetary policy options after the summer of 2021, when inflation had started to accelerate. Here we find that our documented differences in how behavioral and non-behavioral models extrapolate policy transmission translate into large differences in counterfactuals. In the baseline HANK and RANK models, the policymaker can use forward guidance to steer inflation expectations, reducing current inflation at no cost to output in the short run. In our behavioral models, this strategy instead is much less effective. Given this disagreement across models and thus the (relevant) policy causal effects, our method indicates large uncertainty on the counterfactual path of interest rates.

The large posterior uncertainty in this third counterfactual reflects an important gap in our understanding of monetary policy transmission. The available evidence only pins down the causal effects of transitory rate changes; whenever monetary policy needs to respond to any kind of (highly) persistent disturbance, however, the effects of persistent policy changes also matter. Our results here reveal that HANK and RANK models extrapolate to such persistent policy changes in similar ways, while less forward-looking behavioral models behave very differently. Thus, for the purpose of policy evaluation, learning about the strength of persistent policy changes appears to be of chief importance.

Further literature. We contribute to a recent literature on policy shock impulse responses as "sufficient statistics" for policy counterfactuals (see McKay and Wolf, 2023; Barnichon and Mesters, 2023; Hebden and Winkler, 2021). Relative to McKay and Wolf (2023) and Barnichon and Mesters (2023), the present paper differs in two key ways. First, we focus on a different set of policy counterfactual questions—we seek to evaluate conditional and unconditional counterfactual macroeconomic dynamics without requiring any knowledge of the primitive underlying shocks driving the cycle, thereby addressing one the literature's main concerns with the standard DSGE paradigm (Chari et al., 2009). Second, if needed, we complement empirical policy shock evidence with structural assumptions on policy transmission, allowing us to study a wider range of counterfactual policy rules. This analysis allows us to shine light on the commonalities and differences in how popular macro models extrapolate beyond the available evidence, echoing the "sufficient statistics" results of the more recent trade and New Keynesian pricing literatures (e.g., as in Arkolakis et al., 2012; Auclert et al., 2022). Our conclusions on policy shock extrapolation across different models in particular connect with the "forward guidance puzzle" literature (Del Negro et al., 2023). Finally, the combination of direct empirical evidence and model-based policy causal effect extrapolation at the heart of our approach—plus our emphasis on invertibility and econometrician information sets—also distinguishes our analysis from Hebden and Winkler (2021), who rely exclusively on model-implied policy causal effects for policy evaluation.

Outline. We begin in Section 2 with the identification results. We present our methodology and discuss its theoretical properties in Section 3. Our applications to monetary policy counterfactuals follow in Sections 4 and 5. Section 6 concludes.

2 Identification result

This section presents the identification results that underlie our methodology. For exposition, we start in Section 2.1 with a simple static environment. We then in Sections 2.2 to 2.5 extend the analysis to the general infinite-horizon case.

2.1 Static model

To build intuition, we find it useful to first present a simplified version of our identification results in a static model environment. Going from this simple static benchmark to the full-blown dynamic case will then be relatively straightforward, with sequences replacing scalars and lag polynomials replacing matrices.

Environment. We consider a static stochastic economy that admits representation as a structural vector moving-average (SVMA):

$$y_t = \Theta \varepsilon_t, \tag{1}$$

where y_t denotes a vector of macroeconomic aggregates, the n_ε-dimensional shock vector ε_t is distributed as

$$\varepsilon_t \sim N(0, I), \tag{2}$$

and the $n_y \times n_\varepsilon$-dimensional matrix Θ denotes the impulse response of macroeconomic outcomes y_t to a date-t vector of shocks ε_t.[4]

The SVMA system (1) allows for an unrestricted linear, static transmission from shocks to outcomes. We will now impose further restrictions on the economic environment generating this mapping Θ. Specifically, we suppose that the economy can be summarized by

$$\mathcal{H}_x x_t + \mathcal{H}_z z_t + \mathcal{H}_e e_t = \mathbf{0}, \tag{3}$$

$$\mathcal{A}_x x_t + \mathcal{A}_z z_t + \mathcal{A}_v v_t = \mathbf{0}. \tag{4}$$

Here x_t is an n_x-dimensional vector of endogenous variables, z_t is a n_z-dimensional vector of policy instruments, e_t is a n_e-dimensional vector of structural shocks, v_t is an n_v-dimensional vector of policy shocks, and we let $y_t \equiv (x_t', z_t')'$, $\varepsilon_t \equiv (e_t', v_t')'$. Equation (3) summarizes the n_x-dimensional non-policy block of the model, with $\{\mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_e\}$ embedding private-sector relations. Equation (4) is the policy rule, with the instrument z_t set as a function of x_t and v_t; to further simplify the discussion we will assume that $n_z = n_v = 1$—i.e., one policy instrument and one policy shock. While the purpose of this section is illustrative, we do note that some familiar models—most notably the textbook New Keynesian model with transitory shocks (Galí, 2015)—can actually be written in the form (3) - (4) (see Appendix A.1).

Given the shocks $\{e_t, v_t\}$, an equilibrium of this economy is a pair $\{x_t, z_t\}$ that solves (3) - (4). We assume that such an equilibrium exists and is unique, and we let Θ denote this mapping from shocks to outcomes. This delivers the SVMA (1).

For future reference, it will be useful to note that the model environment (3) - (4) embeds three economically meaningful restrictions. The first two are that it is *static* and *linear*. The third one is that policy shapes private-sector outcomes (i.e., equation (3)) *only* through the value of the instrument z_t. In other words, whether or not z_t is set to a certain value because of the systematic component of policy (i.e., $\mathcal{A}_x$ and $\mathcal{A}_z$) or because of a policy shock (i.e., v_t) is irrelevant. Our analysis in Sections 2.2 to 2.5—and the related "VAR-Plus" method—will relax the first restriction, but keep the other two.

We conclude the description of the environment by defining two further objects. The first object is the covariance matrix of the macroeconomic observables y_t, given as

$$\Gamma_y = \Theta\Theta'. \tag{5}$$

[4]The normality assumption is made purely for notational convenience, as this allows us to write conditional expectations instead of linear projections.

The second object is the Wold representation of y_t. Given that our environment is static, we can write the Wold representation with orthogonalized innovations as

$$y_t = \Psi u_t, \tag{6}$$

where $\text{Var}(u_t) = I$ and Ψ is an n_y-dimensional matrix such that $\Psi\Psi' = \Gamma_y$, e.g., the lower-triangular Cholesky factor of Γ_y.

Objects of interest. We now wish to study the evolution of this economy if policy had instead followed the alternative rule

$$\tilde{\mathcal{A}}_x x_t + \tilde{\mathcal{A}}_z z_t = \mathbf{0} \tag{7}$$

in place of (4). Note that this counterfactual rule is followed perfectly, without any additional shocks.[5] Proceeding as above, the macroeconomic observables y_t under the counterfactual policy rule would follow the counterfactual SVMA process

$$\tilde{y}_t = \tilde{\Theta}\varepsilon_t, \tag{8}$$

with the convention that $\varepsilon_t = e_t$, and where $\tilde{\Theta}$ is derived from the solution of the system (3) together with (7). We are now interested in two particular policy counterfactual questions.

1. **Unconditional business cycles.** How would the "average" business cycle have unfolded if the policymaker had instead followed the alternative policy rule (7)? Or, more specifically, we seek the counterfactual second moments of y_t, given as

$$\tilde{\Gamma}_y = \tilde{\Theta}\tilde{\Theta}'. \tag{9}$$

2. **Conditional episodes.** How would the economy have evolved at a certain point in time if the policymaker had instead followed the rule (24)? That is, for a particular time t, we seek to recover

$$\tilde{y}_t = \tilde{\Theta}\varepsilon_t \tag{10}$$

We next discuss the information required to construct these counterfactuals of interest.

[5] Our identification results can easily be extended to allow for shocks to the counterfactual policy rule. We regard the shock-free case as the practically relevant one, however.

Identification result. Let the n_y-dimensional vector θ_v denote the causal effects of the policy shock v_t on macroeconomic outcomes $y_t = (x_t', z_t')'$. Our identification result states conditions under which knowledge of these policy shock causal effects—together with only reduced-form objects—suffices to recover the counterfactuals of interest.

Proposition 1. *Suppose that the SVMA process (1) is invertible; i.e., that*

$$\varepsilon_t \in \text{span}(y_t). \tag{11}$$

Then knowledge of: (i) the Wold representation of y_t (i.e., the covariance matrix Γ_y and the orthogonalized innovations u_t); and (ii) policy causal effects θ_v suffices to construct all policy counterfactuals of interest—$\tilde{\Gamma}_y$ and $\tilde{y}_t$.

Before providing the formal proof, we find it useful to first discuss the high-level intuition underlying this identification result. For this we will proceed in two steps, first assuming that the researcher could actually observe all of the primitive structural shocks ε_t (rather than just y_t and the associated Wold representation). In that case, it is immediate that she could recover the causal effects of those shocks under the baseline policy rule, Θ. She could then leverage the results of McKay and Wolf (2023): since the researcher knows how any possible value of the policy instrument z_t affects macroeconomic outcomes (i.e., θ_v), she can predict how those observed shocks would have counterfactually propagated under the alternative rule (7)—i.e., she has obtained $\tilde{\Theta}$, thereby the counterfactual SVMA representation (8), and thus the desired policy counterfactuals.

The identification result then goes one step further and states that, under the additional assumption of invertibility (i.e., under condition (11)), directly observing the true ε_t's is actually not necessary—it suffices to just observe the reduced-form Wold innovations u_t. The reason for this is simply that, under invertibility, the unknown true shocks ε_t are a one-to-one function of the Wold innovations u_t. Since (counterfactual) predictions based on a one-to-one function of the true shocks equal predictions based on the shocks themselves, the researcher is able to recover the correct counterfactuals.

Proof. We give a constructive argument for our two policy counterfactuals of interest. Each argument will rely on the following building block result. Consider using the policy shock causal effect vector θ_v to predict the counterfactual propagation of the orthogonalized Wold innovations u_t under the counterfactual policy rule (7), proceeding as in McKay and Wolf (2023, Proposition 1). Formally, for $j \in \{1, \ldots, n_y\}$, let the vector Ψ_j denote the impulse response of y_t to the j-th Wold innovation $u_{j,t}$, and then construct the counterfactual impulse responses $\tilde{\Psi}_j$ as

$$\tilde{\Psi}_j = \Psi_j + \theta_v \tilde{v}_j \tag{12}$$

where the artificial policy shock $\tilde{v}_j$ solves the equation

$$\tilde{A}_x \left(\Psi_{x,j} + \theta_{x,v}\tilde{v}_j\right) + \tilde{A}_z \left(\Psi_{z,j} + \theta_{z,v}\tilde{v}_j\right) = 0, \tag{13}$$

and where we have partitioned Ψ_j and θ_v into the responses of x and z. In words, for each Wold innovation j, we add a policy shock $\tilde{v}_j$ to the original Wold impulse response, with the policy shock chosen to ensure that the counterfactual policy rule holds. The $\tilde{\Psi}_j$ are then the impulse responses to the original Wold innovation plus this artificial policy shock.

Combining the $\tilde{\Psi}_{\bullet,j}$'s for all j, we get the counterfactual process

$$\tilde{y}_t = \tilde{\Psi} u_t \tag{14}$$

The key insight underlying Proposition 1 is that the process (14) can be used to recover the true counterfactuals of interest. To see why, note that, under invertibility, the orthogonalized Wold innovations u_t and true structural shocks ε_t are related as

$$u_t = P\varepsilon_t, \tag{15}$$

where P is an orthogonal matrix. It then follows from McKay and Wolf (2023) that

$$\tilde{\Psi} = \tilde{\Theta} P'. \tag{16}$$

Recovering the desired counterfactuals is now straightforward.

1. Consider using the counterfactual process (14) to recover the desired counterfactual second-moment properties. Its implied covariance matrix is

   $$\tilde{\Psi}\tilde{\Psi}' = \tilde{\Theta}P'P\tilde{\Theta}' = \tilde{\Theta}\tilde{\Theta}' = \tilde{\Gamma}_y,$$

   where the first equality uses (16), and the second follows since P is an orthogonal matrix.

2. Consider using (14) to recover the desired counterfactual for a particular historical episode. We have

   $$\tilde{\Psi}u_t = \tilde{\Theta}P'P\varepsilon_t = \tilde{\Theta}\varepsilon_t = \tilde{y}_t,$$

where the first equality uses (15) together with (16), and the second follows since P is an orthogonal matrix.

We have thus recovered both $\tilde{\Gamma}_y$ as well as $\tilde{y}_t$, completing the argument.

$\square$

Summary & outlook. The main insight of Proposition 1 is that, for many counterfactual policy questions, knowledge of the primitive structural shocks driving aggregate fluctuations is not necessary—it suffices to combine simple reduced-form projections (i.e., the Wold representation of the data) with the causal effects of policy shocks. The remainder of this section generalizes these insights to dynamic settings, paving the way for our empirical "VAR-Plus" methodology. The analysis will mirror our discussion in this section: first the environment, then the objects of interest, and finally the identification result.

2.2 Environment

Our dynamic identification results apply to a family of linearized infinite-horizon structural macroeconomic models with aggregate risk. The results are most easily stated and proved using linearized perfect-foresight ("sequence-space") notation.[6] Using this notation will also allow us to connect seamlessly with the static analysis in Section 2.1.

Our description of the economic environment proceeds in two steps. First, we begin by introducing the structural vector moving-average (SVMA) representation of our economy. Second, we present the linearized perfect-foresight system whose transition paths equal the impulse responses collected in the SVMA coefficients.

Stochastic economy. We assume that our stochastic economy admits representation as a general SVMA(∞):

$$y_t = \sum_{\ell=0}^{\infty} \Theta_\ell \varepsilon_{t-\ell}. \tag{17}$$

y_t is again a vector of macroeconomic aggregates, the shock vector ε_t is distributed as

$$\varepsilon_t \sim N(0, I), \tag{18}$$

[6]By certainty equivalence, solutions to linearized perfect-foresight systems correspond to impulse response functions in linearized economies with aggregate risk (Fernández-Villaverde et al., 2016; Auclert et al., 2021).

and the $n_y \times n_\varepsilon$-dimensional matrices Θ_ℓ denote the impulse response of the vector of macroeconomic observables y_t at horizon ℓ to a date-t vector of shocks ε_t. Relative to the simple static case, the only difference is that we have now moved from an SVMA(0) to an SVMA(∞). We will throughout impose the high-level assumption that the matrices Θ_ℓ are absolutely summable across ℓ. Finally, in all of the following, the notation $\mathbb{E}_t[\bullet]$ will be reserved for expectations conditioning on the sequence of shocks $\{\varepsilon_{t-\ell}\}_{\ell=0}^{\infty}$ up to date t. Consistent with the classical Frisch (1933) impulse-propagation paradigm, the SVMA(∞) system (17) allows for an unrestricted dynamic linear transmission from shocks ε_t to outcomes y_t.

Impulse-response system. Leveraging the equivalence between linearized systems with aggregate risk and perfect-foresight transition paths, we obtain the impulse responses Θ_ℓ as solutions of a linear, perfect-foresight, infinite-horizon economy. Below boldface denotes time paths for $t = 0, 1, 2, \ldots$, and all variables are expressed in deviations from the deterministic steady state. The economy is summarized by the system

$$\mathcal{H}_w \boldsymbol{w} + \mathcal{H}_x \boldsymbol{x} + \mathcal{H}_z \boldsymbol{z} + \mathcal{H}_e e_0 = \mathbf{0}, \tag{19}$$

$$\mathcal{A}_x \boldsymbol{x} + \mathcal{A}_z \boldsymbol{z} + \mathcal{A}_v v_0 = \mathbf{0}. \tag{20}$$

where as before $y_t = (x_t', z_t')'$ and $\varepsilon_t = (e_t', v_t')'$. The dynamic system (19) - (20) generalizes the static case in two ways. First, it adds the n_w-dimensional vector of endogenous variables w_t. The distinction between w and x is that all of the variables in x are observable (to the econometrician), while those in w are not. Second, it is an infinite-horizon system, with sequences replacing scalars.[7] The shocks are now dated at 0, and the system characterizes impulse responses to these innovations. Given the shocks $\{e_0, v_0\}$, an equilibrium is a set of bounded sequences $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}\}$ that solve (19) - (20). We assume that the policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$ is such that an equilibrium exists and is unique for any $\{e_0, v_0\}$. We write the implied mapping from shocks to outcomes as

$$y_\ell = \Theta_\ell \cdot \varepsilon_0. \tag{21}$$

Stacked together, those perfect-foresight mappings from date-0 shocks to date-ℓ outcomes deliver the SVMA(∞) representation (17).

[7] The boldface vectors $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}\}$ stack time paths for all variables (e.g., $\boldsymbol{x} = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_{n_x}')'$). The maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_e\}$ and $\{\mathcal{A}_x, \mathcal{A}_z, \mathcal{A}_v\}$ are conformable and map bounded sequences into bounded sequences.

While adding dynamics, the full system (19) - (20) maintains the other two key properties of the static set-up of Section 2.1: linearity, and how policy is allowed to shape private-sector behavior.[8] For our purposes here, the key thing to note is that many of the explicit, parametric structural models used for counterfactual policy analysis in the classical "medium-scale DSGE" approach actually satisfy these two properties, from representative-agent New Keynesian models (Christiano et al., 2005; Smets and Wouters, 2007), to heterogeneous-agent environments (Kaplan et al., 2018), and including some models with behavioral frictions (like Gabaix, 2020). A thorough discussion of the scope and limitations of systems like (19) - (20) is presented in McKay and Wolf (2023).

Some definitions. We conclude our description of the environment by introducing the autocovariance function and Wold representation, mirroring the discussion of the static model. Under our assumptions on (17), the autocovariance function $\Gamma_y(\bullet)$ of macroeconomic observables y_t exists and by standard arguments is given as

$$\Gamma_y(\ell) = \sum_{m=0}^{\infty} \Theta_m \Theta'_{m+\ell}. \tag{22}$$

Next, the Wold representation of y_t is

$$y_t = \sum_{\ell=0}^{\infty} \Psi_\ell u_{t-\ell}, \tag{23}$$

where $u_t^\dagger \equiv y_t - \mathbb{E}\left(y_t \mid \{y_\tau\}_{-\infty<\tau\leq t-1}\right)$ denotes one-step-ahead forecast errors, $\text{Var}(u_t^\dagger) = \Sigma_u$, and $u_t \equiv \text{chol}(\Sigma_u)^{-1} u_t^\dagger$ are orthogonalized Wold innovations, with $\text{Var}(u_t) = I$ and $\text{chol}(\bullet)$ giving the lower-triangular Cholesky factor. Our assumptions on (17) ensure that this Wold representation exists, features no deterministic term, and that $\Psi(L)$ is square-summable.

2.3 Objects of interest

As in the static case, we wish to study the evolution of the economy if policy was set as

$$\tilde{\mathcal{A}}_x \boldsymbol{x} + \tilde{\mathcal{A}}_z \boldsymbol{z} = \boldsymbol{0} \tag{24}$$

[8] In the dynamic case, the second property means that policy affects private-sector outcomes only through current and (expected) future values of the policy instrument, $\boldsymbol{z}$. Mathematically, this is again reflected in the private-sector linear maps $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_e\}$ being invariant to the policy rule, so that policy enters the private-sector block (19) only through the path $\boldsymbol{z}$.

in place of (20). The macroeconomic observables y_t under the counterfactual policy rule would then follow the counterfactual SVMA process

$$\tilde{y}_t = \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \varepsilon_{t-\ell}, \tag{25}$$

with the convention that now $\varepsilon_t = e_t$, and where the shock impulse responses $\tilde{\Theta}_\ell$ are derived from the solution of the perfect-foresight system (19) together with (24).

Differently from the static case, our dynamic analysis here will require us to tackle some subtleties on when precisely the counterfactual rule (24) is followed. In particular, the counterfactual SVMA (25) embeds the assumption that the counterfactual rule (24) is actually followed *forever*. In some of our "conditional" counterfactuals, however, we will assume that the policymaker instead unexpectedly changes to the alternative rule (24) at some date t^*, having followed the original rule (20) up to $t^* - 1$. In that case we will have

$$\tilde{y}_t \quad = \quad \underbrace{\sum_{\ell=0}^{t-t^*} \tilde{\Theta}_\ell \varepsilon_{t-\ell}}_{\text{new shocks after } t^*} \quad + \quad \underbrace{\tilde{y}_t^*}_{\text{initial conditions}} \tag{26}$$

The first term in (26) is straightforward: all newly arriving shocks ε_t propagate according to the new counterfactual impulse responses $\tilde{\Theta}_\ell$. The second term reflects initial conditions: at date t^*, the policymaker revises the planned policy path to ensure that current and expected future values of x and z are related according to (24). Letting $y_t^* = \mathbb{E}_{t^*-1}[y_t]$ denote date-$t^* - 1$ expectations under the initially prevailing rule, the initial conditions term $\tilde{y}_t^*$ can thus be obtained by solving the system[9]

$$\mathcal{H}_w(\tilde{\boldsymbol{w}}^* - \boldsymbol{w}^*) + \mathcal{H}_x(\tilde{\boldsymbol{x}}^* - \boldsymbol{x}^*) + \mathcal{H}_z(\tilde{\boldsymbol{z}}^* - \boldsymbol{z}^*) \quad = \quad \mathbf{0}, \tag{27}$$

$$\tilde{\mathcal{A}}_x \tilde{\boldsymbol{x}}^* + \mathcal{A}_z \tilde{\boldsymbol{z}}^* \quad = \quad \mathbf{0}. \tag{28}$$

We are now in a position to state our counterfactuals of interest.

1. **Unconditional business cycles.** This first "average" counterfactual is largely analogous to the static case—the only change is that static covariances are replaced by a general

[9]Here, boldface denotes sequences from t^* onwards. (28) says the new, counterfactual policy rule holds. By (27), the revised forecasts remain consistent with all private-sector relationships. Finally, under our assumptions on equilibrium existence and uniqueness, it follows that (27) - (28) has a unique solution.

dynamic autocovariance function. Specifically, we seek the counterfactual second moments of y_t, given as

$$\tilde{\Gamma}_y(\ell) = \sum_{m=0}^{\infty} \tilde{\Theta}_m \tilde{\Theta}'_{m+\ell}. \tag{29}$$

Unconditional "average" counterfactuals of this sort have attracted interest in prior work; examples include Rotemberg and Woodford (1997) or Del Negro and Schorfheide (2004).

2. **Conditional episodes.** For counterfactuals conditional on particular historical episodes, the analysis is more involved than in the static case. First, we need to take into account when precisely the switch to the counterfactual rule occurs. Second, we will allow for a distinction between realized and forecasted outcomes.

   (i) *Conditional forecasts.* Consider some date t^*, and suppose the policymaker from t^* commits to the new rule (24). We may ask how, from that point onward, the economy would be *predicted* to evolve; i.e., we would like to recover the expectation

   $$\mathbb{E}_{t^*}\left[\tilde{y}_{t^*+h}\right] = \tilde{\Theta}_h \varepsilon_{t^*} + \tilde{y}^*_{t^*+h}. \tag{30}$$

   Such conditional forecasts are key inputs for central banks (see Svensson, 1997) and have been studied widely in the academic literature (e.g., Antolin-Diaz et al., 2021).

   (ii) *Historical evolution.* Consider a particular episode, $t \in [t_1, t_1+1, \ldots, t_2]$. We may ask how the economy would have evolved over that time window if the policymaker had followed the rule (24) from date t_1 onward; i.e., we seek to recover

   $$\tilde{y}_t = \sum_{\ell=0}^{t-t_1} \tilde{\Theta}_\ell \varepsilon_{t-\ell} + \tilde{y}^1_t, \quad \forall t \in [t_1, t_1+1, \ldots, t_2] \tag{31}$$

   where $\tilde{y}^1_t \equiv \mathbb{E}_{t_1-1}\left[\tilde{y}_t\right]$ reflects initial conditions as of date t_1. Counterfactuals for particular historical episodes have also been the subject of much prior work. A recent example is Eberly et al. (2020).

Together, this list spans most of the policy counterfactuals studied in the literature.[10]

[10] Some papers—including McKay and Wolf (2023)—mainly focus on counterfactuals for particular shocks (e.g., an oil shock). Our counterfactuals here are instead unconditional or conditional on particular historical episodes, and thus require the counterfactual propagation of the *full* set of shocks hitting the macro-economy. As a result, the assumption of invertibility will take center stage when we turn to identification in Section 2.4. Such across-shock counterfactuals are arguably the most relevant ones for practical policy evaluation.

2.4 Identification result

This section presents the headline identification result underlying our methodology. We proceed as in the static case—first we introduce some additional notation on how policy can affect the economy, and then we state, prove, and discuss the identification result.

POLICY CAUSAL EFFECTS. Recall that the policy rule (20) is subject to the n_v-dimensional vector of policy shocks v_t. To state our identification result, we will follow McKay and Wolf (2023) and consider instead a *full* menu of policy shocks that perturb the policy rule at each possible horizon; that is, we have

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

where the policy shock vector $\boldsymbol{\nu}$ is now unrestricted—i.e., we allow for arbitrarily flexible wedges in the rule at each date $t = 0, 1, 2, \ldots$. Analogously to the discussion in Section 2.2, the solution of the system (19) - (20') given an arbitrary policy shock vector $\boldsymbol{\nu}$ alone yields

$$\boldsymbol{y} \;=\; \Theta_\nu \;\cdot\; \boldsymbol{\nu}. \tag{32}$$

Θ_ν gives us the space of allocations implementable through policy—i.e., the paths of macroeconomic aggregates corresponding to any possible path of the policy instrument.[11] Relative to the static case, this is now of course a much higher-dimensional object: the causal effects of policy are summarized not in a single policy shock impulse response vector θ_v, but in the infinite-dimensional linear map Θ_ν.

THE IDENTIFICATION RESULT. We are now in a position to state the identification result. Exactly as in the static case it reveals that, under the assumption of invertibility, knowledge of (i) reduced-form projections and (ii) the causal effects of policy on macroeconomic outcomes, Θ_ν, suffice to evaluate the counterfactuals of interest.

Proposition 2. *Suppose that the SVMA(∞) process* (17) *is invertible; i.e., that*

$$\varepsilon_t \in \text{span}(\{y_\tau\}_{-\infty<\tau\leq t}) \tag{33}$$

[11]Note that the policy causal effects Θ_ν are defined relative to the rule $\{\mathcal{A}_x, \mathcal{A}_z\}$. As discussed at length in McKay and Wolf (2023), this reference policy rule is best viewed as a possible "basis". Our subsequent identification result would apply for Θ_ν defined with respect to any (determinacy-inducing) policy rule.

Then knowledge of: (i) the Wold representation y_t (i.e., the history of innovations $\{u_{t-\ell}\}_{\ell=0}^{\infty}$ together with $\Psi(L)$); and (ii) policy causal effects Θ_ν suffices to construct all policy counterfactuals of interest—$\tilde{\Gamma}_y(\ell)$, $\tilde{y}_t$, and $\mathbb{E}_t[\tilde{y}_{t+h}]$.

Proposition 2 is the natural dynamic generalization of Proposition 1: invertibility means that the structural shocks ε_t lie in the span of current *and lagged* macroeconomic observables, and the causal effects of policy are now multi-dimensional, as discussed above. The intuition is also exactly analogous. First, if the researcher could observe all of the structural shocks ε_t, then identification would be straightforward and follow directly from McKay and Wolf (2023). Second, under invertibility, knowledge of the reduced-form Wold innovations u_t also suffices. In the dynamic case, all of our counterfactuals of interest are still just forecasts—and by invertibility, forecasts with respect to the econometrician's information set equal forecasts with respect to the full history of structural shocks ε_t. Since (counterfactual) forecasts based on a one-to-one function of the true shocks equal forecasts based on the shocks themselves, the researcher is able to recover the correct counterfactuals.

We note that the first part of Proposition 2—the identification argument for unconditional business cycles—is already contained in McKay and Wolf (Appendix A.5, 2023), while the arguments for the conditional episodes are new to the present paper.[12]

Proof. The proof proceeds as in the static case. Consider using the policy transmission map Θ_ν to predict the counterfactual propagation of the Wold innovations u_t under the counterfactual policy rule (24), proceeding as in McKay and Wolf (2023, Proposition 1). Formally, for $j \in \{1, \dots, n_y\}$, let $\Psi_{\bullet,j}$ be the impulse response of y_t to the j-th Wold innovation $u_{j,t}$, and then construct the counterfactual impulse responses $\tilde{\Psi}_{\bullet,j}$ as

$$\tilde{\Psi}_{\bullet,j} = \Psi_{\bullet,j} + \Theta_\nu \tilde{\boldsymbol{\nu}}_j \tag{34}$$

where the artificial policy shocks $\tilde{\boldsymbol{\nu}}_j$ solve the system of equations

$$\tilde{A}_x \left(\Psi_{\bullet,x,j} + \Theta_{x,\nu}\tilde{\boldsymbol{\nu}}_j\right) + \tilde{A}_z \left(\Psi_{\bullet,z,j} + \Theta_{z,\nu}\tilde{\boldsymbol{\nu}}_j\right) = \mathbf{0} \tag{35}$$

Combining the $\tilde{\Psi}_{\bullet,j}$'s for all j, we get the counterfactual process

$$\tilde{y}_t = \sum_{\ell=0}^{\infty} \tilde{\Psi}_\ell u_{t-\ell} \tag{36}$$

[12] Importantly, and differently from Hebden and Winkler (2021), our results for conditional episodes are cast in terms of econometrician observables, i.e., the autocovariance function of the data, and thus stress the centrality of the invertibility assumption. We discuss the implications of its possible violation in Section 2.5.

Under invertibility, the Wold innovations u_t and true structural shocks ε_t are related as

$$u_t = P\varepsilon_t, \tag{37}$$

where P is an orthogonal matrix. It then again follows from McKay and Wolf (2023) that the counterfactual Wold lag polynomial $\tilde{\Psi}(L)$ satisfies

$$\tilde{\Psi}(L) = \tilde{\Theta}(L)P'. \tag{38}$$

We now recover each of the desired counterfactuals.

1. Consider using the counterfactual process (36) to recover the desired counterfactual second-moment properties. Its implied autocovariance function is

$$\sum_{m=0}^{\infty} \tilde{\Psi}_m \tilde{\Psi}'_{m+\ell} = \sum_{m=0}^{\infty} \tilde{\Theta}_m P' P \tilde{\Theta}'_{m+\ell} = \sum_{m=0}^{\infty} \tilde{\Theta}_m \tilde{\Theta}'_{m+\ell} = \tilde{\Gamma}_y(\ell),$$

where the first equality uses (38), and the second follows since P is an orthogonal matrix.

2. Applying Proposition 1 of McKay and Wolf (2023) to the system (27) - (28) that defines initial conditions $\tilde{\boldsymbol{y}}^*$, we see that we can recover initial conditions at t^* as

$$\tilde{\boldsymbol{y}}^* = \boldsymbol{y}^* + \Theta_\nu \tilde{\boldsymbol{\nu}}^*, \tag{39}$$

where the artificial policy shocks $\tilde{\boldsymbol{\nu}}^*$ now solve

$$\tilde{A}_x \left(\boldsymbol{x}^* + \Theta_{x,\nu}\tilde{\boldsymbol{\nu}}^*\right) + \tilde{A}_z \left(\boldsymbol{z}^* + \Theta_{z,\nu}\tilde{\boldsymbol{\nu}}^*\right) = \mathbf{0}.$$

Note that our informational requirements (i) - (ii) suffice to construct $\tilde{\boldsymbol{\nu}}^*$ and thus allow us to also evaluate the initial conditions term $\tilde{\boldsymbol{y}}^*$. In particular, invertibility here is crucial to ensure that $\boldsymbol{x}^*$ and $\boldsymbol{z}^*$ are equal to date-$t^* - 1$ forecasts based on the Wold representation (23), given as $y^*_{t^*+h} = \sum_{\ell=1}^{\infty} \Psi_{h+\ell} u_{t^*-\ell}$. We can now recover the two counterfactuals.

   (i) Consider using (36) and (39) to recover the conditional forecast $\mathbb{E}_t\left[\tilde{y}_{t+h}\right]$. We have

$$\tilde{\Psi}_h u_{t^*} + \tilde{y}^*_{t^*+h} = \underbrace{\tilde{\Psi}_h P}_{=\tilde{\Theta}_h} \varepsilon_{t^*} + \tilde{y}^*_{t^*+h} = \mathbb{E}_{t^*}\left[\tilde{y}_{t^*+h}\right],$$

   as desired.

(ii) Consider using (36) and (39) to recover the historical counterfactual $\tilde{y}_t$. We have

$$\sum_{\ell=0}^{t-t_1} \tilde{\Psi}_\ell u_{t-\ell} + \tilde{y}_t^1 = \sum_{\ell=0}^{t-t_1} \underbrace{\tilde{\Psi}_\ell P}_{=\tilde{\Theta}_\ell} \varepsilon_{t-\ell} + \tilde{y}_t^1 = \tilde{y}_t,$$

as desired.

□

Section 3 will show how to leverage this identification result for our "VAR-Plus" methodology. Before doing so, however, we in Section 2.5 briefly elaborate further on the role of the crucial invertibility assumption (33).

2.5 The invertibility assumption

What happens if the invertibility assumption in Proposition 2 fails? Mathematically, without invertibility, the orthogonalized reduced-form Wold innovations u_t are related to the true structural shocks ε_t as (e.g., see Wolf, 2020)

$$u_t = P(L)\varepsilon_t, \tag{40}$$

where $P(L)$ is a so-called Blaschke matrix—i.e., the matrix-polynomial generalization of an orthogonal matrix. Following the same steps as in the proof of Proposition 2, the term $P'P$ would now be replaced by $P^*(L^{-1})P(L)$ (where * denotes conjugate transposition). However, and unlike $P'P$, this term is generally *not* equal to the identity matrix. As a result, the proof of Proposition 2 breaks down (see Appendix A.2 for further details).

Of course, even without invertibility, it is always possible to just mechanically follow the steps described in the proof of Proposition 2 and so arrive at some (mis-specified) predicted counterfactuals. By the logic of the argument given there, whether or not the end result of doing so will actually be close to the true counterfactuals will be governed by whether forecasts based on the information set $\{y_{t-\ell}\}_{\ell=0}^{\infty}$ are close to full-information forecasts based on $\{\varepsilon_{t-\ell}\}_{\ell=0}^{\infty}$. We illustrate this observation in Appendix A.2. There, we consider the structural model of Smets and Wouters (2007) as an artificial data-generating process, and implement the steps of Proposition 2 for different information sets. We show that even relatively small information sets can deliver forecasts close to the full-information limit, and thus deliver approximate counterfactuals that are almost indistinguishable from the true ones.

The practical takeaway from this discussion is that, when leveraging the identification result in Proposition 2, researchers should ensure that their conclusions are not sensitive to further additions to the information set $\{y_{t-\ell}\}_{\ell=0}^{\infty}$. We will make sure that this is the case in our applications in Section 5.

3 Policy counterfactuals via "VAR-Plus"

By Section 2, constructing our policy counterfactuals of interest requires two key inputs: (i) reduced-form projections, and (ii) policy dynamic causal effects. We now discuss our "VAR-Plus" approach to constructing those inputs. We first present the method in Sections 3.1 to 3.3, and then discuss its theoretical properties in Section 3.4.

3.1 Estimation of reduced-form projections

As the first step in our methodology, the researcher selects a set of macroeconomic observables y_t and then estimates their Wold representation (23)—i.e., she recovers the orthogonalized Wold innovations u_t and the lag polynomial $\Psi(L)$. In principle there are many ways of doing so. One particularly simple and convenient alternative is to estimate a VAR(p) in y_t:

$$y_t = \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + u_t^\dagger \tag{41}$$

The orthogonalized Wold innovations u_t are then equal to $\text{chol}(\Sigma_u)^{-1} u_t^\dagger$ (where $\Sigma_u = \text{Var}(u_t^\dagger)$), and the Wold lag polynomial $\Psi(L)$ is given as $(I - A(L))^{-1}\,\text{chol}(\Sigma_u)$. In practice, the vector of observables y_t should be chosen to be large enough—and contain enough forward-looking variables—so that the invertibility assumption is plausible (Fernández-Villaverde et al., 2007; Forni et al., 2019). Overall this step is entirely reduced-form, only requiring the econometrician to consider a sufficiently large information set.

3.2 Recovering policy causal effects

Our methodology next requires the policy causal effects Θ_ν. For this we will proceed in two steps. First, we use semi-structural time-series methods to get empirical evidence at least on *parts* of Θ_ν—the second part of the "VAR" step. Second, we use economic structure—in the form of one or multiple models of policy transmission—to first match and then extrapolate beyond that evidence, giving the rest of Θ_ν—the "Plus" part of our approach.

Empirical evidence on Θ_ν. In the first step, the researcher uses the standard time-series toolkit—typically in the form of a Structural Vector Autoregression (SVAR) or Local Projection (LP)—to estimate the dynamic causal effects of a list of n_ν distinct policy shocks. For example, for monetary policy applications, she may estimate the causal effects of short-lived and more persistent innovations to the federal funds rate, following identification strategies as in Romer and Romer (2004) or Gertler and Karadi (2015). We then stack those estimated impulse responses of n_m targeted outcome variables over H impulse response horizons to the n_ν identified shocks in the $n_\nu \times n_m \times H$ vector $\hat{\theta}_\nu$.

Under standard asymptotic sampling theory, the asymptotic distribution of the policy shock causal effect vector $\hat{\theta}_\nu$ satisfies (e.g., see Christiano et al., 2010)

$$\hat{\theta}_\nu \overset{a}{\sim} N(\theta_\nu, V_{\theta_\nu}). \tag{42}$$

Our methodology requires the researcher to have an at least approximately consistent estimator of the asymptotic covariance matrix V_{θ_ν}. We discuss standard options and our preferred approach for doing so in Appendix B.2.

Impulse response extrapolation. The second step begins with the researcher writing down a list $\mathcal{M}$ of structural models of policy transmission, denoted by $\mathcal{M}_j$ for $j = 1, 2, \ldots, M$. In the notation of Section 2.2, a "model" is a tuple $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$—i.e., a set of private-sector relations, but *not* any shocks to those relations, nor a policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$. Each model has a parameter vector ψ_j mapping into $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$, a prior distribution $p(\psi_j \mid \mathcal{M}_j)$ for the model parameters, and a prior probability $p(\mathcal{M}_j)$. We write $\theta_\nu(\psi_j, \mathcal{M}_j)$ as the model-implied analogue of the empirically observed policy shock causal effect vector; briefly, this object is defined as impulse responses to a change in policy in the model that comes as close as possible to the empirical targets. A detailed discussion of how to construct this object for any given model is provided in Appendix B.3.[13]

Each model $\mathcal{M}_j$ among the list of contemplated models is estimated through standard impulse-response matching techniques (Rotemberg and Woodford, 1997; Christiano et al., 2005, 2010). Cast as a standard limited-information Bayesian estimation strategy, we can

[13]In the standard impulse response-matching literature, the researcher writes down a policy rule, and then restricts attention to contemporaneous shocks to that rule. We instead find the best fit to the empirical targets within the overall *space* implementable by policy, allowing us to not need to commit to any rule.

define an approximate likelihood of the "data," $\hat{\theta}_\nu$, as a function of ψ_j given $\mathcal{M}_j$:

$$p(\hat{\theta}_\nu \mid \psi_j, \mathcal{M}_j) \propto \exp\left[-0.5\left(\hat{\theta}_\nu - \theta_\nu(\psi_j, \mathcal{M}_j)\right)' V_{\theta_\nu}^{-1}\left(\hat{\theta}_\nu - \theta_\nu(\psi_j, \mathcal{M}_j)\right)\right]. \tag{43}$$

Combining the prior together with the likelihood (43), we obtain the posterior for ψ_j conditional on model $\mathcal{M}_j$ and given the policy shock causal effect data $\hat{\theta}_\nu$:

$$p(\psi_j \mid \hat{\theta}_\nu, \mathcal{M}_j) = \frac{p(\hat{\theta}_\nu \mid \psi_j, \mathcal{M}_j)p(\psi_j \mid \mathcal{M}_j)}{p(\hat{\theta}_\nu \mid \mathcal{M}_j)} \tag{44}$$

and where

$$p(\hat{\theta}_\nu \mid \mathcal{M}_j) = \int p(\hat{\theta}_\nu \mid \psi_j, \mathcal{M}_j)p(\psi_j \mid \mathcal{M}_j)d\psi_j \tag{45}$$

is the marginal density of $\hat{\theta}_\nu$ given model $\mathcal{M}_j$. Computation of these objects is standard, relying on the usual random walk Metropolis-Hastings algorithm both to draw from the posterior distribution and to compute the marginal likelihood. See Appendix B.3.

The final step is to recover posterior model probabilities—i.e., the posterior distribution across the model space $\mathcal{M}$. We have

$$p(\mathcal{M}_j \mid \hat{\theta}_\nu) = \frac{p(\hat{\theta}_\nu \mid \mathcal{M}_j)p(\mathcal{M}_j)}{\sum_{i=1}^M p(\hat{\theta}_\nu \mid \mathcal{M}_i)p(\mathcal{M}_i)} \tag{46}$$

Putting everything together, the researcher is left with a posterior distribution over models and parameter vectors, $p(\psi_j, \mathcal{M}_j \mid \hat{\theta}_\nu)$. Each parameterized model implies a policy transmission map

$$\Theta_\nu \quad = \quad \Theta_\nu(\psi_j, \mathcal{M}_j). \tag{47}$$

We have thus overall arrived at the desired posterior distribution over the causal effects of policy on macroeconomic aggregates Θ_ν, $p(\Theta_\nu)$.[14]

3.3 Constructing policy counterfactuals

As the third step in our proposed approach, it remains to put together the estimated inputs—i.e., reduced-form Wold innovations and projection coefficients $\{u_t, \Psi(L)\}$ as well as policy

[14] In practice, we define Θ_ν with respect to a standard, determinacy-inducing policy rule; i.e., we recover Θ_ν by combining the model-implied $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$ with a policy rule (20'), proceeding as in Section 2.4. As discussed in McKay and Wolf (2023), our results will be invariant to this choice of "basis."

shock causal effects Θ_ν—to construct the three policy counterfactuals of interest. The formulas for mapping $\{u_t, \Psi(L)\}$ together with Θ_ν into our three desired counterfactuals are provided in the proof of the identification result of Section 2.4.

An important practical implementation challenge is how to take into account estimation uncertainty for the inputs $\{u_t, \Psi(L), \Theta_\nu\}$. For this we propose to proceed as follows. First, for the reduced-form inputs $\{u_t, \Psi(L)\}$, we simply look at point estimates. This is in keeping with standard practice in the policy counterfactual literature, which tends to take as given point estimates for the baseline second moments and forecasts (e.g., Rotemberg and Woodford, 1997; Eberly et al., 2020). Second, given those point estimates, we construct the policy counterfactuals by drawing Θ_ν from the posterior distribution estimated in our second step. Given the point estimates of $\{u_t, \Psi(L)\}$, the posterior distribution over Θ_ν thus maps into a posterior distribution over the counterfactuals $\{\tilde{\Gamma}(\ell), \mathbb{E}_t[\tilde{y}_{t+h}], \tilde{y}_t\}$.[15]

3.4 Discussion of our approach

Our "VAR-Plus" methodology contributes to the recent literature on counterfactual policy evaluation with less reliance on model structure (e.g., Barnichon and Mesters, 2023; McKay and Wolf, 2023). Relative to that literature, our approach is more broadly applicable, owing to: (i) the generalized identification result underlying it; and (ii) the use of *partial* model structure for policy causal effect extrapolation, if necessary. As such, our strategy provides an alternative to the standard "quantitative DSGE" approach to policy counterfactual analysis. Under that approach, the researcher writes down a fully specified parametric model of the macro-economy—including in particular the driving stochastic shock processes—, estimates that model using likelihood-based methods, and finally uses it as a laboratory for policy analysis (e.g., see Smets and Wouters, 2007; Justiniano et al., 2010). We will argue that our methodology has several appealing properties relative to this standard approach, notably in regard to its robustness to model mis-specification.

Robustness. Our method begins with simple reduced-form and semi-structural time series analysis—i.e., the "VAR" part—, but then also relies on model structure to recover (at least some) policy causal effects—i.e., the "Plus" part. As such, it is inevitably subject to concerns about model mis-specification. However, that being said, it is strictly more robust than the

[15] Conceptually it would be straightforward to allow for joint estimation uncertainty, by using the same VAR for estimation of both $\{u_t, \Psi(L)\}$ and $\hat{\theta}_\nu$. This would give a joint distribution over $\{u_t, \Psi(L), \hat{\theta}_\nu\}$ and so, via the model estimation step, over $\{u_t, \Psi(L), \Theta_\nu\}$. We leave this extension to future work.

dominant full-information DSGE approach, in two main senses.

First, if the counterfactual of interest is spanned by the available evidence $\hat{\theta}_\nu$ on policy shock causal effects, then our approach is actually robust to *arbitrary* forms of model mis-specification within the general linearized environment (19) - (20) that underlies our identification result.[16] In essence, in that case, our approach is in fact purely semi-structural—the VAR step is already sufficient, without any further need for the "Plus" part. Our applications will reveal that this case is actually quite plausible in practice.

Second, since the model-based "Plus" part of our method does not require the researcher to specify either the shocks driving the macro-economy or the prevailing (in-sample) policy rule, our strategy is robust to arbitrary mis-specification *in those model blocks*. Prior work has argued that mis-specification of the shock processes is often a particularly relevant concern (e.g., see Chari et al., 2009): it is not clear where exactly in the model those shocks should enter, how many there should be, and what stochastic processes they should follow. Our method sidesteps those concerns: the identification result reveals that only the reduced-form *projections* generated by those unknown shocks matter; as a result, the only purpose of model structure is policy causal effect extrapolation, which requires no further assumptions on the unknown shock processes.[17] Similar concerns have also been voiced about mis-specification of policy rules; in fact, in practice, it is unlikely that historical policy conduct can actually be reduced to any simple rule. Either way, the likely mis-specification in those model blocks will generally affect inference on policy transmission, threatening the validity of the standard DSGE approach. We elaborate further on this discussion in Appendix B.4.

Other appealing properties. We conclude by briefly noting that our approach also has several other, secondary appealing properties. First, it is arguably easier to implement than the usual likelihood-based DSGE approach. The most direct argument here is that the researcher only needs to *partially* specify the model; in particular, she need not contemplate where to add shocks, and how precisely to model them. It is furthermore well-known that even models saturated with a large number of shocks struggle to actually match well the second-moment properties of observed time series data (e.g., see Angeletos et al., 2020); our approach sidesteps this challenge since second-moment properties are taken directly from the

[16]Strictly speaking, this of course presupposes that at least one of the contemplated models can match the targeted empirical evidence $\hat{\theta}_\nu$ well. In our applications we will make sure that this is the case.

[17]To evaluate welfare, knowledge of the nature of the primitive shocks driving the economy is important, as also argued by Chari et al. (2009). We here are instead concerned with the more modest objective of evaluating counterfactual paths of measurable macro outcomes (like output and inflation).

data. Second, our approach is amenable to comparison across models and thus in particular allows the researcher to easily incorporate model uncertainty. The only purpose of model structure is policy causal effect extrapolation, and so model uncertainty matters if and only if two models of policy transmission disagree on how they extrapolate beyond the available empirical evidence. Our applications will illustrate this point, showcasing where precisely existing popular models disagree in terms of policy causal effect extrapolation. Third, the likelihood-based DSGE approach is well-known to be vulnerable to weak identification concerns (e.g., Fernández-Villaverde et al., 2016). A key advantage of impulse response-matching approaches is that they need only partially specify the model, thereby reducing the number of parameters that need to be identified. We discuss this point further in Appendix B.4.

Outlook. The remainder of the paper will showcase our methodology with monetary policy applications. We first discuss our approach to getting monetary policy causal effects—i.e., policy shock impulse response estimation, matching and extrapolation—in Section 4, and then present the applications in Section 5. The straightforward reduced-form forecasting step is briefly discussed in Section 5 itself.

4 Extrapolating monetary policy causal effects

Our analysis follows the general discussion in Section 3.2. We first in Section 4.1 review our empirical policy shock transmission targets $\hat{\theta}_\nu$, and then in Sections 4.2 and 4.3 implement the model estimation and extrapolation steps for a range of candidate structural models.

4.1 Monetary shock estimation targets

Our construction of the empirical targets $\hat{\theta}_\nu$ of monetary policy transmission follows recent advances in the time-series monetary policy shock literature. We identify a monetary policy shock series following Aruoba and Drechsel (2022), who extend the classic Romer and Romer (2004) analysis to allow for a larger policymaker information set. We then study the dynamic causal effects of this monetary policy shock series on the output gap as well as on inflation and the federal funds rate. The vector $\hat{\theta}_\nu$ stacks impulse responses of these three variables for the first five years after the shock.

Implementation & results. We only provide a brief discussion of our econometric implementation here, with further details relegated to Appendix C.1. We measure the output gap as de-trended real (per capita) GDP, following the procedure of Hamilton (2018); inflation and the federal funds rate enter in annualized percentage terms. We use quarterly data for the period 1969Q1–2006Q4, and we estimate shock dynamic causal effects by ordering the shock series first in a recursive VAR that contains the shock and our outcomes of interest, following the suggestions of Plagborg-Møller and Wolf (2021) and Li et al. (2023). The estimated VAR includes two lags, a constant, as well as a linear time trend.[18]

Results are reported as the grey lines (medians) and light grey areas (confidence bands) in Figure 1 (on p.32). The estimated causal effects of monetary policy shocks look largely as in prior work: a transitory hike in rates leads to a gradual and moderately persistent fall in output, as well as a more delayed fall in prices.

4.2 Structural models of monetary policy transmission

We consider several models of monetary transmission to match and extrapolate beyond the evidence $\hat{\theta}_\nu$ of Section 4.1. Our first model is a standard representative-agent model with nominal rigidities ("RANK"), augmented with several other frictions to allow a quantitative fit to our empirical evidence, following Christiano et al. (2005). Our second model enriches the consumer block to feature heterogeneous households with uninsurable income risk as in Kaplan et al. (2018) ("HANK"). Those first two models arguably capture the dominant approaches to quantitative business-cycle modeling of the past two decades, and thus constitute a natural starting point for our analysis of monetary policy causal effect extrapolation. We will, however, also consider behavioral variants of these models, with price- and wage-setters forming expectations with cognitive discounting, as in Gabaix (2020). We will see that such behavioral frictions can matter greatly for policy causal effect extrapolation at intermediate and long horizons, and that this matters for policy counterfactual evaluation.

The remainder of this section proceeds as follows. We begin by sketching the baseline representative-agent, rational-expectations model. We then explain how the heterogeneous-agent and behavioral models depart from this benchmark. Details are in Appendix C.2.

Baseline RANK model. Our first model is a standard quantitative business-cycle model, as familiar from the "medium-scale DSGE" tradition, and in particular rich enough to allow us to match the empirical monetary policy shock evidence documented in Section 4.1.

[18] The shock series of Aruoba and Drechsel (2022) is slightly shorter than our sample; we simply set the shock series equal to zero when not available, as in Känzig (2021).

Following Christiano et al., the model features capital accumulation subject to investment adjustment costs and with variable capacity utilization, nominal rigidities (with indexation) in price- and wage-setting, and habit formation in consumer preferences. We now provide a brief sketch of each of the constituent model blocks.

- *Households & unions.* The economy is populated by a representative household with preferences for consumption and leisure given by

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left[u\left(c_t - hc_{t-1}\right) - v\left(\ell_t\right)\right] \right] \tag{48}$$

  where h is a habit parameter, and we assume that $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$ and $v(x) = \nu \frac{x^{1+\varphi}}{1+\varphi}$. This agent chooses paths for consumption and assets a_t^H to maximize her utility subject to a standard no-Ponzi condition as well as the budget constraint

$$c_t + a_t^H = w_t(1-\tau_t^\ell)\ell_t + d_t^H - \tau_t + \frac{1+r_{t-1}^n}{1+\pi_t} a_{t-1}^H, \tag{49}$$

  where w_t is the real wage, τ_t^ℓ is the labor tax rate, d_t^H is real dividend income, τ_t is a transfer, r_t^n is the nominal interest rate, and π_t is the price inflation rate.

  Labor supply is intermediated by labor unions (as in Erceg et al., 2000), with households taking ℓ_t as given when solving their consumption-savings problem. The unions face Calvo-style frictions in adjusting their wages, with full indexation to lagged price inflation (as in Christiano et al., 2005). This yields the wage Phillips curve

$$\pi_t^w - \pi_{t-1} = \kappa_w \chi_t + \beta \mathbb{E}_t \left[\pi_{t+1}^w - \pi_t\right], \tag{50}$$

  where $\chi_t = \varphi \ell_t - (\lambda_t + w_t) + \frac{\bar{\tau}^\ell}{1-\bar{\tau}^\ell}\tau_t^\ell$ is the difference between the household's marginal rate of substitution and the after-tax wage, λ_t is the marginal utility of consumption, and π_t^w is the log change in the aggregate nominal wage index. As usual κ_w is a composite parameter given as $\kappa_w = (1-\theta_w)(1-\theta_w\beta)/\theta_w$, where $1-\theta_w$ is the probability that wages are re-optimized each period.

- *Production.* There is a unit continuum of perfectly competitive intermediate goods producers. They produce using capital and labor, and with a variable rate of capital capacity utilization (with curvature ζ). The intermediate good is sold at real relative price p_t^I to retailers; the retailers costlessly differentiate the good, set prices subject to Calvo-style frictions, and sell their variety to a competitive final goods aggregator. Prices that are not re-optimized are fully indexed to lagged inflation (as in Christiano et al., 2005), and so overall we get a Phillips curve of the form

$$\pi_t - \pi_{t-1} = \kappa_p p_t^I + \beta \mathbb{E}_t \left[\pi_{t+1} - \pi_t \right], \tag{51}$$

where the slope parameter κ_p is given as $\kappa_p = (1-\theta_p)(1-\theta_p\beta)/\theta_p$, with $1-\theta_p$ denoting the probability that prices are re-optimized.

The intermediate goods producers purchase capital goods from perfectly competitive capital goods producers. Those capital goods producers purchase the final good, turn it into the capital good subject to adjustment costs on their level of investment (with curvature κ), and then sell the capital good at real relative price q_t.

- *Policy.* There is a monetary and a fiscal authority. The fiscal authority issues nominal bonds with exponential maturity structure, spends a constant amount in real terms, and then adjusts labor taxes gradually to maintain long-run budget balance. In the representative-agent economy described here, this fiscal rule has real effects through the distortionary effects of taxes on labor supply. In the heterogeneous-agent economy sketched below, it also matters by affecting the timing of household income.

  The monetary authority sets nominal interest rates. Importantly, for our purposes—i.e., estimating the model to match $\hat{\theta}_\nu$ and then extrapolating to all of Θ_ν—we need not specify any particular monetary policy rule, as discussed above.

Stacking all model blocks except the behavior of the monetary authority, we obtain $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$—i.e., the first model block (19). Solving the system for every possible path of monetary policy shocks and thus nominal rates, we obtain the policy effects Θ_ν.

We estimate the model to ensure consistency with the empirical monetary policy shock targets $\hat{\theta}_\nu$. In particular, we estimate five parameters: the strength of habits (h), the degrees of price and wage rigidity (θ_p and θ_w), the curvature of investment adjustment costs (κ), and the curvature of capacity utilization costs (ζ). We estimate the model using a standard Bayesian implementation of impulse response-matching, so we impose priors on these five parameters; all remaining parameters are calibrated. See Appendix C.2 for details.

HANK MODEL. Our second model is a heterogeneous-agent ("HANK") model. It differs from the representative-agent baseline in that the representative consumer is replaced by a unit continuum of households subject to uninsurable idiosyncratic income risk and borrowing constraints (e.g., Kaplan et al., 2018), delivering elevated average marginal propensities to consume (MPCs). To ensure consistency with the empirically observed gradual response of output to changes in monetary policy, we furthermore assume that households are inattentive to macroeconomic conditions, as in Auclert et al. (2020). Unlike habits, this modeling choice delivers sluggish responses to changing aggregates while still maintaining large MPCs out of transitory income changes.

We provide a sketch of the household consumption-savings problem here, with further details in Appendix C.2. Households are subject to idiosyncratic income risk (with the risk process taken from Kaplan et al., 2018), and hours worked are intermediated by labor unions, as in the baseline representative-agent model.[19] Households save in government bonds, while firm capital and equity is held by financial intermediaries; those intermediaries gradually pay out dividends to households in proportion to their productivity. Letting $1 - \theta$ denote the probability that a household updates its information about aggregate conditions, and letting s denote the number of periods since the last update, the consumption-savings problem can be stated recursively as

$$V_t(a, e, s) = \max_{c,a'} \{u(c) - v(\ell_t) + \beta\mathbb{E}_{t-s} [\theta V_{t+1}(a', e', s + 1) + (1 - \theta)V_{t+1}(a', e', 0)]\}$$

subject to the budget constraint

$$c + a' = \left((1 - \tau_{\ell,t})w_t\ell_t + d_t^H\right) e + \frac{1 + r_{t-1}^n}{1 + \pi_t}a + \tau_t$$

and the borrowing constraint $a' \geq \underline{a}$, and where e denotes idiosyncratic household productivity. The borrowing constraint $\underline{a}$ is set as in Kaplan et al. (2018). The estimated model parameters are then the same as in the representative-agent model, except for the household information stickiness parameter (θ) replacing the degree of habit formation (h).

Cognitive discounting. Standard New Keynesian models, including the representative- and heterogeneous-agent variants presented so far, imply that inflation is strongly forward-looking. This model feature implies that small changes in future monetary policy can have large and immediate effects on inflation. The literature on the forward guidance puzzle (e.g.,

[19] We assume that unions evaluate the marginal utility of income using $c^{-\gamma}$ where c is aggregate consumption. As the Phillips curves are then unchanged, this assumption limits the effects of inequality to the demand side of the model (as in McKay and Wolf, 2022).

Del Negro et al., 2023) has questioned this feature of standard models.

For our final set of model variants, we will consider versions of our representative- and heterogeneous-agent baselines in which price- and wage-setting becomes less forward-looking. We follow Gabaix (2020) in assuming that agents engage in "cognitive discounting." According to this view, agents do not trust that they understand the structure of the economy and thus shrink their evaluation of future outcomes towards the economy's steady state. In particular, an innovation occurring s periods in the future is down-weighed by a factor m^s, where $m \in [0,1]$ controls the strength of cognitive discounting, and with $m = 1$ corresponding to the rational-expectations benchmark. With such cognitive discounting among price- and wage-setters, the Phillips curves become (see Appendix C.2)

$$\pi_t - \pi_{t-1} = \kappa_p p_t^I + \beta^p \mathbb{E}_t \left[\pi_{t+1} - \pi_t\right] \tag{52}$$

$$\pi_t^w - \pi_{t-1} = \kappa_w \chi_t + \beta^w \mathbb{E}_t \left[\pi_{t+1}^w - \pi_t\right] \tag{53}$$

with $\beta^p = \beta\theta_p m(1+\kappa_p/(1-\beta\theta_p m)) \leq \beta$ and $\beta^w = \beta\theta_w m(1+\kappa_w/(1-\beta\theta_w m)) \leq \beta$. Relative to our first two models, the behavioral models thus add one further parameter: the degree of cognitive discounting m. For our behavioral models we will consider the case of m fixed and set to $m = 0.65$, at the lower end of the range considered by Gabaix (2020). We make this choice because, as we will see, our data are only weakly informative about m.

4.3 Estimation results and policy extrapolation

We now use the empirical targets of Section 4.1 together with the four models of Section 4.2 to implement the methodology of Section 3, assuming a uniform prior across models. We proceed in two steps. First, we present results of the impulse response matching exercise—i.e., the ability of our models to match the targets, and the resulting posterior distribution across models. Second, we discuss how the different models extrapolate from the matched transitory policy shock causal effects to more persistent policy changes.

Model estimation via impulse response matching. We begin in Figure 1 by showing the monetary policy shock estimation targets (grey) and as well as the matched impulses at the posterior mode for each of our four models (beige and red, solid and dashed); the full posterior distributions for the estimated parameters are presented in Appendix C.3. We see that all of the models are able to match the empirical targets quite well: a transitory hike in nominal interest rates leads to a hump-shaped decline in output as well as a delayed decline in

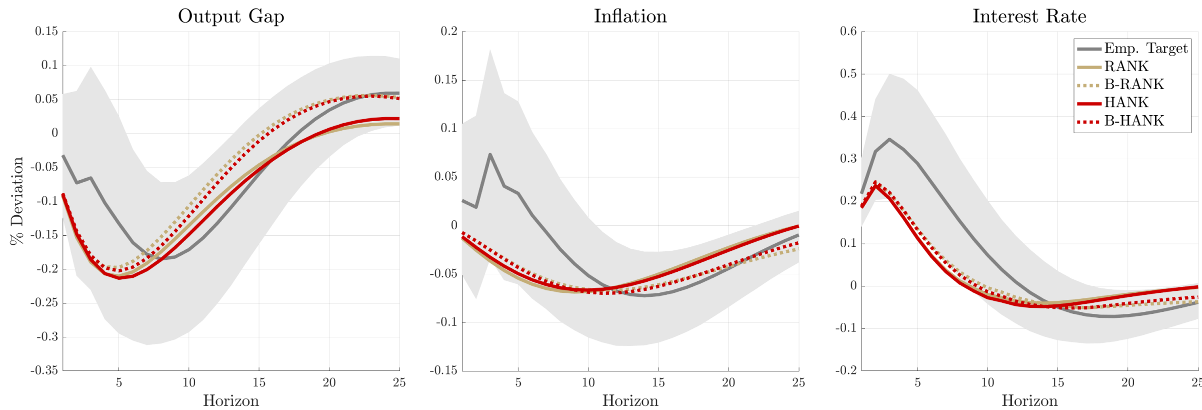

Figure 1: The grey line and shaded areas indicate the estimated impulse responses for a monetary policy shock (see Section 4.1), and their respective 16th and 84th percentile confidence bands. The remaining lines indicate the model-implied impulse responses at the estimated posterior modes. Beige: representative-agent consumer block. Red: heterogeneous-agent consumer block. Solid: no cognitive discounting. Dotted: cognitive discounting with $m = 0.65$.

Model	Baseline	Cognitive discounting	Total
RANK	0.4189	0.2435	0.6624
HANK	0.2183	0.1192	0.3376
Total	0.6372	0.3628	1.0000

Table 4.1: Posterior probabilities across the four models, assuming a uniform prior. The posterior model probabilities are computed as in (46).

inflation. This similar fit suggests that the data do not strongly distinguish between our four models; this visual impression is confirmed in Table 4.1, which shows posterior probabilities across models, computed following (46).[20]

Having estimated our models to match the empirical targets $\hat{\theta}_\nu$, we can now ask how the various estimated models extrapolate beyond that evidence on transitory interest rate changes to complete the full set of policy dynamic causal effects Θ_ν.

Monetary policy causal effect extrapolation. We will proceed in two steps: first comparing the baseline representative- and heterogeneous-agent models, and then asking how

[20]The analysis in Bundick and Smith (2020) as well as our results in Figures 2 and 3 suggest that empirical evidence on relatively near-term forward guidance would similarly be insufficient to discriminate between the models considered here. Furthermore, in results not reported here, we have also found that the fit across models is similar for the impulse responses of consumption and investment.

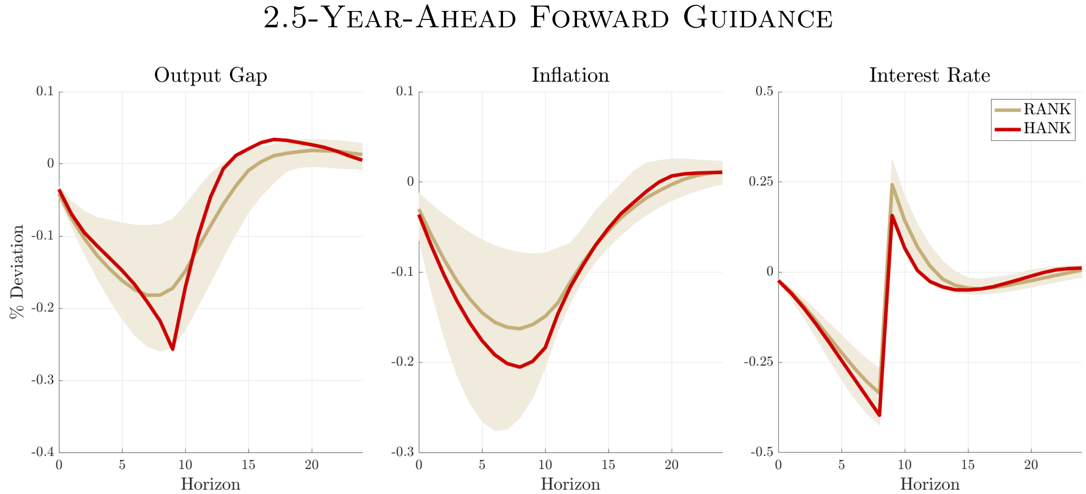

Figure 2: Policy causal effect extrapolation in our estimated RANK and HANK models. The figure shows output and inflation impulse responses (left and middle) to a forward guidance policy shock that leads to the interest rate movements depicted on the right. Beige: RANK model (solid = median, shaded = 16th and 84th percentile confidence bands). Red: HANK model (median).

cognitive discounting changes things. Results are displayed in Figures 2 and 3, which show impulse responses to forward guidance shocks—i.e., nominal interest rate movements that are (much) more delayed than our transitory targets $\hat{\theta}_\nu$.[21]

- *RANK vs. HANK.* We consider a far-ahead "forward guidance"-type monetary policy intervention: deviations from a standard monetary policy rule announced at $t = 0$ and occurring ten quarters later. The right panel shows the response of nominal interest rates, while the left and middle panels display the causal effects on output and inflation in RANK (beige) and HANK (red) associated with those interest rate paths.

  The main takeaway is that the two models—which by construction agree on the effects of the targeted interest rate path in Figure 1—also approximately agree on the dynamic causal effects of this much more delayed monetary intervention. In fact this is not just the case for the particular nominal rate paths that are shown in Figure 2; it holds robustly across different possible paths, i.e., across the entirety of Θ_ν.

- *Baseline vs. cognitive discounting.* Figure 3 repeats the same exercise, but now comparing the baseline (beige) and behavioral (blue) representative-agent models.[22]

[21] Details on how we construct those particular delayed interest rate paths are provided in Appendix C.3.

[22] We use the term "behavioral" as shorthand for "with cognitive discounting." The baseline HANK model also has elements that could be considered "behavioral" (sticky information among consumers).

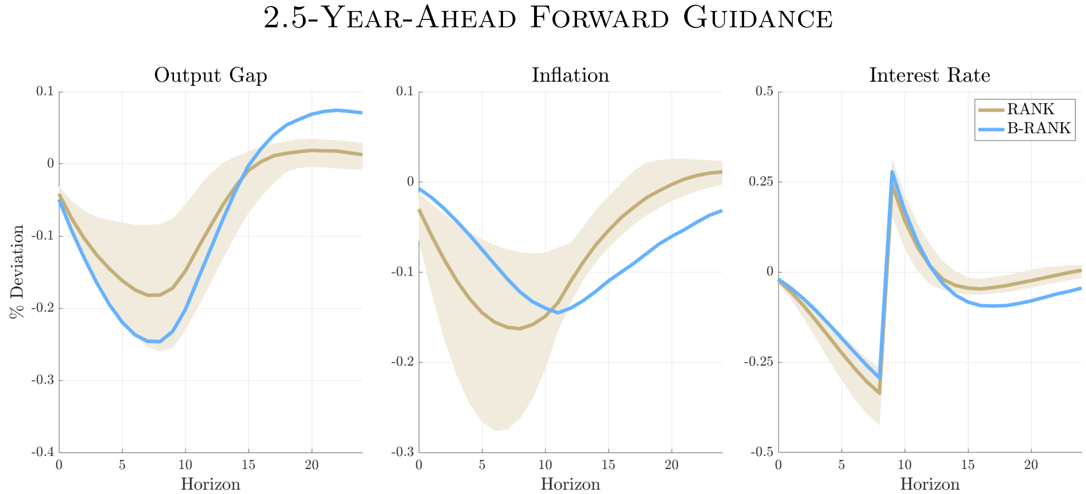

Figure 3: Policy causal effect extrapolation in our estimated RANK and B-RANK models. The figure shows output and inflation impulse responses (left and middle) to a forward guidance policy shock that lead to the interest rate movements depicted on the right. Beige: RANK model (solid = median, shaded = 16th and 84th percentile confidence bands). Blue: B-RANK model (median).

We see that now there are more material differences across the two models. In particular, and as expected given the additional discounting in the Phillips curves (52) - (53), we see that the inflation responses in the behavioral model are more delayed. While Figure 3 establishes this conclusion for the representative-agent models, we note that the same is also true for the heterogeneous-agent models.

Takeaways. In this section we have used our four structural models to match and extrapolate beyond the empirical evidence on monetary policy transmission, delivering the required full matrix of monetary policy causal effects on macroeconomic outcomes, Θ_ν. With an eye towards our applications in Section 5 we emphasize the following three conclusions.

1. All models can (approximately) match the targeted evidence. Thus, for any counterfactual that can (approximately) be enforced through the empirical policy evidence $\hat{\theta}_\nu$ alone, our conclusions in Section 5 will be entirely semi-structural—the first model mis-specification robustness property discussed in Section 3.4.

2. Representative- and heterogeneous-agent models, once estimated to match the same empirical evidence, also appear to extrapolate similarly beyond those targeted causal effects, to complete the entirety of Θ_ν. Thus, by the identification result in Section 2.4, it follows immediately that our two estimated RANK and HANK models of monetary transmission will imply very similar monetary policy counterfactuals in *all* possible applications.

3. For very persistent monetary policy interventions, important differences between our rational and behavioral models emerge, with the inflation response in the former much more forward-looking. Thus, for counterfactuals involving persistent monetary interventions, we expect meaningful differences across models, and thus large posterior uncertainty.

5 Applications to monetary policy counterfactuals

This section presents our various monetary policy counterfactual applications. We begin in Section 5.1 by estimating the required reduced-form projections and forecasts. Sections 5.2 to 5.4 then combine those reduced-form inputs with the distribution over policy causal effects from Section 4 to construct the counterfactuals.

5.1 Reduced-form projections

Following the discussion in Section 3.1, we begin by estimating a relatively large-dimensional reduced-form VAR. Implementation details are relegated to Appendix D.1.

The estimated reduced-form VAR is designed to be consistent with the theoretical discussion in Section 2. We consider a large number of observables, as in Angeletos et al. (2020). Differently from those authors, we transform all of the included variables to stationarity (if necessary), as in Hamilton (2018); in particular, we treat the thus-transformed real output series as a measure of the output gap. Our sample period stretches from 1960:Q1 – 2019:Q4—a long post-war but pre-covid sample. For our covid inflation counterfactual application (in Section 5.4) we then extend that sample to the chosen forecast date (2021:Q2).

5.2 Average business cycle

We use our methodology to study how a change in monetary policy design would have affected the "average" U.S. post-war business cycle. We communicate our results by focusing on the counterfactual volatilities of the output gap, inflation, and the federal funds rate.

The counterfactual monetary policy rule that we consider is the one that minimizes the following standard quadratic objective:

$$\mathcal{L} = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left\{ \lambda_\pi \pi_t^2 + \lambda_y y_t^2 + \lambda_i (i_t - i_{t-1})^2 \right\} \right], \tag{54}$$

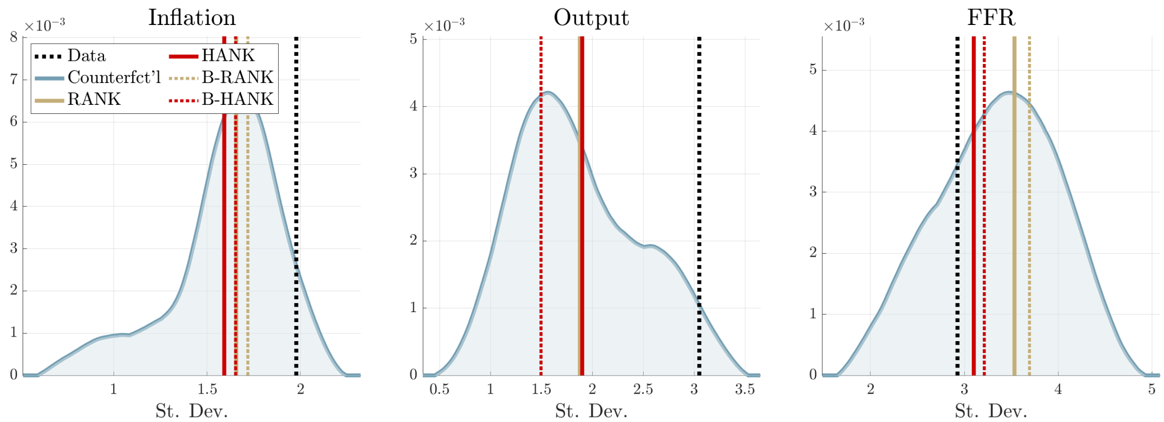

Figure 4: Counterfactual unconditional volatilities of inflation, output, and the federal funds rate, under the policy rule that minimizes (54). Black dashed: data point estimate under observed policy. Blue: posterior Kernel density of counterfactual volatilities drawing from posterior across all models and parameters. Beige: posterior mode of counterfactual using RANK models (baseline and behavioral). Red: posterior mode of counterfactual using HANK models (baseline and behavioral).

where inflation is measured in deviation from target.[23] Note that, even though our counterfactual rule is expressed implicitly as the minimizer of the loss function (54), we actually do not give it a normative interpretation; rather, we interpret our counterfactual as a reasonable approximation to the strategy of flexible inflation targeting, as discussed in Woodford (2003). Consistent with the discussion in Federal Reserve Tealbook (2016), we consider an equal-weights parameterization, with $\lambda_\pi = \lambda_y = \lambda_i = 1$ (and no discounting, i.e., $\beta = 1$).[24]

Main results. Results are reported in Figure 4. The figure shows actual (black-dashed) and counterfactual (blue, beige, red) volatilities of inflation, the output gap, and the federal funds rate. The black-dashed lines are constructed by translating the estimated reduced-form VAR into its implied Wold representation, and from there computing the three volatilities. For the counterfactuals, we first draw from the posterior over Θ_ν. For each draw, we then compute the volatilities under the counterfactual policy (following Proposition 2); finally, we construct a smoothed Kernel density estimate of the resulting posterior distribution.

The headline finding is that the counterfactual policy could have actually achieved materially lower inflation and (in particular) output gap volatilities, at the cost of only moderately

[23] That is, given any matrix of monetary policy causal effects Θ_ν, we set the counterfactual policy rule coefficients $\{\mathcal{A}_x, \mathcal{A}_z\}$ exactly as in Proposition 2 of McKay and Wolf (2023).

[24] Strictly speaking, with $\beta = 1$, the value of (54) is infinite. Our reported conclusions should thus be interpreted as corresponding to the limit $\beta \to 1$ (from below).

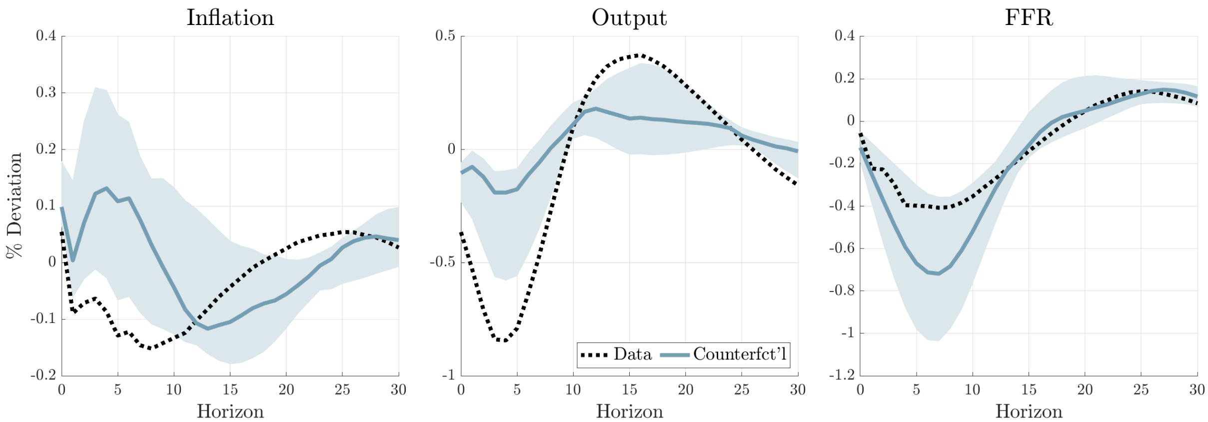

Figure 5: Counterfactual impulse responses of inflation, output, and the federal funds rate to the main business-cycle shock (see Angeletos et al., 2020), under the policy rule that minimizes (54). Black dashed: data point estimate under observed policy. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded).

more volatile nominal interest rates.[25] Furthermore, we note that the posterior uncertainty surrounding those counterfactuals—i.e., the light blue areas—primarily reflects uncertainty over parameters *within* models rather than differences *across* models. To demonstrate this, the beige and red lines show the counterfactuals at the posterior modes of the baseline and behavioral RANK and HANK models; as we can see, they are very close.

Inspecting the mechanism. To understand why such material volatility reductions would have been feasible, it will prove instructive to study the counterfactual propagation of one particular (reduced-form) combination of the Wold residuals—the "main business-cycle shock" of Angeletos et al. (2020). The black-dashed lines in Figure 5 show the propagation of this shock under the in-sample monetary reaction: inflation drops just a little, output drops materially, and monetary policy somewhat leans against this contraction. Our contemplated counterfactual policy simply leans against this shock much more, stabilizing output at the cost of moderately higher inflation and larger nominal interest rate movements.

Figure 5 also suggests that our finding of feasible volatility reductions is not really driven by model-based extrapolation of monetary policy effects—i.e., by the "Plus" step of our methodology. In that figure, in response to the "main business-cycle shock," our counterfactual policy calls for a decline in near-term interest rates, with a dynamic profile that is reasonably close to the policy shock we estimated empirically (see Figure 1), suggesting that

[25]It is important to note that this volatility reduction does not just reflect infeasible rate cuts during the period of a binding lower bound on nominal interest rates (see Appendix D.2).

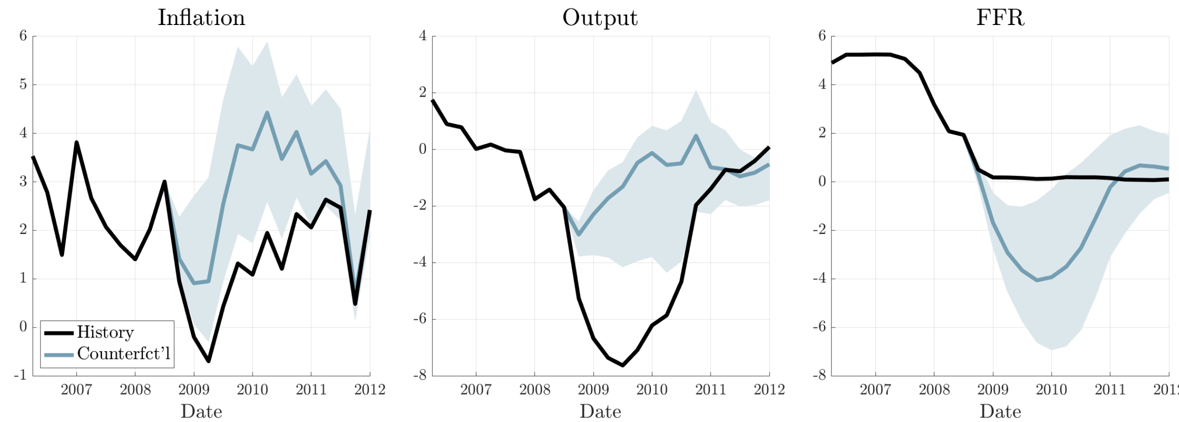

Figure 6: Counterfactual evolution of inflation, output, and the federal funds rate in the Great Recession, under the policy rule that minimizes (54) without any effective lower bound on rates. Black: data. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded).

model-based causal effect extrapolation is actually not adding all that much. To corroborate this intuition, we in Appendix D.2 repeat our analysis using only the empirically estimated policy causal effects $\hat{\theta}_\nu$ (rather than the entire model-implied matrix Θ_ν). As expected, we indeed find that meaningful volatility reductions, in particular for output, are feasible. The uncertainty documented in Figure 4 thus mainly reflects uncertainty about the (empirically estimable) causal effects of transitory interest rate changes, rather than disagreement across models in how they extrapolate to the effects of persistent rate changes.

5.3 Great Recession

We next use our methodology to evaluate how the economy would have evolved during the Great Recession if policy had instead followed the same flexible inflation targeting framework as above (i.e., minimizing (54)), and *without* any effective lower bound on nominal rates.[26] Specifically, we assume that the central bank follows this alternative, unconstrained rule from 2008:Q4 onwards, and does so throughout 2012:Q1—an example of our "historical evolution" counterfactual type. A counterfactual of this sort is informative about the plausible costs of a binding lower bound constraint. Results are reported in Figure 6.

We see that, absent any effective binding lower bound, a policy that follows the rule of minimizing (54) would have involved a very aggressive cut in interest rates, down to around

[26] We note that our methodology remains applicable to model environments with a linear non-policy block (19) and a non-linear policy rule, allowing for a binding lower bound on nominal interest rates. The argument is analogous to that in Appendix A.9 of McKay and Wolf (2023).

-4 per cent. Such (infeasible) nominal interest rate cuts would have materially reduced the output gap, at the cost of somewhat elevated inflation. Just as in Section 5.2, and for the same reasons as discussed there, the uncertainty displayed in Figure 6 chiefly reflects uncertainty about the causal effects of transitory rate changes (and not extrapolation uncertainty across models)—again, the counterfactual is largely pinned down by the empirically targeted causal effects of transitory rate changes. See Appendix D.3 for further details.

Our results are informative about the broader policy response to the Great Recession. Given constraints on nominal interest rates, policymakers attempted to substitute through other stimulative measures, notably unconventional monetary policy as well as fiscal stimulus. If we interpret (54) as the objective for monetary policy, our counterfactual suggests that the unconventional monetary policy response was insufficient—in nominal interest rate space, additional stimulus of around -4 per cent would have been necessary.

5.4 Post-covid inflation

As the third and final application of our methodology, we use it to evaluate monetary policy options at the height of the post-covid inflation. Using our estimated reduced-form VAR, we construct forecasts of inflation, the output gap, as well as the federal funds rate from 2021:Q2 onwards. We then apply our methodology to construct the forecast under the policy that minimizes the dual-mandate inflation targeting objective (54).

Main results. Figure 7 shows the historical evolution of inflation, the output gap, and the federal funds rate (black), their baseline forecasts from 2021:Q2 onwards (grey-dashed), and their forecasts under the counterfactual policy (blue). The three panels distinguish the models used to extrapolate the policy causal effects: the two rational-expectations models (top), the two behavioral models (middle), and our full set of all models (bottom). We see that, under the baseline forecast, inflation is expected to be persistently elevated, the output gap is slightly negative, and nominal interest rates rise sharply. Our focus is now on how the inflation targeting monetary policy moves the economy away from these baseline forecasts. Since the economy is predicted to be persistently away from steady state, we expect those counterfactuals to be sensitive to the degree of foresight embedded in the model(s) used for policy extrapolation, consistent with the discussion in Section 4.3.

Consider first the top panel of Figure 7, which shows counterfactuals constructed with policy extrapolation via our rational-expectations models. We see that policy succeeds in reducing inflation sharply with only a small reduction in output. Furthermore, and coun-

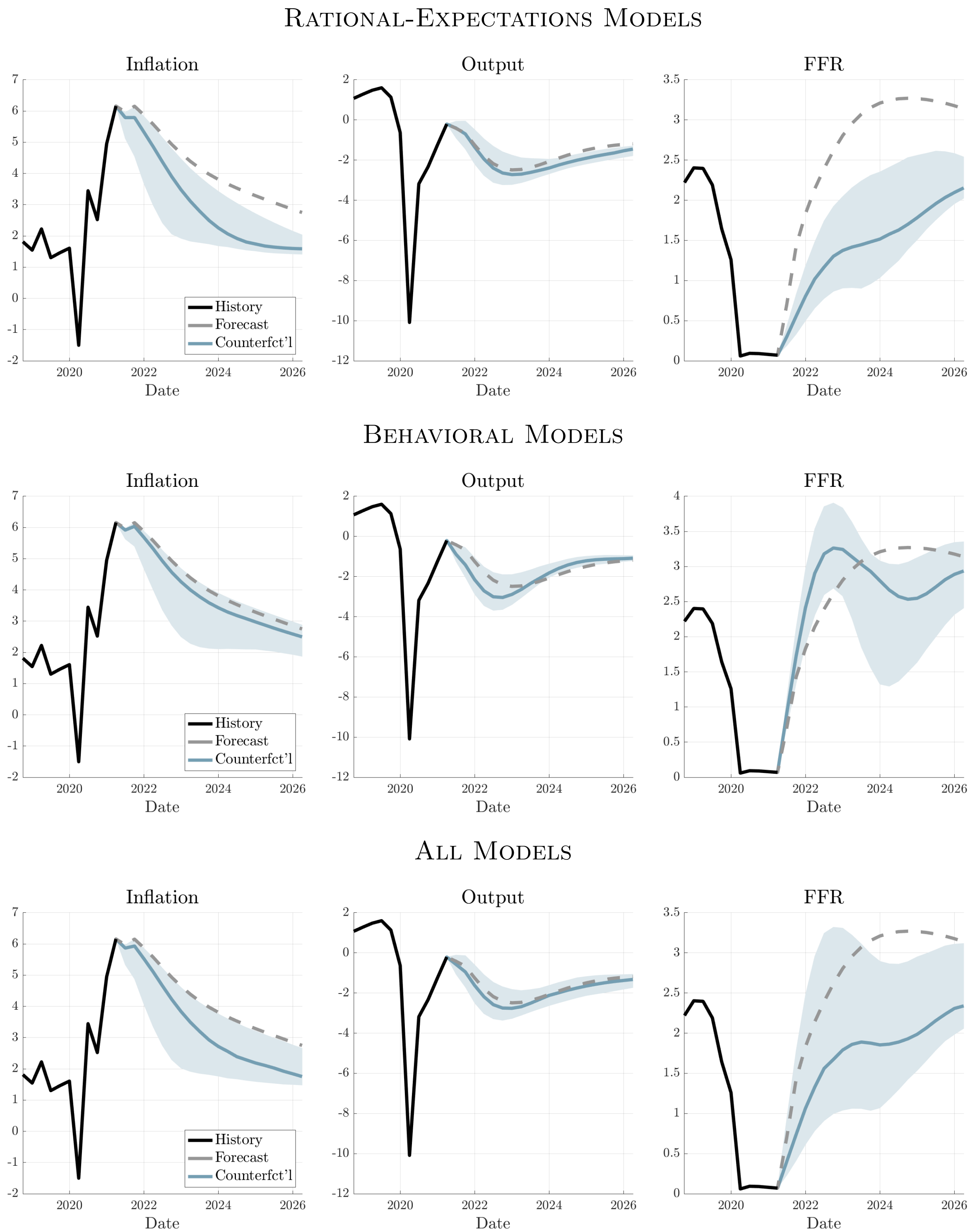

Figure 7: Counterfactual projections of inflation, output, and the federal funds rate in the post-covid inflationary episode (from 2021:Q2), under the policy rule that minimizes (54). Policy causal effects from rational-expectations models (top), behavioral models (middle), and all models jointly (bottom). Black: data. Grey: actual (VAR-implied) forecast. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded).

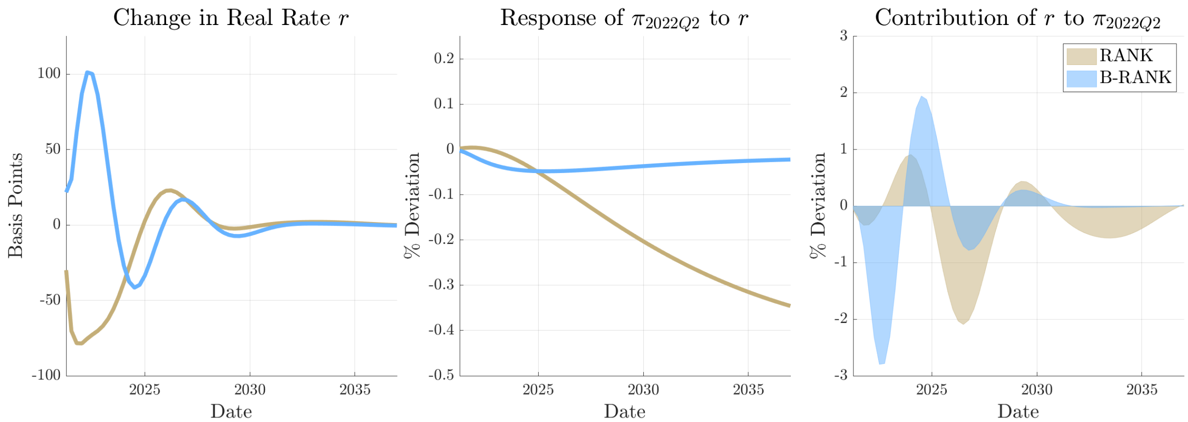

Figure 8: Behavior of real interest rates under the counterfactual monetary policy. Left panel: difference between the counterfactual and the forecast real interest rate path. Middle panel: change in inflation at 2022:Q2 in response to a policy-induced real interest rate change by horizon of the rate change, as indicated on the x-axis. Right panel: contribution of real interest rate changes at different horizons to the change in inflation at 2022:Q2. Beige: posterior mode of the baseline RANK model. Blue: posterior mode of the behavioral RANK model.

terintuitively, a *lower* interest rate path achieves this disinflation. This result reflects the extremely forward-looking nature of the model: the policymaker achieves low inflation in the short-run via negative output gaps in the far-future, implemented through future increases in real rates; in fact, small future output gaps move current inflation so much that *lower* short-term real rates can actually be used to stabilize output in the short run. The beige lines in Figure 8 illustrate this intuition: real rates initially decline and only later rise (left panel); short-run inflation is much more sensitive to real rates in the far-future than to real rates today (middle panel); combining the two, it follows that near-term disinflation can be achieved through moderate medium- and long-term real rate hikes (right panel).

Consider next the middle panel of Figure 7, which constructs our policy counterfactuals using instead the behavioral models to extrapolate policy causal effects. The counterfactual looks very different: the federal funds rate is hiked somewhat *more* aggressively than in the baseline forecast, bringing inflation down somewhat faster, and at the cost of moderately lower output. Intuitively, the policymaker now cannot rely on very far-ahead real interest rate movements to stabilize inflation in the short run. Instead, she faces an undesirable short-run trade-off between output and inflation, and now responds to it through *higher* short-term real rates. A visual illustration is provided with the blue lines in Figure 8: real interest rates rise immediately (left panel); short-run inflation is not nearly as sensitive to long-run real rate fluctuations (middle panel); as a result, the short-term inflation reduction largely reflects short-term real rate movements.

Finally, the bottom panel of Figure 7 puts everything together, showing counterfactuals with policy causal effects extrapolated from our full set of four models. The main message of the figure is the very large uncertainty about the inflation targeting nominal rate response. As discussed in Section 4.3, our estimation targets—the causal effects of transitory interest rate changes—do not discriminate between this set of models. As seen above, however, those models do have quite dramatically different implications for what (quite persistent) paths of nominal interest rates will best stabilize output and inflation. The end result are quite wide uncertainty bands, with the 68 per cent probability band for the federal funds rate in 2023 reaching from below 1 per cent to 3.3 per cent.

Discussion & future work. Our results for the post-covid counterfactual in Figure 7 connect very closely with the "forward guidance puzzle" literature (e.g., see Del Negro et al., 2023; Carlstrom et al., 2015; McKay et al., 2016). During and after the Great Recession, there was considerable interest in understanding forward guidance as an unconventional policy tool. Our results demonstrate that the model pathologies uncovered in that literature matter even away from any binding lower bound—they will apply whenever the economy is persistently away from its steady state. As shown in Section 4.3, the available empirical evidence on transitory changes in monetary policy does not discriminate between models with dramatically different causal effects of persistent policy changes; as a result, for monetary policy counterfactuals involving persistent interest rate changes, our method will invariably indicate large posterior policy counterfactual uncertainty.

This discussion suggests that, for the purposes of counterfactual policy evaluation, there will be high returns to future empirical work on the causal effects of persistent changes in monetary policy. Failing that, relying on other pieces of evidence to discriminate across the kinds of models considered here should prove useful. In particular, our results indicate that researchers should aim to distinguish between models with and without behavioral frictions; on the other hand, market incompleteness—the focus of the recent "HANK" literature—seems less central for the purpose of monetary policy causal effect extrapolation.

6 Conclusions

The main contribution of this paper is to propose a new "VAR-Plus" approach to evaluating the counterfactual evolution of the macro-economy under alternative assumptions on policy design. Leveraging a theoretical identification result, our method relies on reduced-form or semi-structural empirical evidence whenever possible, and then complements that evidence with additional structural assumptions whenever necessary. We have argued that this new approach is more robust to model mis-specification than the dominant DSGE paradigm: if empirical evidence already suffices to pin down the counterfactual, then our approach is in fact semi-structural; and even if the evidence does not suffice, our approach still relies on strictly less model structure, in particular not requiring the researcher to develop a complete account of the origins of business-cycle fluctuations.

Our analysis suggests two avenues for future work. First, *empirically*, there are very high returns to analyses identifying the causal effects of persistent changes in monetary policy—i.e., the causal effects of forward guidance-type policies. Second, *theoretically*, it would be useful to gain a more complete understanding of the range of models that can be consistent with the available evidence on monetary policy propagation, and then of how they differ in extrapolating beyond that evidence.

References

Angeletos, G.-M., F. Collard, and H. Dellas (2020): "Business-cycle anatomy," *American Economic Review*, 110, 3030–70.

Angeletos, G.-M. and C. Lian (2023): "Determinacy without the Taylor principle," *Journal of Political Economy*, 131, 2125–2164.

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

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018): "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications," *Econometrica*, 86, 685–720.

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

Aruoba, S. B. and T. Drechsel (2022): "Identifying Monetary Policy Shocks: A Natural Language Approach," CEPR Discussion Paper No. DP17133.

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2021): "Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models," *Econometrica*, 89, 2375–2408.

Auclert, A., R. D. Rigato, M. Rognlie, and L. Straub (2022): "New Pricing Models, Same Old Phillips Curves?" Tech. rep., National Bureau of Economic Research.

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

Barnichon, R. and G. Mesters (2023): "A Sufficient Statistics Approach for Macro Policy," *American Economic Review*, 113, 2809–2845.

Bundick, B. and A. L. Smith (2020): "The dynamic effects of forward guidance shocks," *Review of Economics and Statistics*, 102, 946–965.

Carlstrom, C. T., T. S. Fuerst, and M. Paustian (2015): "Inflation and output in New Keynesian models with a transient interest rate peg," *Journal of Monetary Economics*, 76, 230–243.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2009): "New Keynesian models: Not yet useful for policy analysis," *American Economic Journal: Macroeconomics*, 1, 242–266.

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

Christiano, L. J., M. Trabandt, and K. Walentin (2010): "DSGE models for monetary policy analysis," in *Handbook of monetary economics*, Elsevier, vol. 3, 285–367.

Crump, R. K., M. Del Negro, K. Dogra, P. Gundam, D. Lee, R. Nallamotu, and B. Pacula (2023): "The New York Fed DSGE Model Perspective on the Lagged Effect of Monetary Policy," Tech. rep., Federal Reserve Bank of New York.

Del Negro, M., M. P. Giannoni, and C. Patterson (2023): "The forward guidance puzzle," *Journal of Political Economy Macroeconomics*, 1, 43–79.

Del Negro, M. and F. Schorfheide (2004): "Priors from general equilibrium models for VARs," *International Economic Review*, 45, 643–673.

Eberly, J. C., J. H. Stock, and J. H. Wright (2020): "The Federal Reserve's Current Framework for Monetary Policy: A Review and Assessment," *International Journal of Central Banking*, 16, 5–71.

Erceg, C. J., D. W. Henderson, and A. T. Levin (2000): "Optimal monetary policy with staggered wage and price contracts," *Journal of Monetary Economics*, 46, 281–313.

Federal Reserve Tealbook (2016): "Report to the FOMC on Economic Conditions and Monetary Policy," .

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of understanding VARs," *American economic review*, 97, 1021–1026.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016): "Solution and estimation methods for DSGE models," in *Handbook of macroeconomics*, Elsevier, vol. 2, 527–724.

Forni, M., L. Gambetti, and L. Sala (2019): "Structural VARs and noninvertible macroeconomic models," *Journal of Applied Econometrics*, 34, 221–246.

Frisch, R. (1933): "Propagation problems and impulse problems in dynamic economics," *Essays in Honor of Gustav Cassel*.

Gabaix, X. (2020): "A behavioral New Keynesian model," *American Economic Review*, 110, 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. and P. Karadi (2015): "Monetary policy surprises, credit costs, and economic activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

Geweke, J. (1999): "Using simulation methods for bayesian econometric models: inference, development,and communication," *Econometric Reviews*, 18, 1–73.

Hamilton, J. D. (2018): "Why You Should Never Use the Hodrick-Prescott Filter," *The Review of Economics and Statistics*, 100, 831–843.

Hebden, J. and F. Winkler (2021): "Impulse-Based Computation of Policy Counterfactuals," FEDS Working Paper.

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

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

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary policy according to HANK," *American Economic Review*, 108, 697–743.

Kilian, L. and H. Lütkepohl (2017): *Structural vector autoregressive analysis*, Cambridge University Press.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2023): "Local projections vs. VARs: Lessons from Thousands of DGPs," Working Paper.

McKay, A., E. Nakamura, and J. Steinsson (2016): "The power of forward guidance revisited," *American Economic Review*, 106, 3133–3158.

McKay, A. and C. Wolf (2022): "Optimal policy rules in hank," Tech. rep., Working Paper, FRB Minneapolis.

McKay, A. and C. K. Wolf (2023): "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?" Working Paper.

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

——— (2022): "Instrumental variable identification of dynamic variance decompositions," *Journal of Political Economy*, 130, 2164–2202.

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

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

Ruge-Murcia, F. J. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31, 2599–2636.

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

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

Wolf, C. K. (2020): "Svar (mis) identification and the real effects of monetary policy shocks," *American Economic Journal: Macroeconomics*, 12, 1–32.

——— (2023): "The missing intercept: A demand equivalence approach," Tech. rep., National Bureau of Economic Research.

Woodford, M. (2003): *Interest and prices: Foundations of a theory of monetary policy*, Princeton University Press.

Appendix for:
Evaluating Policy Counterfactuals:
A "VAR-Plus" Approach

This appendix contains supplemental material for the article "Evaluating Policy Counterfactuals: A "VAR-Plus" Approach." We here provide: (i) some additional theoretical results to complement the discussion in Section 2; (ii) practical implementation details for our method as described in Section 3; (iii) further information on the structural models that we use for monetary policy shock impulse response matching and extrapolation in Section 4; and (iv) supplementary details for our empirical applications in Section 5.

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

Contents

A Supplementary theoretical results

This appendix provides supplementary theoretical results. Appendix A.1 discusses examples of specific structural models nested in our general framework of Section 2, and Appendix A.2 elaborates on how our identification result is affected by non-invertibility.

A.1 Nested models

We begin with an example of a structural model that is nested in our simple static environment (3) - (4)—the textbook small-scale New Keynesian model:

$$\begin{aligned}
y_t &= \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\gamma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right]\right) + e_t^d \\
\pi_t &= \kappa y_t + \beta \mathbb{E}_t\left[\pi_{t+1}\right] + e_t^s \\
i_t &= \phi_\pi \pi_t + v_t,
\end{aligned}$$

where e_t^d and e_t^s denote demand and supply shocks, respectively, v_t is the policy shock, and the rest of the notation is as in the standard textbook (Woodford, 2003; Galí, 2015). Assuming all innovations are transitory, and focusing on the model's fundamental solution (Angeletos and Lian, 2023) (or assuming a policy rule that uniquely implements that solution), the system then further simplifies to

$$\begin{aligned}
y_t &= -\frac{1}{\gamma} i_t + e_t^d \\
\pi_t &= \kappa y_t + e_t^s \\
i_t &= \phi_\pi \pi_t + v_t,
\end{aligned}$$

It is immediate that this system fits into the structure (3) - (4) with $x_t = (y_t, \pi_t)'$ and $z_t = i_t$.

Dropping the assumption of transitory shocks (or adding further sources of endogenous persistence, like for example an inertial policy rule), we arrive at a system that instead fits into the more general dynamic form (19) - (20). Analogous arguments apply to even richer quantitative (linearized) DSGE models, HANK-type environments, or models with certain behavioral frictions. Detailed worked-out examples along those lines are provided in McKay and Wolf (Appendix A.1, 2023), and thus omitted here.

A.2 Non-invertibility

We here provide additional details on our discussion of non-invertibility in Section 2.5. We first elaborate on why the proof of Proposition 2 fails in the absence of invertibility, and then provide a quantitative illustration of the effects of non-invertibility using a structural macro model as a laboratory. We close with some general discussion.

Proposition 2 without invertibility. As discussed in Section 2.5, without invertibility, the orthogonalized reduced-form residuals u_t satisfy

$$u_t = P(L)\varepsilon_t.$$

The Wold lag polynomial $\Psi(L)$ then satisfies

$$\Psi(L)P(L) = \Theta(L),$$

Using that $P(L)P^*(L^{-1}) = I$, it then follows from the arguments in McKay and Wolf (2023) that the artificial Wold lag polynomial $\tilde{\Psi}(L)$ constructed in our proof of Proposition 2 satisfies

$$\tilde{\Psi}(L) = \tilde{\Theta}(L)P^*(L^{-1}).$$

Proceeding from here, however, the proof strategy of Proposition 2 fails since it is generally the case that $P^*(L^{-1})P(L) \neq I$.

Non-invertibility in a structural model. We now use a structural model—the well-known medium-scale DSGE of Smets and Wouters (2007)—as a laboratory to illustrate what happens to our procedure in the absence of invertibility. The nature of the exercise is as follows. We consider a researcher that wishes to predict the counterfactual average second-moment properties of short-term nominal rates, output, and inflation under a counterfactual monetary policy rule. The monetary authority is assumed to follow rules of the form

$$i_t = \rho_i i_{t-1} + (1 - \rho_i)\left(\phi_\pi \pi_t + \phi_y y_t\right). \tag{A.1}$$

We assume that the researcher observes data generated from the posterior mode parameterization of Smets and Wouters (2007), but with the monetary policy rule taking the particular form (A.1) with $\phi_\pi = 1.5$ and $\rho_i = \phi_y = 0$. She wishes to predict the counterfactual second-moment properties of interest if instead the monetary authority followed the rule (A.1) but with $\rho_i = 0.8$, $\phi_\pi = 1.5$ and $\phi_y = 0.5$.[27]

To construct the desired counterfactuals, the researcher leverages Proposition 2. As inputs, she uses (i) the full, true matrix of policy shocks Θ_ν to the baseline policy rule, together with (ii) reduced-form projections based on information sets of different size. In particular, we consider four possible information sets: interest rates, output, and inflation alone ("baseline"); the baseline plus hours worked; the baseline plus investment and consumption; and finally the baseline plus hours worked, wages, investment as well as consumption. Note that, among these information sets, only the fourth one satisfies the invertibility assumption underlying Proposition 2. We thus know that only the counterfactuals constructed based on that fourth information set are guaranteed to equal the true counterfactuals.

Results are shown in Figure A.1. The top panel depicts spectral densities of interest rates, output, and inflation over business-cycle frequencies (i.e., from 6 to 32 quarters). The pink lines indicate spectral densities under the initially prevailing monetary policy rule, while the other lines indicate the true (solid) and predicted (dashed) counterfactual spectral densities. In all panels, we have normalized the peak spectral density (looking across both the old policy rule and the true counterfactual) to 1. Three findings stand out. First, the true spectral densities under the old and counterfactual policy rules are very different. This is by design—we picked two rules that respond very differently to cyclical fluctuations, and thus also imply very different overall cyclical fluctuations. Second, the predicted counterfactual based on the full information set equals the true one. Third, for all information sets, the predicted counterfactual spectral densities are close to each other, and so to the true counterfactual. In other words, even for the smallest information set, the estimand of our econometric strategy is very close to the actual true counterfactual.

To understand this finding, recall from the discussion surrounding Proposition 2 and in Section 2.5 that our identification result relies on forecasts with respect to the econometrician information set being close to full-information forecasts. The results in the top panel of Figure A.1 thus suggest that even the smaller information sets considered by the researcher allow forecasts almost as good as the full information set. The bottom panel of Figure A.1—which shows residual forecast uncertainty for interest rates, output, and inflation at different horizons (x-axis), and for different information sets (different lines)—confirms this intuition.

[27] We have chosen these two policy rules because they imply quite starkly different second-moment properties, with the first one aggressively stabilizing inflation, while the second smoothes interest rates and also stabilizes output. We have found results very similar to those reported in this section for a wide range of baseline and counterfactual policy rules.

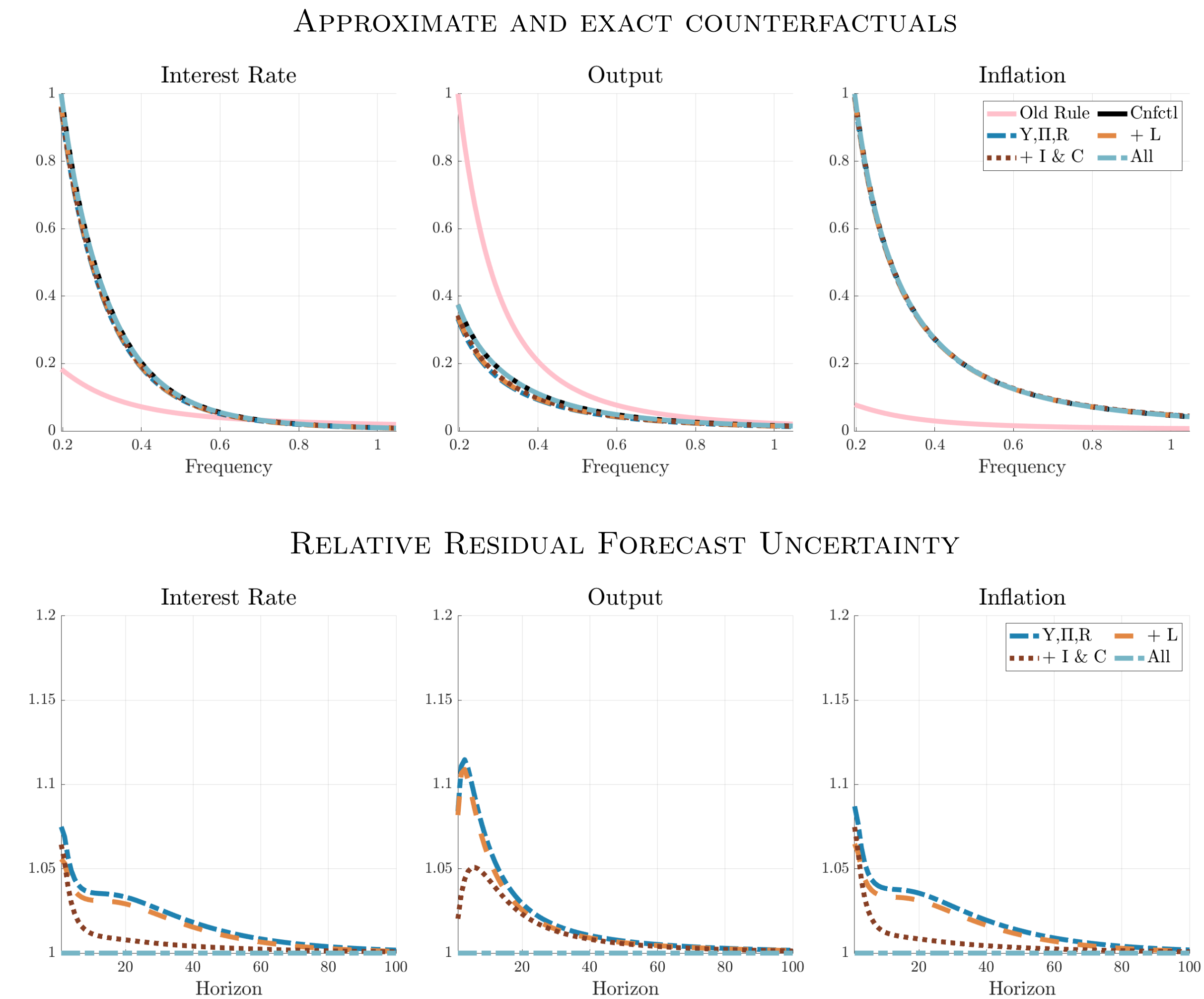

Figure A.1: Top panel: business-cycle spectral densities for interest rates, output, and inflation under the old rule (solid pink) and under the counterfactual rule, true (solid dark blue) and predicted using Proposition 2 for different information sets (solid-dashed blue, dashed orange, dotted red, solid-dashed cyan). Bottom panel: residual forecast variances for the same variables and for the same information sets, as a function of the forecast horizon (x-axis), and relative to the forecast variance for the full information set.

For that panel, we have normalized the residual forecast uncertainty under the full information set to 1 at each horizon h. As $h \to \infty$, the residual forecast variances for all information sets of course limit to the same number—the unconditional variance. For intermediate h, forecasting uncertainty is strictly larger for smaller information sets. We see, however, that the differences are moderate, with forecast variances that are only at most around 10 per cent larger than with the full information set. If instead those differences had been bigger, then the approximations in the top panel also would have been less accurate.

Discussion. The main practical takeaway from the previous discussion is that, when implementing our proposed “VAR-Plus” methodology, researchers should ensure that their conclusions are robust to expansions of the underlying information set y_t. We furthermore note that the invertibility assumption is in fact testable in the case of at least partially observable structural macroeconomic shocks. In particular, the testable implication of invertibility is that such shocks should not Granger-cause the included macroeconomic observables (see the discussion in Plagborg-Møller and Wolf, 2022).

B Further details for our method

This appendix provides supplementary details for our proposed methodology. Appendix B.1 discusses estimation of reduced-form projections, Appendix B.2 elaborates on how to obtain the (asymptotic) distribution of the policy shock targets $\hat{\theta}_\nu$, and Appendix B.3 gives further details on the model estimation step. Finally, in Appendix B.4, we provide an extended discussion of the desirable (robustness) properties of our approach.

B.1 Reduced-form projections

Our approach begins with estimation of the reduced-form VAR in (41). A textbook discussion of how to estimate reduced-form VARs and translate the autoregressive lag polynomial $A(L)$ into the implied Wold lag polynomial $\Psi(L)$ is provided, for example, in Kilian and Lütkepohl (2017). Appendix D.1 discusses the concrete implementation in our application, including details on data, lag length selection, and variable transformations.

B.2 Impulse response target estimation

Consistent with the recommendations of Plagborg-Møller and Wolf (2021) and Li et al. (2023), the researcher first uses a structural VAR to estimate the effects of policy shocks, identified using one (or several) of the semi-structural time series identification approaches. We propose to estimate this VAR using standard Bayesian techniques, delivering draws $i = 1, 2, \ldots, N$ of the policy shock causal effect vector $\hat{\theta}_{i,\nu}$.

We obtain $\hat{\theta}_\nu$ as the posterior mode of the estimated policy shock causal effects. For V_{θ_ν} we proceed as follows. We construct

$$\bar{V}_{\theta_\nu} \equiv \sum_{i=1}^{N} \left(\hat{\theta}_{i,\nu} - \hat{\theta}_\nu\right) \left(\hat{\theta}_{i,\nu} - \hat{\theta}_\nu\right)'.$$

Since the small-sample properties of estimating V_{θ_ν} in this way are poor, we instead work with a sample size-dependent transformation of $\bar{V}_{\theta_\nu}$, following Christiano et al. (2010):

$$V_{\theta_\nu} = f(\bar{V}_{\theta_\nu}, T)$$

where T is the sample size. The transformation $f(\bullet)$ has the following properties. First, V_{θ_ν} and $\bar{V}_{\theta_\nu}$ have the same diagonal entries. Second, for off-diagonal entries that correspond to the ℓth and jth lagged response of a common variable to a common shock, it scales down the entry of $\bar{V}_{\theta_\nu}$ by

$$\left(1 - \frac{|\ell - j|}{\bar{H}_T}\right)^{\eta_{1,T}}, \quad \ell, j = 1, 2, \ldots, \bar{H}_T \tag{B.1}$$

where $\bar{H}_T \leq H$ and $\bar{H}_T \to H$, $\eta_{1,T} \to 0$ as $T \to \infty$. Third, for all other off-diagonal entries corresponding ℓth and jth lagged responses, it scales down the entry of $\bar{V}_{\theta_\nu}$ by

$$\zeta_T \left(1 - \frac{|\ell - j|}{\bar{H}_T}\right)^{\eta_{2,T}}, \quad \ell, j = 1, 2, \ldots, \bar{H}_T \tag{B.2}$$

where $\zeta_T \to 1$ and $\eta_{2,T} \to 0$ as $T \to \infty$. Intuitively, this transformation dampens some (off-diagonal) elements in $\bar{V}_{\theta_\nu}$, with the dampening factor removed as the sample size increases. Finally, all covariances that are further apart than $\bar{H}_T$ periods are set to zero. One popular approach—followed, for example, in Christiano et al. (2005)—is to set $\eta_{1,T} = \infty$ and $\zeta_T = 0$ (thus $\eta_{2,T}$ and $\bar{H}_T$ are immaterial), so that V_{θ_ν} is simply a diagonal matrix composed of the diagonal components of $\bar{V}_{\theta_\nu}$. The opposite extreme is to not dampen at all, setting $V_{\theta_\nu} = \bar{V}_{\theta_\nu}$.

In our applications we will follow an intermediate strategy. We set $\zeta_T = 1$ in order to treat autocorrelations and correlations across different variables equally; we furthermore use a triangular kernel, so $\eta_{1,T} = \eta_{2,T} = 1$, and a bandwidth of $\bar{H}_T = 8$.[28] We depart from the standard diagonal weighting matrix because of the model selection step: using a diagonal matrix would lead to artificially sharp model selection, since small differences in fit of different models will lead to starkly different posterior odds. Accounting for the correlation patterns present in the IRF estimates reflects the informativeness of the data more accurately.

B.3 Model estimation

We here provide further implementation details for the model estimation step. We proceed in two steps. First, for a given model $\mathcal{M}_j$ and parameter vector ψ_j, we explain how to obtain $\theta_\nu(\psi_j, \mathcal{M}_j)$. This step is non-standard; in particular, we explain why we need not specify a policy rule to do so. Second, we discuss how we draw from the posterior and estimate the marginal likelihood. That step is entirely standard, so we will be brief.

Obtaining $\theta_\nu(\psi_j, \mathcal{M}_j)$. In order to evaluate the likelihood, we first need to obtain $\theta_\nu(\psi_j, \mathcal{M}_j)$. We obtain it in the following way:

[28] We note that our results are robust to different choices of bandwidth or to the use of other kernels.

1. Given $(\psi_j, \mathcal{M}_j)$, solve for the impulse responses of the targeted outcome variables with respect to policy news shocks for all horizons, $\boldsymbol{\nu}$. To do so we close the model with some determinacy-inducing policy rule; as discussed in Footnote 11, the choice of baseline rule is immaterial. Denote the (truncated) impulse response function matrices of interest as $\Theta_{\boldsymbol{x}_m,\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)$ for variable $\boldsymbol{x}_m$. Stack all of those impulse response matrices vertically in the same order as for $\hat{\theta}_\nu$, and denote the stacked matrix as $\Theta_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)$. This is a $(n_m T) \times T$ matrix, where T is the truncation horizon.[29]

2. We then, for each of the n_ν empirically identified policy shocks, find the unique *vector* of policy shocks in the model that matches the empirical impulse response targets as well as possible. Formally, for each empirical target shock $n = 1, \dots, n_\nu$, define a $T \times 1$ vector of news shocks $\tilde{\boldsymbol{\nu}}_n$. Vertically stack all these vectors of policy news shocks in the $(n_\nu T) \times 1$ vector $\tilde{\nu} = [\tilde{\boldsymbol{\nu}}'_1, \dots, \tilde{\boldsymbol{\nu}}'_{n_\nu}]'$. Define also for convenience the following $(n_m T) \times (n_\nu T)$ matrix: $\boldsymbol{\Phi}(\psi_j, \mathcal{M}_j) = I_{n_\nu} \otimes \Theta_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)$ where I_{n_ν} is an n_ν-dimensional identity matrix. We then obtain the best-fit vector of news shocks $\tilde{\nu}^*$ as

$$
\begin{aligned}
\tilde{\nu}^*(\psi_j, \mathcal{M}_j) \quad &= \quad \underset{\tilde{\nu}}{\operatorname{argmax}} \ \ \tilde{p}(\hat{\theta}_\nu, \tilde{\theta}_\nu, V_{\theta_\nu}) \\
\text{s.t} \qquad & \qquad \tilde{\nu}_{H+1:T,n} = 0 \quad \text{for all } 1, \dots, n_\nu \\
& \qquad \tilde{\theta}_\nu = \boldsymbol{\Phi}(\psi_j, \mathcal{M}_j)\tilde{\nu}
\end{aligned}
$$

where $\tilde{p}(\hat{\theta}_\nu, V_{\theta_\nu}, \tilde{\theta}_\nu)$ is the assumed density for "data" $\hat{\theta}_\nu$ with mean $\tilde{\theta}_\nu$ and covariance matrix V_{θ_ν}, and $\tilde{\nu}_{H+1:T,n}$ denotes elements $H+1, H+2, \dots, T$ of vector $\tilde{\boldsymbol{\nu}}_n$.[30] Given that f is assumed to be the density of a multivariate normal and V_{θ_ν} is taken as given, the maximizer $\tilde{\nu}^*(\psi_j, \mathcal{M}_j)$ can be found in closed form (since the maximization problem is a simple restricted linear quadratic problem).

In our empirical applications, we use only one identified monetary policy shock. To gain further intuition it is instructive to analyze this one-shock case in more detail. Take the top left $H \times H$ elements of each of the stacked impulse response matrices $\Theta_{\boldsymbol{x}_m,\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)$, stack them vertically, and denote the resulting matrix by $\Theta^H_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)$. Replacing the multivariate normal density, transforming appropriately, and focusing

[29] We set a truncation horizon of $T = 300$. Our results are insensitive to that choice.

[30] We impose this constraint to avoid overfitting: in order to match the IRF up to horizon H, we can only use the news shocks up to horizon H, and all other news shocks are set to zero.

only on the first H news shocks (since all others are set to zero) the problem to be solved can be written as:

$$\tilde{\nu}^*_{1:H}(\psi_j, \mathcal{M}_j) = \underset{\tilde{\nu}_{1:H}}{\text{argmax}} -\frac{1}{2}\left(\hat{\theta}_\nu - \Theta^H_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)\tilde{\nu}_{1:H}\right)' V^{-1}_{\theta_\nu}\left(\hat{\theta}_\nu - \Theta^H_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)\tilde{\nu}_{1:H}\right)$$

It is straightforward to show that the solution in this case is given by:

$$\tilde{\nu}^*_{1:H}(\psi_j, \mathcal{M}_j) = \left(\Theta^H_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)' V^{-1}_{\theta_\nu} \Theta^H_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)\right)^{-1}\left(\Theta^H_{\boldsymbol{\nu}}(\psi_j, \mathcal{M}_j)' V^{-1}_{\theta_\nu}\hat{\theta}_\nu\right)$$

In words, we can find the best-fitting shock vector $\tilde{\nu}^*_{1:H}$ through a "regression" of the empirical target impulse responses on the space of impulse response sequences implementable through policy.

3. With $\tilde{\nu}^*$ at hand, compute the model-implied impulse response functions as $\theta_\nu(\psi_j, \mathcal{M}_j) = \boldsymbol{\Phi}(\psi_j, \mathcal{M}_j)\tilde{\nu}^*$.

We note that this way of constructing the model-implied impulse responses $\theta_\nu(\psi_j, \mathcal{M}_j)$ differs from the standard approach of first (i) specifying a policy rule and then (ii) assuming that the identified policy shock corresponds to a time-0 shock under that rule (e.g., as in Christiano et al., 2005). For this approach to be valid, the assumed rule has to be correctly specified. In contrast, our approach does not require assumptions about the policy rule—we simply construct a sequence of contemporaneous and news policy shocks $\tilde{\nu}^*$ that perturbs the expected path of the policy instrument analogously to the empirically estimated policy instrument impulse response.[31]

Posterior distribution & marginal likelihood. We use a standard Random Walk Metropolis Hastings algorithm, with a multivariate normal for the proposal distribution. The variance-covariance matrix is initially assumed to be equal to the prior variance-covariance matrix, scaled by a constant c_1^2.[32] We use the first N_a draws to estimate the variance-covariance matrix of the proposal distribution, updating the proposal variance-covariance matrix to the observed variance-covariance matrix of parameters in the first N_a draws (scaled

[31] This claim works exactly in population as $T, H \to \infty$. However, due to the finite horizon of the impulse-response matching, the baseline assumed rule may matter due to truncation. In the models we consider, the matched impulse-response and inferred structural parameters are almost exactly the same under a variety of parameters for the assumed determinacy-inducing rule.

[32] For our HANK models, we in this step use a standard deviation of 0.1 for the informational stickiness parameter (instead of 0.2), to avoid getting too many draws outside of the parameter support.

by c_2^2). Once updated, we sample another $N_b + N_c$ draws, burn the first N_b and keep the last N_c draws, which we use as out posterior distribution. We set $N_a = N_c = 100000$, $N_b = 50000$, $c_1 = 0.8$ and $c_2 = 0.7$ for all models. Our acceptance rates for all models considered range between 20 and 30 percent.

In order to then implement our applications, we need to store the impulse response matrices of the outcomes of interest with respect to the full sequence of news shocks. Given that storing hundreds of thousands of draws of $T \times T$ matrices is very expensive in terms of memory, we store only the $T_u \times T_u$ top left elements, for only a number of N_d draws. We set $T_u = 200$ and $N_d = 1000$. Specifically, we store one draw out of each $N_c/N_d = 100$, to get draws that are closer to uncorrelated. Finally, given those posterior draws, we estimate the marginal likelihood using the harmonic mean estimator of Geweke (1999).[33]

B.4 Properties of our method

We here provide some further details supplementing the discussion in Section 3.4, elaborating on advantages of our approach relative to the "medium-scale DSGE" strategy.

Robustness to model mis-specification. Under standard full-information approaches to model estimation (like Smets and Wouters, 2007), mis-specification in one part of the model will affect inference for the other parts. The argument is straightforward, so our discussion here will be brief; we will furthermore focus our discussion on mis-specification in shock processes, as such mis-specification is particularly likely in practice (Chari et al., 2009). Analogous arguments apply to mis-specification in policy rules.

Suppose the true data-generating process is

$$y_t = \Theta^*(L)\xi_t \tag{B.3}$$

$$\xi_t = B^*(L)\varepsilon_t \tag{B.4}$$

where $\varepsilon_t \sim N(0, I)$. Relative to (17), the system (B.3) - (B.4) is written to explicitly separate the exogenous process (i.e., equation (B.4)) from the endogenous model propagation (i.e., equation (B.3)). For example, ε_t could be an innovation to total factor productivity, while ξ_t is the exogenous TFP level itself. For future reference we define $\Psi^*(L) = \Theta^*(L)B^*(L)$.

[33] We set the truncation parameter such that we use only half of the sample. We use the full sample consisting of N_c draws to estimate the marginal likelihoods.

The researcher instead entertains models indexed by parameters $\psi = (\psi_1', \psi_2')'$:

$$y_t = \Theta_{\psi_1}(L)\xi_t \tag{B.5}$$
$$\xi_t = B_{\psi_2}(L)\varepsilon_t \tag{B.6}$$

where again $\varepsilon_t \sim N(0, I)$. We assume that there is no mis-specification in the endogenous propagation part of the model: there is a (in fact unique) ψ_1^* such that $\Theta_{\psi_1^*}(L) = \Theta^*(L)$. Shock propagation, however, is mis-specified; for example, the researcher may assume that all shocks follow AR(1) processes, while in fact they follow richer ARMA(p,q) processes. For future reference we again write $\Psi_\psi(L) = \Theta_{\psi_1}(L)B_{\psi_2}(L)$.

Finally, to make our arguments as stark as possible, we suppose that there exists a unique $\psi^\dagger$ such that

$$\Psi^*(e^{-i\omega}) = \Psi_{\psi^\dagger}(e^{-i\omega}) \qquad \forall \omega \in [0, 2\pi].$$

Thus, when evaluated at $\psi^\dagger$ (and only then), the two processes (B.3) - (B.4) and (B.5) - (B.6) imply the exact same second moments, so conventional likelihood-based estimation will asymptotically yield $\psi = \psi^\dagger$. But since $B^*(L) \neq B_{\psi_2^\dagger}(L)$, we will generically have $\Theta^*(L) \neq \Theta_{\psi_1^\dagger}(L)$—i.e., mis-specification in the endogenous shock propagation part, including in particular the policy space Θ_ν. Since our approach does not require the researcher to take any stance on the shock process part $B(L)$, it is by design robust to such concerns.

Weak identification. Standard full-information approaches to estimation of DSGE models are often subject to concerns of weak identification (e.g., see Fernández-Villaverde et al., 2016). Our approach is arguably less subject to this concern, simply because it only requires the researcher to partially specify the model, thus reducing the number of parameters that need to be identified. We here provide a simple example illustration of this insight.

Consider the following two-variable, two-equation static model:

$$\begin{aligned} y_t &= -\frac{1}{\gamma} i_t + \sigma_d \varepsilon_t^d, \\ i_t &= \phi_y y_t + \sigma_m \varepsilon_t^m, \end{aligned}$$

where y_t and i_t denote outcome variables (output and interest rates), and $(\varepsilon_t^d, \varepsilon_t^m)$ are shocks.

Note that the solution is given as

$$\begin{pmatrix} y_t \\ i_t \end{pmatrix} = \underbrace{\frac{1}{1+\frac{\phi_y}{\gamma}} \begin{pmatrix} -\frac{1}{\gamma}\sigma_m & \sigma_d \\ \sigma_m & \phi_y \sigma_d \end{pmatrix}}_{\equiv \Theta} \begin{pmatrix} \varepsilon_t^m \\ \varepsilon_t^d \end{pmatrix}$$

Consider first a researcher following our approach. The ratio of the impulse responses of interest rates and output to a monetary policy shock ε_t^m point-identifies γ, and so the space of output and interest rate allocations implementable through policy, as required by our identification result. Now consider instead identification based on second moments; i.e., we seek to find a tuple $\{\gamma, \phi_y, \sigma_d, \sigma_m\}$ such that

$$\Sigma = \Theta(\gamma, \phi_y, \sigma_d, \sigma_m)\Theta(\gamma, \phi_y, \sigma_d, \sigma_m)'$$

where $\Sigma \equiv \Theta\Theta'$ is the true variance-covariance matrix. It is straightforward to verify that these moment conditions are insufficient to point-identify the model, and in particular do not point-identify γ.[34]

Inability to match second moments. Recall that, by our identification result, the researcher needs to know (i) second moments of the observed aggregate time series data y_t and (ii) policy causal effects Θ_ν. Our discussion of the standard "medium-scale DSGE" approach so far has focused on sources of mis-specification for (ii), even if second moments (i) are matched perfectly. Prior work, however, also suggests that even rich estimated structural DSGE models may struggle to deliver (i); for example, the "business-cycle anatomy" of Angeletos et al. (2020) uncovers important disagreement between observed and model-implied autocovariance functions of macroeconomic aggregates. Our approach sidesteps such concerns simply because it takes the second moments directly from the data.

Efficiency losses. While strictly more robust to model mis-specification, our approach will of course be less efficient than full-information likelihood-based method *if the entirety*

[34] To see this, start with some arbitrary $\gamma > 0$. Note that

$$\frac{\text{Var}(i_t) + \gamma \,\text{Cov}(y_t, i_t)}{\text{Var}(y_t) + \frac{1}{\gamma}\text{Cov}(y_t, i_t)} = \frac{\phi_y^2 + \phi_y \gamma}{1 + \frac{\phi_y}{\gamma}}$$

Solve this equation for ϕ_y, recover σ_d from $\text{Var}(i_t) + \gamma \,\text{Cov}(y_t, i_t)$, and finally get σ_m from $\text{Var}(i_t)$. The resulting parameter vector leads the model to correctly match the desired Σ.

of the model is well-specified. That being said, we note that prior work has generally found the efficiency gains associated with (correctly specified) full-information maximum-likelihood approaches to DSGE estimation to be rather limited (e.g., see Ruge-Murcia, 2007).

C Monetary policy transmission extrapolation

This appendix complements our discussion in Section 4 on model estimation and monetary policy causal effect extrapolation. First, in Appendix C.1, we present additional details on our empirical monetary policy shock estimation. Second, in Appendix C.2, we provide the detailed equations for our list of models $\mathcal{M}$. Third, in Appendix C.3, we present supplementary results for our model estimation exercise.

C.1 Empirical evidence on monetary shock propagation

We here provide further details on how we construct our monetary shock estimation targets in Section 4.1. We elaborate on data construction and econometric implementation.

Data. We are interested in impulse responses of three outcome variables: the output gap, inflation, and the policy rate. For the output gap, we take log real output per capita (obtained from FRED, series `A939RX0Q048SBEA`), and then de-trend following Hamilton (2018). Next, for inflation, we log-difference the GDP deflator (obtained from FRED, series `GDPDEF`), and then annualize. Finally, we consider the federal funds rate as our measure of the policy rate, also obtained from FRED (`FEDFUNDS`) and annualized. All series are quarterly, and we consider a sample from 1969:Q1—2006:Q4.

Our measure of a monetary shock series is obtained from the replication files of Aruoba and Drechsel (2022). We aggregate by averaging the monthly series, and we set all missing values of this monetary shock IV to zero, as in Känzig (2021).

Econometric implementation. We estimate a VAR in the shock series together with our three outcome variables of interest, consistent with the recommendations of Li et al. (2023). As in Plagborg-Møller and Wolf (2021), we order the shock series first in a recursive identification of our VAR, delivering invertibility-robust estimates of the causal effects of the monetary shock. We include two lags, a linear time trend, and use a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization (Arias et al., 2018). Our estimation results are robust to these particular choices. This procedure yields draws of the policy shock causal effect vector $\hat{\theta}_\nu$, which are then used to construct V_{θ_ν} following the steps outlined in Appendix B.2.

C.2 Models

This section provides some supplementary details for our structural models of monetary policy transmission sketched in Section 4. We list all model equations; however, since the models are relatively standard, the derivations will be rather brief. Throughout this section, we use tildes to denote log-deviations from steady state.

C.2.1 Baseline RANK

Households & unions. Households choose sequences of consumption c_t and assets a_t^H to maximize (48) subject to (49). The Euler Equation in log-deviations from steady state is:

$$\tilde{\lambda}_t = \mathbb{E}_t[\tilde{r}_{t+1} + \tilde{\lambda}_{t+1}]$$

with $\tilde{r}_{t+1} = \tilde{r}_t^n - \pi_{t+1}$, $\frac{P_{t+1}}{P_t} = \exp(\pi_{t+1})$, and

$$\tilde{\lambda}_t = -\frac{1}{(1-\beta h)(1-h)}\gamma(\tilde{c}_t - h\tilde{c}_{t-1}) + \frac{1}{(1-\beta h)(1-h)}\beta h\gamma(\mathbb{E}_t\left[\tilde{c}_{t+1}\right] - h\tilde{c}_t).$$

A detailed derivation of the wage Phillips curve (50) is deferred until Appendix C.2.3, given that the full information case is nested in the derivation that includes cognitive discounting.

Production and pricing. The production function for an intermediate good producer i is:

$$Y_t(i) = \bar{A}(u_t(i)k_{t-1}(i))^{\alpha}(\ell_t(i))^{1-\alpha}$$

where $\bar{A}$ denotes aggregate productivity, $k_{t-1}(i)$ is capital stock of firm i, $u_t(i)$ is capacity utilization, and $\ell_t(i)$ denotes labor hired. All intermediate good producers are symmetric and so we drop the i subscript. Capital is purchased one period in advance. The intermediate good producer solves:[35]

$$\max_{\ell_t,k_t,u_t} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(\Pi_{j=0}^{t}(1+r_j))^{-1}[p_t^I Y_t - w_t\ell_t - a(u_t) - q_t(k_t - (1-\delta)k_{t-1})\right]$$

[35]We discount future pay-offs using the real rate of interest. Up to first order, this is equivalent to using the representative household's implied stochastic discount factor.

where $a(u_t)$ is an utility cost of adjusting capacity, and $q_t k_t$ is the total cost of capital purchases for next period.[36] The first-order conditions are:

$$\begin{aligned} w_t &= p_t^I(1-\alpha)\bar{A}\left(\frac{\ell_t}{u_t k_{t-1}}\right)^{-\alpha} \\ a'(u_t) &= p_t^I \alpha \bar{A}\left(\frac{\ell_t}{u_t k_{t-1}}\right)^{1-\alpha} \\ q_t &= \mathbb{E}_t\left(\frac{1}{1+r_{t+1}}\left[p_{t+1}^I \alpha \bar{A}\left(\frac{\ell_{t+1}}{u_{t+1}k_t}\right)^{1-\alpha} + (1-\delta)q_{t+1}\right]\right) \end{aligned}$$

Log-linearizing around the steady state:

$$\begin{aligned} \tilde{y}_t &= \alpha(\tilde{u}_t + \tilde{k}_{t-1}) + (1-\alpha)\tilde{\ell}_t \\ \tilde{w}_t &= \tilde{p}_t^I + \alpha(\tilde{u}_t + \tilde{k}_{t-1}) - \alpha\tilde{\ell}_t \\ \zeta\tilde{u}_t &= \tilde{p}_t^I + (\alpha-1)(\tilde{u}_t + \tilde{k}_{t-1}) + (1-\alpha)\tilde{\ell}_t \\ \tilde{q}_t &= \mathbb{E}_t\left[-\tilde{r}_{t+1} + \left(1 - \frac{1-\delta}{1+\bar{r}}\right)(\tilde{p}_{t+1}^I + (\alpha-1)(\tilde{k}_t + \tilde{u}_{t+1}) + (1-\alpha)\tilde{\ell}_{t+1}) + \frac{1-\delta}{1+\bar{r}}\tilde{q}_{t+1}\right] \end{aligned}$$

where $\zeta = a''(1)/a'(1)$ is the curvature parameter of the capacity utilization cost function. Following Smets and Wouters (2007), we parametrize $\zeta = \frac{\psi}{1-\psi}$ and then use the same prior on ψ as in that paper.

Retail firms solve their dynamic pricing problem subject to Calvo frictions. The resulting Phillips curve is (51); detailed derivations are deferred until Appendix C.2.3.

Capital good producers solve

$$\max_{i_t} \mathbb{E}_0\left[\sum_{t=0}^{\infty}(\Pi_{j=0}^t(1+r_j))^{-1}\left(q_t i_t - S\left(\frac{i_t}{i_{t-1}}\right)\right)\right],$$

where i_t is the production of new capital goods (sold to the intermediate goods producers), and $S(x)$ is the adjustment cost function. The first-order condition is given by:

$$q_t = \frac{1}{i_{t-1}}S'\left(\frac{i_t}{i_{t-1}}\right) - \mathbb{E}_t\left[\frac{1}{1+r_{t+1}}S'\left(\frac{i_{t+1}}{i_t}\right)\frac{i_{t+1}}{i_t^2}\right]$$

We assume that $S(1) = S'(1) = 0$ and $\kappa = S''(1) > 0$. Log-linearizing around the steady

[36] The cost is written in terms of utility, so it does not enter the market-clearing condition.

state yields:

$$q_t = \kappa(\tilde{i}_t - \tilde{i}_{t-1}) - \frac{\kappa}{1+\bar{r}}(\tilde{i}_{t+1} - \tilde{i}_t)$$

Finally, capital evolves according to $k_t = (1-\delta)k_{t-1} + i_t$ or in log-linearized terms:

$$\tilde{k}_t = (1-\delta)\tilde{k}_{t-1} + \delta\tilde{i}_t$$

We note that therefore goods market-clearing implies that, to first order:

$$\tilde{y}_t = \bar{c}\tilde{c}_t + \bar{i}\tilde{i}_t$$

Policy. The government budget constraint is

$$w_t \ell_t \tau_t^\ell + b_t = (1+r_t)b_{t-1} + \tau_t + g_t$$

where r_t is the real return on government debt b_t, τ_t denotes lump-sum transfers, τ_t^ℓ denotes distortionary labor taxes, and g_t denotes government expenditure. Log-linearizing:

$$\bar{w}\bar{\ell}\bar{\tau}^\ell(\tilde{w}_t + \tilde{\ell}_t + \tilde{\tau}_t^\ell) + \bar{b}\tilde{b}_t = (1+\bar{r})\bar{b}(\tilde{b}_{t-1} + \tilde{r}_t) + \bar{\tau}\tilde{\tau}_t + \bar{g}\tilde{g}_t$$

Second, the realized real return on government debt satisfies

$$1 + r_t = \frac{\bar{r}+\eta}{\exp(\pi_t)}\frac{1}{p_{t-1}} + \frac{1-\eta}{\exp(\pi_t)}\frac{p_t}{p_{t-1}}$$

where p_t is the real relative price of government debt and η is the decay rate of the coupon, with $\eta = 0$ corresponding to perpetuities and $\eta = 1$ corresponding to one-period debt. Log-linearizing:

$$\tilde{r}_t = -\pi_t - \tilde{p}_{t-1} + \frac{1-\delta}{1+\bar{r}}\tilde{p}_t$$

The central bank sets the nominal rate on one-period government debt, which is in zero net supply. By perfect foresight arbitrage we have

$$1 + r_t = \frac{1 + r_{t-1}^n}{\exp(\pi_t)}, \quad t = 1, 2, \ldots$$

and so, in log-deviations

$$\tilde{r}_t = \tilde{r}_{t-1}^n - \pi_t, \quad t = 1, 2, \ldots$$

or

$$\tilde{r}_t^n = -\tilde{p}_t + \frac{1-\eta}{1+\bar{r}}\mathbb{E}_t\left[\tilde{p}_{t+1}\right]$$

It remains to determine how taxes are set. We assume:

$$\begin{aligned}\tilde{\tau}_t &= \tilde{g}_t = 0 \\ \bar{w}\bar{\ell}\tilde{\tau}_t^\ell &= \bar{b}\tau_b^\ell\tilde{b}_{t-1}\end{aligned}$$

That is, all the adjustment is done via distortionary taxes. The resulting law of motion for government debt is

$$\tilde{b}_t = (1+\bar{r}-\bar{\tau}^\ell\tau_b^\ell)\tilde{b}_{t-1} + (1+\bar{r})\tilde{r}_t.$$

Policy rule for computation. For our numerical analysis, we close the model with a determinacy-inducing Taylor rule, as discussed in Appendix B.3:

$$\tilde{r}_t^n = (1-\rho)\left(\rho\tilde{r}_{t-1}^n + \phi_\pi\pi_t + \phi_y\tilde{y}_t + \phi_{\Delta y}(\tilde{y}_t - \tilde{y}_{t-1})\right)$$

As emphasized throughout, our model estimation step and policy counterfactual applications do not depend on this choice of basis rule, simply because we allow for arbitrarily general policy shocks, allowing us to implement arbitrary paths of interest rates. For Figures 2 and 3, we subject this rule to ten-quarter-ahead forward guidance shocks.

Steady state. We normalize the level of disutility of labor such that $\bar{\ell} = 1$. Given that assumption, the Euler equation pins down the real rate as $1+\bar{r} = \beta^{-1}$. We can then find $\bar{k}$, which immediately yields $\bar{i}$, $\bar{y}$ and $\bar{w}$. We calibrate the level of outstanding government debt, labor taxes and transfers (see Appendix C.3), and pick the steady state level of government consumption such that the intertemporal government budget constraint holds.

C.2.2 Baseline HANK

The only two differences relative to the baseline RANK model are that: (i) we replace the representative agent with a heterogeneous agents block, as already described in the main text; (ii) we now need to specify how dividends are paid to the households.

Household and unions. The details of household block is described in the main text. In order to compute the solution with informational rigidities, we follow Auclert et al. (2020):

we first solve for the Jacobians of the household block under full information, and then transform them to obtain the solution under sticky information.

Dividend distribution. Households receive dividends through a financial intermediary. Let a_t^I denote total assets held by the financial intermediary. Those assets evolve as

$$a_t^I = (1 + r_t)a_{t-1}^I + (d_t - d_t^H)$$

where d_t denotes dividends paid by firms to the intermediary and d_t^H denotes payments from the intermediary to the households. We assume the following distribution rule:

$$(d_t^H - \bar{d}) = \delta_1(d_t - \bar{d}) + \delta_2(1 + r_t)a_{t-1}^I$$

Note that $\delta_1 = 1$ corresponds to the usual case of dividends paid out straight to households, with $a_t^I = 0$ always. The linearized relations are

$$\hat{a}_t^I = (1 - \delta_2)(1 + \bar{r})\hat{a}_{t-1}^I + (1 - \delta_1)\bar{d}\tilde{d}_t$$

and

$$\bar{d}\tilde{d}_t^H = \delta_1 \bar{d}\tilde{d}_t + \delta_2(1 + \bar{r})\tilde{a}_{t-1}^I$$

where $\hat{x} = x - \bar{x}$. We linearize (instead of log-linearizing) with respect to a_t^I since $\bar{a}^I = 0$.

Steady state. We proceed exactly as in the RANK case. Given a calibrated real interest rate, we pick β such that in equilibrium households want to hold the calibrated level of liquid assets, which are given by the outstanding stock of government debt. Apart from the value of β, the steady state is exactly the same as in the RANK case.

C.2.3 Adding cognitive discounting

This subsection derives the price- and wage-NKPCs under cognitive discounting and price indexation. We derive the NPKC under partial indexation and cognitive discounting, where ζ is the degree of price indexation; $\zeta = 1$ corresponds to the case considered in the text.

Pricing. The problem of a retailer is to choose P_t^* to maximize

$$\mathbb{E}_t \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p)^{\tau-t} M_{\tau|t} \left(\frac{P_{\tau|t}}{P_\tau} - \mu_\tau \right) \left(\frac{P_{\tau|t}}{P_\tau} \right)^{-\epsilon_p} Y_\tau,$$

where $\bar{\beta} = \frac{1}{1+\bar{r}}$, $P_{\tau|t}$ is the price at date τ of a firm that last updated its price at t, μ_τ is the real marginal cost of producing at τ, P_τ is the aggregate price index, Y_τ is aggregate demand, $M_{\tau|t} = u_c(c_\tau)/u_c(c_t)$, and $1 - \theta_p$ is the probability of resetting the price. Due to price indexation, we have

$$P_{\tau|t} = P_t^* \underbrace{\exp\left(\zeta(\pi_t + \pi_{t+1} + \cdots \pi_{\tau-1})\right)}_{\equiv I_{\tau|t}}.$$

The first-order condition of the price-setting problem is

$$(\epsilon_p - 1)\mathbb{E}_t \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p)^{\tau-t} M_{\tau|t} \left(\frac{P_{\tau|t}}{P_\tau} \right)^{-\epsilon_p} Y_\tau \frac{I_{\tau|t}}{P_\tau} = \epsilon_p \mathbb{E}_t \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p)^{\tau-t} M_{\tau|t}\mu_\tau \left(\frac{P_{\tau|t}}{P_\tau} \right)^{-\epsilon_p - 1} Y_\tau \frac{I_{\tau|t}}{P_\tau}.$$

Log-linearizing both sides of this equation around a zero-inflation steady state we have

$$\mathbb{E}_t \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p)^{\tau-t} \left[\tilde{\mu}_\tau - \tilde{P}_{\tau|t} + \tilde{P}_\tau \right] = 0$$

or

$$\mathbb{E}_t \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p)^{\tau-t} \left[\tilde{\mu}_\tau - \tilde{P}_t^* - \sum_{s=t+1}^{\tau} \zeta\pi_{s-1} + \tilde{P}_\tau \right] = 0$$

and so

$$\tilde{P}_t^* - \tilde{P}_t = (1 - \bar{\beta}\theta_p)\mathbb{E}_t \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p)^{\tau-t} \left[\tilde{\mu}_\tau - \sum_{s=t+1}^{\tau} \zeta\pi_{s-1} + \tilde{P}_\tau - \tilde{P}_t \right]$$

or

$$\tilde{P}_t^* - \tilde{P}_t = (1 - \bar{\beta}\theta_p)\mathbb{E}_t \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p)^{\tau-t} \left[\tilde{\mu}_\tau + \sum_{s=t+1}^{\tau} (1 - \zeta L)\pi_s \right]$$

where L is the lag operator. We now apply cognitive discounting (as in Gabaix, 2020):

$$\tilde{P}_t^* - \tilde{P}_t = (1 - \bar{\beta}\theta_p) \sum_{\tau \geq t}^{\infty} (\bar{\beta}\theta_p m)^{\tau-t} \left[\tilde{\mu}_\tau + \mathbb{E}_t \sum_{s=t+1}^{\tau} (1 - \zeta L)\pi_s \right] \tag{C.1}$$

where m is the cognitive discount factor.

The aggregate price index evolves as

$$P_t = \left[\theta_p(P_{t-1}(\exp(\zeta\pi_{t-1})))^{1-\varepsilon} + (1-\theta_p)(P_t^*)^{1-\varepsilon}\right]^{1/(1-\varepsilon)}$$

Solving this for P_t^*:

$$P_t^* = \left[\frac{P_t^{1-\varepsilon} - \theta_p(P_{t-1}(\exp(\zeta\pi_{t-1}))^{1-\varepsilon}}{1-\theta_p}\right]^{1/(1-\varepsilon)}$$

Dividing by P_t:

$$\frac{P_t^*}{P_t} = \left[\frac{1-\theta_p(\exp(\pi_t))^{\varepsilon-1}(\exp(\zeta\pi_{t-1}))^{1-\varepsilon}}{1-\theta_p}\right]^{1/(1-\varepsilon)}$$

Re-arranging:

$$(1-\theta_p)\left(\frac{P_t^*}{P_t}\right)^{1-\varepsilon} = 1-\theta_p(\exp(\pi_t))^{\varepsilon-1}(\exp(\zeta\pi_{t-1}))^{1-\varepsilon}$$

Log-linearizing:

$$\begin{aligned} \pi_t &= \frac{1-\theta_p}{\theta_p}\left(\tilde{P}_t^* - \tilde{P}_t\right) + \zeta\pi_{t-1} \\ (1-\zeta L)\pi_t &= \frac{1-\theta_p}{\theta_p}\left(\tilde{P}_t^* - \tilde{P}_t\right) \end{aligned} \tag{C.2}$$

Combining (C.1) and (C.2) we arrive at

$$(1-\zeta L)\pi_t = \frac{(1-\theta_p)(1-\bar{\beta}\theta_p)}{\theta_p}\mathbb{E}_t\sum_{\tau\geq t}^{\infty}(\bar{\beta}\theta_p m)^{\tau-t}\left[\tilde{\mu}_\tau + \sum_{s=t+1}^{\tau}(1-\zeta L)\pi_s\right]. \tag{C.3}$$

Define $\breve{\pi}_t = (1-\zeta L)\pi_t$ as the quasi-differenced rate of inflation. We can then rewrite the preceding equation as

$$\breve{\pi}_t = \frac{(1-\theta_p)(1-\bar{\beta}\theta_p)}{\theta_p}\mathbb{E}_t\left[\sum_{\tau=t}^{\infty}(\bar{\beta}\theta_p m)^{\tau-t}\tilde{\mu}_\tau + \frac{1}{1-\bar{\beta}\theta_p m}\sum_{\tau=t+1}^{\infty}(\bar{\beta}\theta_p m)^{\tau-t}\breve{\pi}_\tau\right]$$

or

$$\breve{\pi}_t = \underbrace{\frac{(1-\theta_p)(1-\bar{\beta}\theta_p)}{\theta_p}}_{\kappa_p}\mathbb{E}_t\left[\sum_{\tau=t}^{\infty}(\bar{\beta}\theta_p m)^{\tau-t}\left(\tilde{\mu}_\tau + \frac{\breve{\pi}_\tau}{1-\bar{\beta}\theta_p m}\right) - \frac{\breve{\pi}_t}{1-\bar{\beta}\theta_p m}\right]$$

and so

$$\check{\pi}_t \left[1 + \frac{\kappa_p}{1 - \bar{\beta}\theta_p m}\right] = \kappa_p \left[\sum_{\tau=t}^{\infty} (\bar{\beta}\theta_p m)^{\tau-t} \left(\tilde{\mu}_\tau + \frac{\check{\pi}_\tau}{1 - \bar{\beta}\theta_p m}\right)\right]$$

Differencing forward and re-arranging:

$$\left[1 + \frac{\kappa_p}{1 - \bar{\beta}\theta_p m}\right] \left(\check{\pi}_t - \bar{\beta}\theta_p m \mathbb{E}_t \check{\pi}_{t+1}\right) = \kappa_p \left(\tilde{\mu}_t + \frac{\check{\pi}_t}{1 - \bar{\beta}\theta_p m}\right)$$

and so

$$\check{\pi}_t = \kappa_p \tilde{\mu}_t + \bar{\beta}\theta_p m \left[1 + \frac{\kappa_p}{1 - \bar{\beta}\theta_p m}\right] \mathbb{E}_t \check{\pi}_{t+1}$$

Replacing the definition of $\tilde{\pi}_t$ and noting that $\tilde{\mu}_t = \tilde{p}_t^I$ yields the expression in the main text.

Wage-setting. For tractability we assume that unions evaluate household utility at average consumption and hours worked (rather than averaging across individual household utilities), as in McKay and Wolf (2022). When a union does not update its wage, it adjusts it to $W_{j,t} = W_{j,t-1}(\exp(\zeta_w \pi_{t-1}))$, where π_t is price inflation. We will use the notation

$$W_{\tau|t} \equiv W_t^* \exp(\zeta_w(\pi_t + \cdots + \pi_{\tau-1})$$

for the nominal wage at date τ for a union that set its wage at date t. As before we derive everything allowing for partial indexation, with our analysis in the main text corresponding to the special case of full indexation ($\zeta_w = 1$). Real earnings for union j are

$$\frac{W_{\tau|t}}{P_\tau} \ell_{j\tau} = \left(\frac{W_{\tau|t}}{P_\tau}\right) \left(\frac{W_{\tau|t}}{W_\tau}\right)^{-\epsilon_w} L_\tau = \left(\frac{W_\tau}{P_\tau}\right) \left(\frac{W_{\tau|t}}{W_\tau}\right)^{1-\epsilon_w} L_\tau.$$

Note that $\ell_{j,t}$ denotes hours worked for union j, ℓ_τ is total hours worked by the households, and L_τ is the effective aggregate labor supply. Wage dispersion implies $L_\tau \leq \ell_\tau$; however, since we consider first-order approximations, we can proceed as if $L_\tau = \ell_\tau$.

The union's problem is to choose the nominal reset wage W_t^* to maximize

$$\mathbb{E}_t \sum_{\tau \geq t} \left(\bar{\beta}\theta_w\right)^{\tau-t} \left[\lambda_t \left(\frac{W_\tau}{P_\tau}\right) \left(\frac{W_{\tau|t}}{W_\tau}\right)^{1-\epsilon_w} - \nu_\ell\left(\ell_\tau\right) \left(\frac{W_{\tau|t}}{W_\tau}\right)^{-\epsilon_w}\right] L_\tau$$

where λ_t is the relevant aggregate marginal utility, and $\bar{\beta}$ is the time discount factor used by

the union, assumed to equal to the one used by the firm.[37]
The first-order condition is

$$\mathbb{E}_t \sum_{\tau \geq t} \left(\bar{\beta}\theta_w\right)^{\tau-t} \nu_\ell\left(\ell_\tau\right) \ell_\tau \epsilon_w W_\tau^{\epsilon_w} \prod_{s=t+1}^{\tau} \exp(\zeta_w \pi_{s-1})$$

$$= \mathbb{E}_t \sum_{\tau \geq t} \left(\bar{\beta}\theta_w\right)^{\tau-t} u_c(c_\tau)(\epsilon_w - 1) \frac{W_{\tau|t}}{P_\tau} W_\tau^{\epsilon_w} \ell_\tau \prod_{s=t+1}^{\tau} \exp(\zeta_w \pi_{s-1}).$$

Log-linearizing the first-order condition around a zero-inflation steady state:

$$\mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\theta_w)^{\tau-t} \left(\phi\tilde{\ell}_\tau - \tilde{W}_{\tau|t} + \tilde{p}_\tau - \tilde{\lambda}_\tau\right) = 0$$

or

$$\mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\theta_w)^{\tau-t} \left(\phi\tilde{\ell}_\tau - \tilde{W}_t^* - \sum_{s=t+1}^{\tau} \zeta_w \pi_{s-1} + \tilde{p}_\tau - \tilde{\lambda}_\tau\right) = 0,$$

where $\phi \equiv \frac{\nu_{\ell\ell}(\bar{\ell})\bar{\ell}}{\nu_\ell(\bar{\ell})}$. Re-arranging

$$\tilde{W}_t^* - \tilde{W}_t = (1 - \bar{\beta}\theta_w)\mathbb{E}_t \sum_{\tau \geq t} \left(\bar{\beta}\theta_w\right)^{\tau-t} \left(\phi\tilde{\ell}_\tau - \tilde{\lambda}_\tau - \sum_{s=t+1}^{\tau} \zeta_w \pi_{s-1} - \tilde{W}_t + \tilde{p}_\tau\right)$$

and so

$$\tilde{W}_t^* - \tilde{W}_t = (1 - \bar{\beta}\theta_w)\mathbb{E}_t \sum_{\tau \geq t} \left(\bar{\beta}\theta_w\right)^{\tau-t} \left(\phi\tilde{\ell}_\tau - \tilde{\lambda}_\tau + \sum_{s=t+1}^{\tau} (\pi_s^w - \zeta_w \pi_{s-1}) - \tilde{w}_\tau\right),$$

where $\tilde{w}_\tau \equiv \tilde{W}_\tau - \tilde{p}_\tau$. We will define $\chi_\tau = \varphi\tilde{\ell}_\tau - \tilde{\lambda}_\tau - \tilde{w}_\tau$ to be the labor wedge. Recall that, under our assumptions, we in the HANK model have that $\tilde{\lambda}_t = -\gamma\tilde{c}_t$ where $\tilde{c}_t$ is log-deviations of aggregate consumption.

From the definition of the wage index we have

$$\pi_t^w = \frac{1-\theta_w}{\theta_w}(\tilde{W}_t^* - \tilde{W}_t) + \zeta_w \pi_{t-1}.$$

[37] In the case of RANK, λ_t is as discussed in Appendix C.2.1, and the firm and union discount factors are always identical. In the case of HANK, we use the marginal utility evaluated at aggregate consumption (i.e., $\lambda_t = c_t^{-\gamma}$), as in McKay and Wolf (2022), and we just set the discount factor for unions equal to the one for firms to keep the models as comparable as possible.

Combining these relations we get

$$\pi_t^w - \zeta_w \pi_{t-1} = \frac{(1-\theta_w)(1-\beta\theta_w)}{\theta_w} \mathbb{E}_t \sum_{\tau \geq t} \left(\bar{\beta}\theta_w\right)^{\tau - t} \left(\chi_\tau + \sum_{s=t+1}^{\tau} (\pi_s^w - \zeta_w \pi_{s-1}) \right)$$

Applying cognitive discounting:

$$\pi_t^w - \zeta_w \pi_{t-1} = \frac{(1-\theta_w)(1-\bar{\beta}\theta_w)}{\theta_w} \mathbb{E}_t \sum_{\tau \geq t} \left(\bar{\beta}\theta_w m\right)^{\tau - t} \left(\chi_\tau + \sum_{s=t+1}^{\tau} (\pi_s^w - \zeta_w \pi_{s-1}) \right)$$

This expression has the same structure as (C.3). Operating exactly in the same way as before we obtain

$$\pi_t^w - \zeta_w \pi_{t-1} = \kappa_w \chi_t + \beta\theta_w m \left[1 + \frac{\kappa_w}{1 - \beta\theta_w m} \right] \mathbb{E}_t \left[\pi_{t+1}^w - \zeta_w \pi_t \right]$$

C.3 Model calibration and estimation

We now discuss the parameterization of our models. We proceed in two steps—first the calibration part, and then the estimation.

Calibration. For all three models, we calibrate the elasticity of intertemporal substitution and the Frisch elasticity to be $\frac{1}{2}$, which are standard values in the literature. For RANK, we set $\beta = 0.99$ (quarterly) in order to get a real interest rate of 4 percent annualized. For HANK, we pick β in order to match the same steady-state level of assets for all models. We calibrate the idiosyncratic income process for HANK from Kaplan et al. (2018).

We set the capital share to $\alpha = 0.36$ and depreciation rate to $\delta = 0.025$ quarterly, which is consistent with the values used in Christiano et al. (2005). The dividend distribution process is parameterized by assuming $\delta_1 = 0.2$ and $\delta_2 = 0.05$, which ensures a gradual payment of dividends and therefore low consumption response from capital gains.[38]

We follow Wolf (2023) for the steady state calibration of the fiscal side of the model. We assume a labor tax rate $\bar{\tau}_\ell$ of 0.3, and set transfers to be 5 percent of GDP. The steady state level of nominal assets is set to 27 percent of GDP, as in Kaplan et al. (2018). Government debt maturity is calibrated to $\eta = 0.2$, which implies an average debt duration of 5 quarters. The steady-state level of government expenditure is set such that the budget constraint holds

[38] As long as the pay-out is gradual, our results are not sensitive to the specific values used.

Parameter	Description	Value	Target
$1/\gamma$	EIS	0.5	Standard
$1/\varphi$	Frisch elasticity	0.5	Standard
$\bar{r}$	Real interest rate (annual)	0.04	Real interest rate
α	Capital share	0.36	Christiano et al. (2005)
δ	Depreciation rate (anual)	0.1	Christiano et al. (2005)
δ_1, δ_2	Dividend pay-out process	$0.2, 0.05$	Capital Gains MPC
$\bar{\tau}_\ell$	Labor tax rate	0.3	Average Labor Tax
$\bar{\tau}/\bar{y}$	Transfers	0.05	Wolf (2023)
$\bar{b}/\bar{y}$	Steady state liquid assets	1.04	Kaplan et al. (2018)
$1/\eta$	Liquid assets duration (quarters)	5	Kaplan et al. (2018)
τ_b^ℓ	Speed of fiscal adjustment	0.15	Gradual fiscal adjustment

Table C.1: Calibrated model parameters.

in steady state, which yields $\frac{\bar{G}}{\bar{Y}} = 0.1395$. We assume that all dynamic fiscal adjustment is done via labor taxes, with $\tau_b^\ell = 0.15$. This implies gradual fiscal adjustment, in line with the range considered in Auclert et al. (2020).

A summary of the calibrated parameter values is provided in Table C.1.[39]

Estimation. Table C.2 summarizes the posterior distributions of all estimated parameters. We see that, for h, κ and ψ, posterior distributions are relatively close to the prior. On the other hand, the distributions of θ_p, θ_w and θ are meaningfully affected. In the cases of θ_p and θ_w, the level of price and wage stickiness required to fit the impulse responses is relatively large, especially for prices; this reflects the known mismatch between micro level and macro level estimates of price rigidity, with macro estimates pointing towards much stickier prices than micro evidence. For the case of θ, a higher degree of informational stickiness is required to fit the empirical impulse responses than the one encoded in the prior. The degree of information rigidity is close to the one inferred in Auclert et al. (2020).

[39]The baseline determinacy-induced monetary policy rule that we consider sets $\rho = 0.85$, $\phi_\pi = 2$, $\phi_y = 0.25$, and $\phi_{\Delta y} = 0.3$. Recall that this choice of rule only matters for our illustrative results in Figures 2 and 3.

		Prior			Posterior				
Model	Parameter	Dist.	Mean	St. Dev	Mode	Mean	Median	5 percent	95 percent
RANK - RE	h	Beta	0.70	0.10	0.7240	0.7082	0.7157	0.5335	0.8571
	θ_p	Beta	0.67	0.20	0.9485	0.8622	0.9136	0.5598	0.9804
	θ_w	Beta	0.67	0.20	0.8860	0.7544	0.8091	0.3706	0.9657
	κ	Normal	5.00	1.50	5.0527	5.2668	5.2512	2.9409	7.6503
	ψ	Beta	0.50	0.15	0.4621	0.4665	0.4645	0.2247	0.7175
HANK - RE	θ	Beta	0.70	0.20	0.9526	0.7985	0.8414	0.4655	0.9813
	θ_p	Beta	0.67	0.20	0.9521	0.8527	0.9070	0.5381	0.9822
	θ_w	Beta	0.67	0.20	0.9031	0.7664	0.8215	0.3860	0.9683
	κ	Normal	5.00	1.50	5.3003	5.2477	5.2396	2.8847	7.6453
	ψ	Beta	0.50	0.15	0.4654	0.4678	0.4648	0.2247	0.7212
RANK - CD	h	Beta	0.70	0.10	0.7102	0.7112	0.7187	0.5373	0.8609
	θ_p	Beta	0.67	0.20	0.8641	0.8664	0.9061	0.6166	0.9769
	θ_w	Beta	0.67	0.20	0.9462	0.7564	0.8118	0.3691	0.9646
	κ	Normal	5.00	1.50	5.2998	5.3312	5.3211	3.0123	7.6935
	ψ	Beta	0.50	0.15	0.4701	0.4714	0.4686	0.2292	0.7244
HANK - CD	θ	Beta	0.70	0.20	0.9600	0.8112	0.8553	0.4866	0.9847
	θ_p	Beta	0.67	0.20	0.8544	0.8467	0.8917	0.5507	0.9794
	θ_w	Beta	0.67	0.20	0.9511	0.7860	0.8483	0.4005	0.9711
	κ	Normal	5.00	1.50	5.3222	5.3264	5.3186	2.9636	7.7030
	ψ	Beta	0.50	0.15	0.4692	0.4607	0.4571	0.2244	0.7119

Table C.2: Prior and posterior distributions of structural parameters. RE denotes that the model assumes rational expectations ($m = 1$), whereas CD indicates that the model features cognitive discounting in price and wage setters (with $m = 0.65$).

D Supplementary details for empirical applications

This appendix provides further details on our monetary policy counterfactual applications in Section 5. Appendix D.1 elaborates on our reduced-form projections, and Appendices D.2 to D.4 then contain several supplementary results for our three applications.

D.1 Reduced-form projections

We here provide supplementary details on how we construct our reduced-form projections. We elaborate on data construction and econometric implementation.

DATA. We consider the same ten observables y_t as in Angeletos et al. (2020). The series are constructed as follows. Unless indicated otherwise, each series is transformed to stationarity following Hamilton (2018), and series names refer to FRED mnemonics.

- *Unemployment rate.* We take the series `UNRATE` from FRED. We do not transform this series further.
- *Output gap.* We take log output per capita from FRED (`A939RX0Q048SBEA`). We interpret the stationarity-transformed series as a measure of the output gap.
- *Investment.* We compute log investment per capita, where investment is defined as the sum of durables and gross private domestic investment. We construct this series as (`PCDG`+`GPDI`)*`A939RX0Q048SBEA/GDP`.
- *Consumption.* We compute log consumption per capita, where consumption is defined as the sum of nondurables and services. We construct this series as (`PCND`+`PCESV`) * `A939RX0Q048SBEA/GDP`.
- *Hours.* We compute log hours worked, where total hours worked are constructed as `PRS85006023` * `CE16OV/CNP16OV`.
- *Utilization-adjusted TFP.* We compute the cumulative sum of the series `DTFPu`, from John Fernald's webpage (`https://www.johnfernald.net/TFP2023.03.07revision`).
- *Labor productivity.* We compute log labor productivity, where labor productivity is obtained as `OPHNFB`.
- *Labor share.* We compute the log labor share, with `PRS85006173` as the labor share.

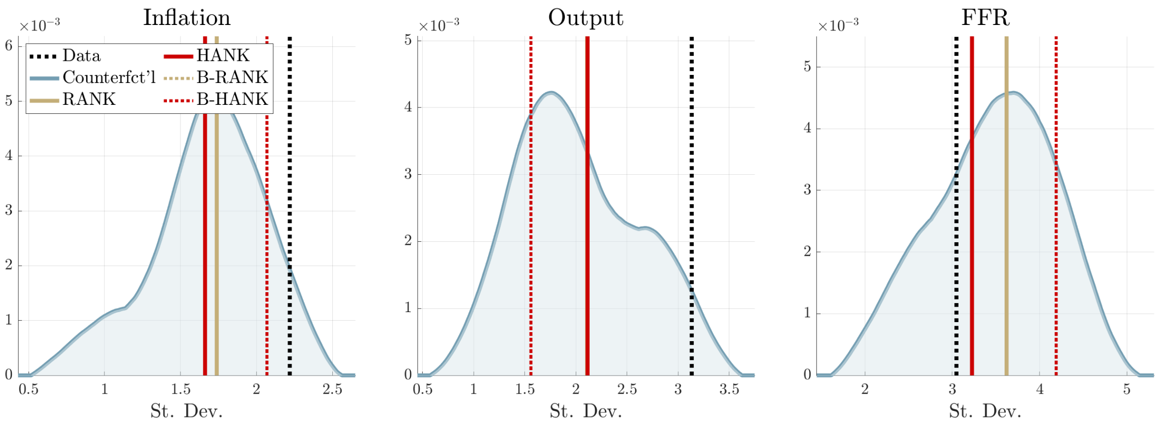

Figure D.1: Counterfactual early-sample (1960:Q1 – 2007:Q1) average volatilities of inflation, output, and the federal funds rate, under the policy rule that minimizes (54). Black dashed: data point estimate under observed policy. Blue: posterior Kernel density of counterfactual volatilities drawing from posterior across all models and parameters. Beige: posterior mode of counterfactual using RANK models (baseline and behavioral). Red: posterior mode of counterfactual using HANK models (baseline and behavioral).

- *Inflation.* We compute the log-differenced GDP deflator (`GDPDEF`), and then annualize, without further transformations.

- *Federal funds rate.* We obtain the series `FEDFUNDS`, without further transformations.

All series are quarterly. For the applications in Sections 5.2 and 5.3, we consider samples from 1960:Q1—2019:Q4. For the covid inflation counterfactual in Section 5.4, we extend the sample to 2021:Q2, the contemplated forecasting date.

Econometric implementation. Throughout we restrict attention to OLS point estimates of our ten-variable reduced-form VARs. We always include a constant and a linear time trend. For the second-moment counterfactual in Section 5.2 we include four lags, to allow for an accurate fit of the second moments. For the forecast-based counterfactuals in Sections 5.3 and 5.4, we include two lags.

D.2 Average business cycle

We here substantiate our claims that the headline finding of Section 5.2—i.e., that meaningful volatility reductions in output and, to a lesser extent, inflation would have been feasible—are driven by neither the Great Recession nor by the policy causal effect extrapolation embedded in our structural models.

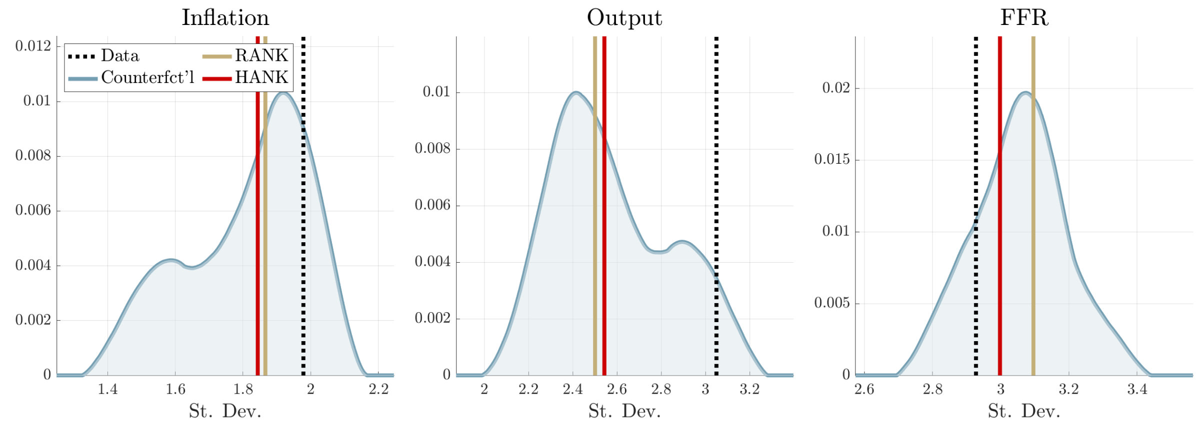

Figure D.2: Counterfactual average volatilities of inflation, output, and the federal funds rate, under the policy that minimizes (54). Black dashed: data point estimate. Blue: posterior Kernel density of estimates drawing from posterior across all models and parameters, using only the matched policy shock impulse responses $\hat{\theta}_\nu$. Beige: posterior mode for the baseline RANK model. Red: posterior mode for the baseline HANK model.

We begin in Figure D.1 by instead constructing counterfactual volatilities with reduced-form forecasts obtained on a sample that only stretches to 2007:Q1. We see that the picture is essentially unchanged relative to Figure 4: inflation and in particular output gap volatility reductions are feasible, at the cost of somewhat more volatile interest rates. This robustness is not surprising: the main business-cycle shock of Angeletos et al. (2020) meaningfully moves aggregate output even on pre-ZLB samples (while having rather little effect on inflation), so the exact same logic from our discussion in Section 5.2 continues to apply.

Next, in Figure D.2, we repeat our baseline analysis, but now minimizing (54) using only the matched policy shock impulse responses $\hat{\theta}_\nu$, rather than the entirety of the model-implied policy causal effect matrices Θ_ν. Qualitatively, the exact same picture emerges as before: inflation and in particular output are less volatile, while interest rates are somewhat more volatile. The intuition is yet again immediate from our analysis of the main business-cycle shock in Figure 5: through nominal interest rate cuts—including those directly matched in $\hat{\theta}_\nu$—the policymaker can meaningfully reduce the volatility of output fluctuations. The only difference is quantitative: the entire matrix of monetary policy causal effects Θ_ν allows the policymaker to tailor her interest rate response even better.

D.3 Great Recession

Figure D.3 constructs the Great Recession counterfactual using only the matched policy shock impulse responses $\hat{\theta}_\nu$ (rather than all of Θ_ν). As before, and exactly as in McKay and

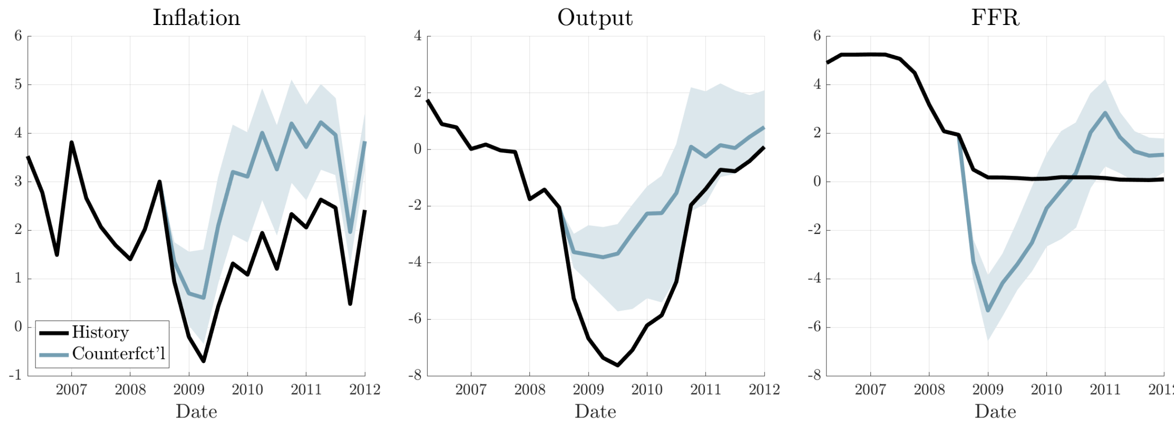

Figure D.3: Counterfactual evolution of inflation, output, and the federal funds rate in the Great Recession, under the policy that minimizes (54) without any effective lower bound on rates. Black: data. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded), using only the matched policy shock impulse responses $\hat{\theta}_\nu$.

Wolf (2023), we use $\hat{\theta}_\nu$ to enforce the counterfactual policy rule of interest as well as possible.

The resulting counterfactuals are again broadly similar to our main results in Figure 6: the nominal interest rate is cut aggressively, leading to more stable output, at the cost of elevated inflation. Moving from the restricted policy causal effect space $\hat{\theta}_\nu$ to the entirety of Θ_ν smoothes out the rate cut and helps somewhat better stabilize output; that being said, the differences are relatively small, suggesting that the counterfactual policy does not rely much on model-implied extrapolation to the causal effects of interest rates forward guidance.

D.4 Post-covid inflation

Figure D.4 complements the analysis in Section 5.4 by constructing the post-covid inflation forecasting counterfactual using only the matched monetary policy shock impulse responses $\hat{\theta}_\nu$ (and not all of Θ_ν). Recall that the disagreement across models discussed in Section 5.4 was precisely related to the model-based extrapolation from $\hat{\theta}_\nu$ to the rest of Θ_ν: in the rational-expectations models, the policymaker leverages small real rate hikes in the far future to stabilize inflation today; in the behavioral models, this is not possible, so interest rates are instead hiked already today.

As expected, Figure D.4 reveals that, when not relying on any extrapolation, the counterfactual looks much closer to the behavioral case: interest rates are actually hiked more aggressively, and that then leads to moderate declines in inflation and output.

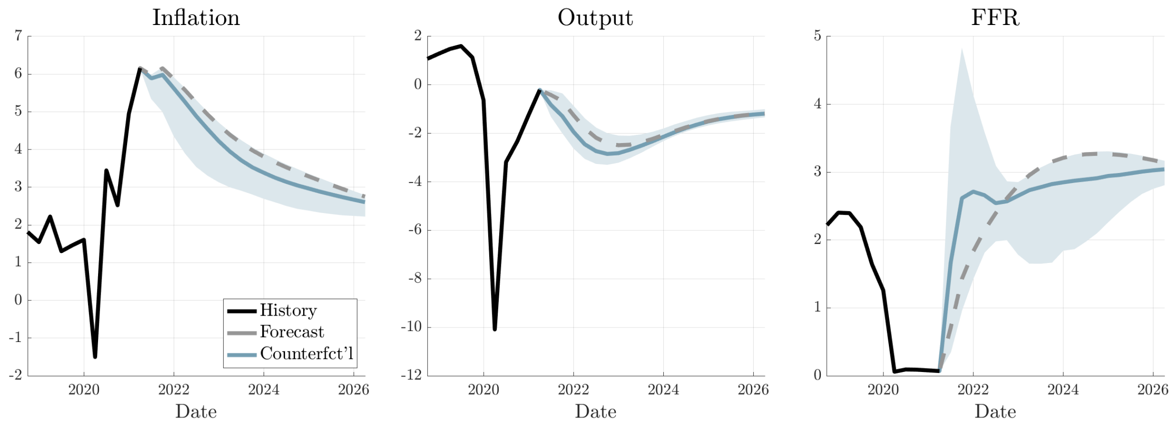

Figure D.4: Counterfactual projections of inflation, output, and the federal funds rate in the post-covid inflationary episode (from 2021:Q2), under the policy that minimizes (54). Policy causal effects from posterior across all models and parameters, but using only the matched policy shock impulse responses $\hat{\theta}_\nu$. Black: data. Grey: actual forecast. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded).