Evaluating Monetary Policy Counterfactuals: (When) Do We Need Structural Models?†

Tomás E. Caravello Alisdair McKay Christian K. Wolf

MIT FRB Minneapolis MIT & NBER

May 6, 2025

Abstract: To evaluate the evolution of the macro-economy under alternative assumptions on monetary policy, it suffices, under weak structural assumptions, to know the causal effects of monetary shocks on macroeconomic outcomes. The existing empirical literature estimates the effects of monetary shocks to the short end of the yield curve, thus allowing the evaluation of counterfactuals that change assumptions on near-term policy. If the contemplated policy changes are instead more persistent, then model structure becomes necessary, for one sole purpose: to extrapolate from the estimated short-end effects to those of policy shocks further out on the yield curve. Among popular models of monetary policy transmission, market incompleteness (i.e., "HANK") does not change this extrapolation much, while behavioral frictions do.

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

---

†Email: tomasec@mit.edu, alisdair.mckay@mpls.frb.org and ckwolf@mit.edu. We received helpful comments from Marios Angeletos, Régis Barnichon, Anmol Bhandari, Francesco Bianchi, Luigi Bocola, Ricardo Caballero, Gabriel Chodorow-Reich, Martin Eichenbaum, Simon Gilchrist, Cosmin Ilut, Giuseppe Moscarini, Mikkel Plagborg-Møller, Giorgio Primiceri, Valerie Ramey, Matt Rognlie, Karthik Sastry, Ben Schumann, Ludwig Straub, Iván Werning, and seminar participants at various venues. We also thank Seungki Hong, Klodiana Istrefi, and Diego Känzig for valuable discussions, and Valeria Morales Vasquez for superb research assistance. 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. An earlier version of this project was circulated under the title "Evaluating Policy Counterfactuals: A 'VAR-Plus' Approach."

1 Introduction

How would the macro-economy have evolved if monetary policy had been conducted differently? Following Lucas (1976), the standard way to evaluate such counterfactuals has been to go fully structural—i.e., build a model of macroeconomic fluctuations, change monetary policy in that model, and finally re-solve it.[1] A recent literature instead proposes to leverage empirically estimable "sufficient statistics" (Barnichon and Mesters, 2023; McKay and Wolf, 2023); the main idea here is that, across a large *family* of models, it is possible to evaluate the desired counterfactuals using the causal effects of monetary policy shocks, which in turn can be recovered using standard time-series methods. The key challenge limiting the applicability of this approach is that, in principle, it requires evidence on the effects of policy shocks along the *entire* yield curve—something the empirical literature clearly fails to deliver.

In this paper we propose a hybrid strategy, more widely applicable than the pure sufficient statistics approach, yet more transparent and robust than the hitherto dominant fully structural alternatives. The basic idea of our strategy is to take as many of the sufficient statistics as we can directly from the data, and then use the minimal structural assumptions necessary to extrapolate from those observed sufficient statistics to the ones we still lack. In the context of monetary policy, the directly measurable statistics are the causal effects of changes in near-term monetary policy, i.e., moving the short end of the yield curve. Extrapolation to the missing effects of policy shocks further out on the yield curve can then be achieved through structural models of monetary transmission. Appealingly, such extrapolation does *not* require the researcher to specify a full model of macroeconomic fluctuations; in particular she can be entirely silent on what shocks drive the business cycle, thus allowing our strategy to sidestep what is a key source of misspecification of complete structural models (e.g., see Chari et al., 2009). We then illustrate our approach with several applications to counterfactual monetary policy evaluation. Some only involve transitory changes in policy, so the empirically estimable sufficient statistics alone pin down the counterfactual; in others extrapolation to the long end of the yield curve matters more, and we compare and contrast how a range of popular models of monetary transmission achieve this extrapolation.

Background: sufficient statistics. We are interested in the counterfactual evolution of the macro-economy under alternative monetary policy rules, both on average—i.e., how

[1] Smets and Wouters (2007) is the exemplar of this approach. Recent examples are Crump et al. (2023) and Bocola et al. (2024), who re-evaluate U.S. monetary policy during the post-covid inflationary episode.

the "typical" business cycle would unfold—and for particular historical episodes. Across a quite general family of linearized structural macroeconomic models, these counterfactuals are pinned down by just two sufficient statistics (generalizing McKay and Wolf, 2023):

(i) *Reduced-form projections.* The first statistic is a set of reduced-form projections. For unconditional average business-cycle counterfactuals, those projections are impulse responses to reduced-form ("Wold") innovations. For counterfactuals conditional on particular historical episodes, the projections are forecasts, from each date in the episode of interest. Importantly, these reduced-form projections need to be relative to an information set that spans the (unknown) structural shocks buffeting the economy; identification of those primitive shocks themselves, however, is *not* required.

(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. In our monetary policy setting, this means the researcher needs to know the effects of changes in the policy rate both today and at all future horizons.

Sufficiency of the first statistic relies on the assumption of "invertibility." In standard macroeconometric practice, the role of the invertibility assumption is to ensure that the structural shock of interest can be recovered from the observables (e.g., Fernández-Villaverde et al., 2007). We show theoretically and through simulations that invertibility plays a very different role here: it ensures that projections are formed with respect to an information set that does not omit any important predictors of the macroeconomic outcomes of interest. Such forecasting, however, is a familiar task that in particular requires no further structural assumptions. It is because of this first "building block" insight that, for the remainder of the paper, we are free to fully focus our attention on the second and more challenging sufficient statistic: the causal effects of monetary policy along the *entire* yield curve.

Empirical evidence—the short end. We review the available empirical evidence on monetary policy shock propagation, and argue that it primarily pins down the causal effects of interest rate changes at the short end of the yield curve. Two representative examples are the classic monetary shock series of Romer and Romer (2004), which induces a very short-lived change in policy rates, and its recent refinement in Aruoba and Drechsel (2024), which has somewhat more persistent effects, consistent with increased gradualism in monetary policy practice. Many other recently popular monetary policy shock series, like those of Gertler and Karadi (2015) or of Miranda-Agrippino and Ricco (2023), also behave similarly. Even the "forward guidance" disturbance in Swanson (2024) only leads to relatively transitory revisions of the expected short-run interest rate path.

Putting together the sufficient statistics result with this empirical evidence, it follows that time series data alone will already largely suffice to evaluate monetary policy counterfactuals that only involve relatively short-lived policy changes. Our empirical applications will reveal that this case is actually quite typical in practice. However, for the more ambitious objective of evaluating *every possible* monetary policy counterfactual, there is of course still a missing piece: the effects of monetary shocks further out on the yield curve.

Model extrapolation—the long end. We propose to learn about the missing longer-horizon effects through model-based extrapolation: i.e., we build models of monetary policy transmission that are consistent with the available evidence on transitory monetary shocks, and then use them to impute the missing long-end effects. We implement the first step using impulse response-matching estimation techniques (Christiano et al., 2005).

Compared to full-information structural approaches, this targeted use of model structure has transparency and robustness benefits. For any given counterfactual, our strategy allows the researcher to decompose her results into the direct contribution of the data vs. the role played by model-based extrapolation, allowing her to see clearly what each part contributes. On robustness, a first obvious advantage is that, if the data alone suffice to evaluate a given counterfactual, then our approach indeed only uses the data. And if instead extrapolation is needed, then we leverage precisely the model structure that is needed to do so, *and nothing more*. For example, the question of what shocks drove historical business cycles simply has no bearing on the required extrapolation, and so it need not be answered. In contrast, full-information likelihood approaches invariably do need to make inferences on the origins of cycles, leading to likely misspecification because of "dubiously structural" shocks (Chari et al., 2009). Our strategy can similarly remain silent on the rule that monetary policy followed historically—yet another likely source of misspecification for full-information approaches. In short, our strategy gains robustness by dispensing with unnecessary assumptions.

We implement this extrapolation strategy using a menu of popular quantitative models of monetary policy transmission. Consistent with prior work, we document that representative-agent (RANK) and heterogeneous-agent (HANK) models alike can match the available empirical evidence on short-end monetary shocks. We then show that, once the two models are disciplined in this way, they will also largely agree on the long-horizon effects. Intuitively, while it is well-known that market incompleteness can in principle dampen or amplify monetary propagation (e.g., see Auclert, 2019), our insistence on identical short-end effects means that any heterogeneity-related channels largely offset, delivering as-if aggregation that echoes prior analytical results in Werning (2015). In particular we find that, in both models, short-run economic conditions are highly sensitive to assumptions about far-ahead interest rates. This contrasts sharply with alternative models featuring strong behavioral frictions in the expectations of pricing setters (as in Gabaix, 2020). We show that such models can be similarly consistent with the available short-end empirical evidence, but predict much weaker effects of far-ahead interest rate changes. This centrality of expectations echoes earlier insights from the "forward guidance puzzle" literature.

Applications. With the required causal effects in hand, we leverage the sufficient statistics result to evaluate three different monetary policy counterfactuals.

1. We ask whether monetary policy could have reduced the volatility of output and inflation over a post-war sample period. Our analysis suggests that substantial volatility reductions for output would have been feasible; strikingly, this takeaway does not at all rely on long-end extrapolation, but follows straight from the data: our reduced-form projections reveal that the economy is subject to frequent, transitory "demand-type" disturbances, and short-lived interest rate changes suffice to better stabilize these shocks.

2. We study how the Great Recession would have evolved in the counterfactual absence of a binding lower bound on nominal interest rates. We find that a standard "dual mandate" central bank would have liked to reduce interest rates substantially into negative territory, suggesting that the implemented unconventional policy measures were insufficient. This second counterfactual is also chiefly governed by changes in monetary policy conduct at the short end, and so yet again the long-end extrapolation matters relatively little.

These two applications are examples for which, in our hybrid approach, the empirically estimable sufficient statistics alone take center stage, with simple, well-accepted time series moments already enough to arrive at sharp counterfactual policy conclusions.

3. We evaluate monetary policy options after the summer of 2021, when inflation had started to accelerate. Under our estimated reduced-form projections the inflation spike is expected to be quite persistent, and so assumptions about longer-horizon changes in monetary policy take center stage. In counterfactuals based on extrapolation through standard RANK

and HANK models, the policymaker uses forward guidance to steer inflation expectations, reducing current inflation at little to no cost to output in the short run. In more behavioral model variants, this strategy is instead much less effective. To the extent that all of these models are regarded as *ex ante* plausible, there is thus necessarily large uncertainty about the optimal (counterfactual) path of interest rates.

Our third application is an example in which the model-based extrapolation to the long end very much matters; and since the models we consider disagree on this extrapolation, the reported counterfactual is highly uncertain. A key step towards reducing that uncertainty is to discriminate between models with and without behavioral frictions—and doing so in fact seems much more important than the incomplete-markets margin (i.e., RANK vs. HANK) that has received so much attention recently. In particular, our analysis demonstrates that differences in transmission *channels*—the direct-indirect dichotomy stressed by Kaplan et al. (2018)—may not matter at all for aggregate dynamics.

Further literature. Hybrid sufficient statistics and structural approaches are common in public economics, e.g., as discussed in Chetty (2009). Our method carries these insights forward to dynamic macroeconomic analysis, delivering the same transparency and robustness gains. Guided by our insights on the role of model structure in shaping counterfactual analysis, we provide novel comparisons on how the dominant structural frameworks in the monetary policy literature achieve the required extrapolation to the long end; as such, our analysis echoes the sufficient statistics results of the recent trade and New Keynesian pricing literatures (e.g., as discussed in Arkolakis et al., 2012; Auclert et al., 2022).

Relative to the sufficient statistics literature in macroeconomic policy evaluation that we cited above, our main value-added is the greater applicability of our hybrid approach. We furthermore provide some additional identification results, including a much more detailed discussion of the understudied yet key role played by the invertibility assumption. Just as in that literature, and differently from Sims and Zha (1995), the estimand of our proposed strategy is the effect of a systematic change in the policy rule that is communicated to and well-understood by the private sector, thus fully respecting the Lucas (1976) critique.

Outline. Section 2 begins with the sufficient statistics identification results. Section 3 then reviews the empirical evidence on monetary policy shock propagation, while Section 4 uses model structure for extrapolation. Our applications to several monetary policy counterfactuals follow in Section 5, and Section 6 concludes.

2 Identification results

This section presents the identification results that underlie our hybrid approach to monetary policy counterfactual analysis. Section 2.1 introduces the setting and then defines the objects of interest, Section 2.2 states the sufficient statistics identification results, and Section 2.3 investigates the role of the invertibility assumption for our arguments. The analysis closely builds on but further extends McKay and Wolf (2023), and is written to be self-contained.

2.1 Setting and objects of interest

We begin with a description of the environment. While the notation is purposefully general, we will connect to our monetary policy applications when appropriate.

We refer the reader to McKay and Wolf (2023) for a further discussion of the scope and limitations of the model setting; for our purposes, it suffices to note that typical linearized business-cycle models—from representative-agent New Keynesian models (Christiano et al., 2005; Smets and Wouters, 2007), to heterogeneous-agent environments (à la Kaplan et al., 2018), and also including certain models with behavioral frictions (e.g., like Gabaix, 2020)—are all nested in the general environment considered here.

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

$$y_t = \sum_{\ell=0}^{\infty} \Theta_\ell \varepsilon_{t-\ell}. \tag{1}$$

y_t is a vector of macroeconomic aggregates, the vector of aggregate shocks ε_t is distributed as $\varepsilon_t \sim N(0, I)$, and the $n_y \times n_\varepsilon$-dimensional matrices Θ_ℓ denote the impulse response of y_t at horizon ℓ to a date-t vector of shocks ε_t.[2] In 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 classic Frisch (1933) impulse-propagation paradigm, the system (1) allows for an unrestricted dynamic linear transmission from shocks ε_t to outcomes y_t.

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

[2] We impose the high-level assumption that the matrices Θ_ℓ are absolutely summable across ℓ.

all variables are expressed in deviations from the deterministic steady state. The economy is summarized by the system

$$\mathcal{H}_w \boldsymbol{w} + \mathcal{H}_x \boldsymbol{x} + \mathcal{H}_z \boldsymbol{z} + \mathcal{H}_e e_0 = \mathbf{0}, \tag{2}$$

$$\mathcal{A}_x \boldsymbol{x} + \mathcal{A}_z \boldsymbol{z} + \mathcal{A}_v v_0 = \mathbf{0}. \tag{3}$$

Here x_t and w_t are n_x- and n_w-dimensional vectors of endogenous variables, respectively, z_t is an n_z-dimensional vector of policy instruments, e_t is an n_e-dimensional vector of exogenous structural shocks, v_t is an n_v-dimensional vector of policy shocks, and we write $y_t = (x_t', z_t')'$, $\varepsilon_t = (e_t', v_t')'$.[3] The distinction between w and x is that the variables in x are observable while those in w are not. Equation (2) summarizes the $n_x + n_w$-dimensional non-policy block of the model, with $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_e\}$ embedding private-sector relations. Equation (3) is the policy rule, with the policy instrument z set as a function of x and v. In the monetary policy applications considered in the remainder of this paper, z will be the policy rate, and x will typically include measures of aggregate output and inflation.[4]

Given the date-0 shocks $\{e_0, v_0\}$, impulse responses are sets of bounded sequences $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}\}$ that solve (2) - (3). We assume that this solution exists and is unique, and write it as

$$y_\ell = \Theta_\ell \cdot \varepsilon_0.$$

Stacked together, those perfect-foresight mappings from date-0 shocks to date-ℓ outcomes deliver the SVMA(∞) representation (1).

Policy counterfactuals. We wish to study the evolution of the economy if policy were instead, and counterfactually, set as

$$\tilde{\mathcal{A}}_x \boldsymbol{x} + \tilde{\mathcal{A}}_z \boldsymbol{z} = \mathbf{0} \tag{4}$$

rather than (3). Assuming that this policy rule similarly induces equilibrium existence and uniqueness, we recover the counterfactual SVMA process

$$\tilde{y}_t = \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \varepsilon_{t-\ell}, \tag{5}$$

[3] 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.

[4] In this application, the fiscal policy rule—which will be held fixed in our counterfactual analysis—would actually be part of the "non-policy" block (2).

with the convention that $\varepsilon_t = e_t$, and where the shock impulse responses $\tilde{\Theta}_\ell$ are derived from the solution of the perfect-foresight system (2) together with (4). We wish to evaluate the effects of the counterfactual policy change both *unconditionally*, for the "average" business cycle, and *conditionally*, for particular historical episodes.

1. **Unconditional business cycles.** In the linear SVMA setting considered here, a complete summary of the counterfactual stochastic properties of our economy under the alternative policy rule (4) is provided by the counterfactual autocovariance function of y_t. By standard arguments it is given as

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

   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.** We distinguish two kinds of conditional counterfactuals—forecasts, and full historical episodes.

   (i) *Conditional forecasts.* Consider some date t^*, and suppose the policymaker from t^* commits to the new rule (4). We may ask how, from that point onward, the economy would be *predicted* to evolve; i.e., we would like to recover the expectation

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

   where $\tilde{y}^*_{t^*+h}$ reflects how the policy change at t^* revises policy promises made and thus expectations formed before t^*, and is defined in Appendix A.1. 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 (4) from date t_1 onward; i.e., we seek to recover

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

   where again the second term reflects revisions of past promises (see Appendix A.1). Policy counterfactuals for particular historical episodes have similarly been the subject of much prior work; e.g., see Eberly et al. (2020) for a recent example.

The next section identifies sufficient statistics that, across the family of environments considered here, suffice to pin down all of these counterfactuals of interest.

2.2 A sufficient statistics identification result

We begin by introducing the two sufficient statistics at the heart of the identification result.

The first sufficient statistic is simply the Wold representation of y_t under the actual (not counterfactual) prevailing policy rule, given as

$$y_t = \sum_{\ell=0}^{\infty} \Psi_\ell u_{t-\ell}, \tag{6}$$

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 under the baseline SVMA (1), $\text{Var}(u_t^\dagger) = \Sigma_u$, 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.

The second sufficient statistic are the causal effects of shocks to the policy instrument z *at all possible horizons*. To define these effects consider the following generalized policy rule:

$$\mathcal{A}_x \boldsymbol{x} + \mathcal{A}_z \boldsymbol{z} + \boldsymbol{\nu} = \mathbf{0}. \tag{3'}$$

While the *actual* policy rule (3) is subject only to the n_v-dimensional vector of policy shocks v_t (which may be low-dimensional, or in fact even empty), the policy shock vector $\boldsymbol{\nu}$ in the artificial rule (3') is instead unrestricted—i.e., we allow for arbitrarily flexible wedges in the policy rule at each date $t = 0, 1, 2, \ldots$. Analogously to Section 2.1, the solution of (2) - (3') given an arbitrary policy shock vector $\boldsymbol{\nu}$ alone yields

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

Θ_ν is the *space* of y-allocations implementable through policy shocks—i.e., the paths of macro aggregates corresponding to any possible *time path* of the policy instrument.[5] For example, in our later monetary policy applications, Θ_ν is the space of paths of macroeconomic outcomes

[5] Θ_ν here is defined through shocks to the particular rule in (3'). As discussed in McKay and Wolf (2023), this choice of rule is immaterial—Θ_ν could be defined through shocks to an alternative determinacy-inducing policy rule, and all subsequent results would go through unchanged. This basic insight will lie at the heart of our particular (rule-free) implementation of impulse response-matching in Section 4.

x (e.g., of aggregate output and inflation) that is implementable through all possible paths of nominal interest rates z; i.e., through monetary policy shocks at every single point along the yield curve, from the short end to the long end.[6]

The identification result now states that, under the additional assumption of *invertibility*, these two objects actually suffice to recover all of our policy counterfactuals of interest. We note that the first part of this result—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.

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

$$\varepsilon_t \in \text{span}(\{y_\tau\}_{-\infty<\tau\leq t}). \tag{7}$$

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* $\mathbb{E}_t\left[\tilde{y}_{t+h}\right]$*.*

Proof. Please see Appendix A.2. □

Under the conditions of Proposition 1, policy shock causal effects together with a standard reduced-form time-series object—the Wold representation of the data—suffice for counterfactual policy evaluation. In particular, *no* knowledge is required of the true structural shocks ε_t that drive the cyclical fluctuations. This insight will be central to the greater robustness of our hybrid approach vis-à-vis standard full-information strategies.

In the constructive proof of the proposition, knowledge of the policy causal effects Θ_ν is used to evaluate the counterfactual propagation of the reduced-form innovations u_t. In words, in the first step of the proof, we simply take the impulse responses corresponding to each reduced-form innovation $u_{i,t}$, and then add to this the impulse responses to an artificial vector of policy shocks $\boldsymbol{\nu}$, chosen to enforce the hypothesized counterfactual policy rule. The key twist is that, *under invertibility*, doing so turns out to be equivalent to counterfactual analysis using the dynamic causal effects not of the reduced-form u_t's, but of the *true* shocks ε_t. The remainder of this section discusses the intuition for this result, focusing on the key role that is played by the invertibility assumption. The remainder of the paper will then be

[6] We stress that, while we use policy shocks $\boldsymbol{\nu}$ at all possible horizons to *define* the full causal effect map Θ_ν, the *actual* policy conduct need not be subject to a similarly rich set of shocks. Of course, if the actual shocks v_t are much lower-dimensional, then empirical evidence alone is insufficient to recover all of Θ_ν, and so model-based extrapolation will be needed—our analysis in Sections 3 and 4.

concerned exclusively with part (ii) of the informational requirements in Proposition 1, the policy causal effects Θ_ν.

2.3 Invertibility and forecasting

This section digs deeper into the role played by invertibility in delivering Proposition 1, and in particular on why invertibility allows us to sidestep the need to say anything about the deep shocks ε_t driving the business cycle. We present both theory and model-based simulations, together establishing that, in practice, all that is required to leverage Proposition 1 is a good forecasting model for the variables of interest y_t (plus of course the policy causal effects).

Forecasting intuition. The key insight to understanding the role of invertibility is to realize that all of our counterfactuals are ultimately nothing but forecasts. For conditional forecasts, this statement is immediate. For historical episodes, to evaluate the desired counterfactual, we need to figure out how the hypothesized change in policy would affect realized outcomes as well as forecasts *at each date* in the sample period under consideration. And unconditionally, the same can be done not for a particular episode, but at the "typical" state of the business cycle. We formalize this intuition in Appendix A.3, where we establish that the Wold representation of y_t suffices for policy evaluation *if and only if* the forecasts it implies are identical to full-information forecasts.[7] Conditional on access to such full-information forecasts, knowledge of the primitive structural shocks ε_t driving the cycle is entirely irrelevant. Intuitively, we do not need to know *why* the macro-economy fluctuates—we just need to know by *how much*, both today and in expectation tomorrow.

This forecasting intuition reveals that, in the context of Proposition 1, invertibility is sufficient but actually not necessary. As mentioned above, in the constructive proof of that proposition, policy shock causal effects Θ_ν are used to predict the counterfactual propagation of reduced-form Wold innovations u_t. The forecasting intuition reveals that doing so can be correct even if invertibility is strictly speaking violated—what is required is simply that Wold projections are formed with respect to an information set that does not omit any relevant predictors of the macroeconomic outcomes of interest. In the remainder of this section we illustrate this theoretical observation with model-based simulations, establishing that even small information sets can suffice for reliable counterfactual analysis, as long as the implied

[7] As discussed in Appendix A.4, our identification results *are* consistent with various kinds of frictions in private-sector expectation formation; however, the forecasts that the econometrician leveraging Proposition 1 constructs still always need to be full-information forecasts.

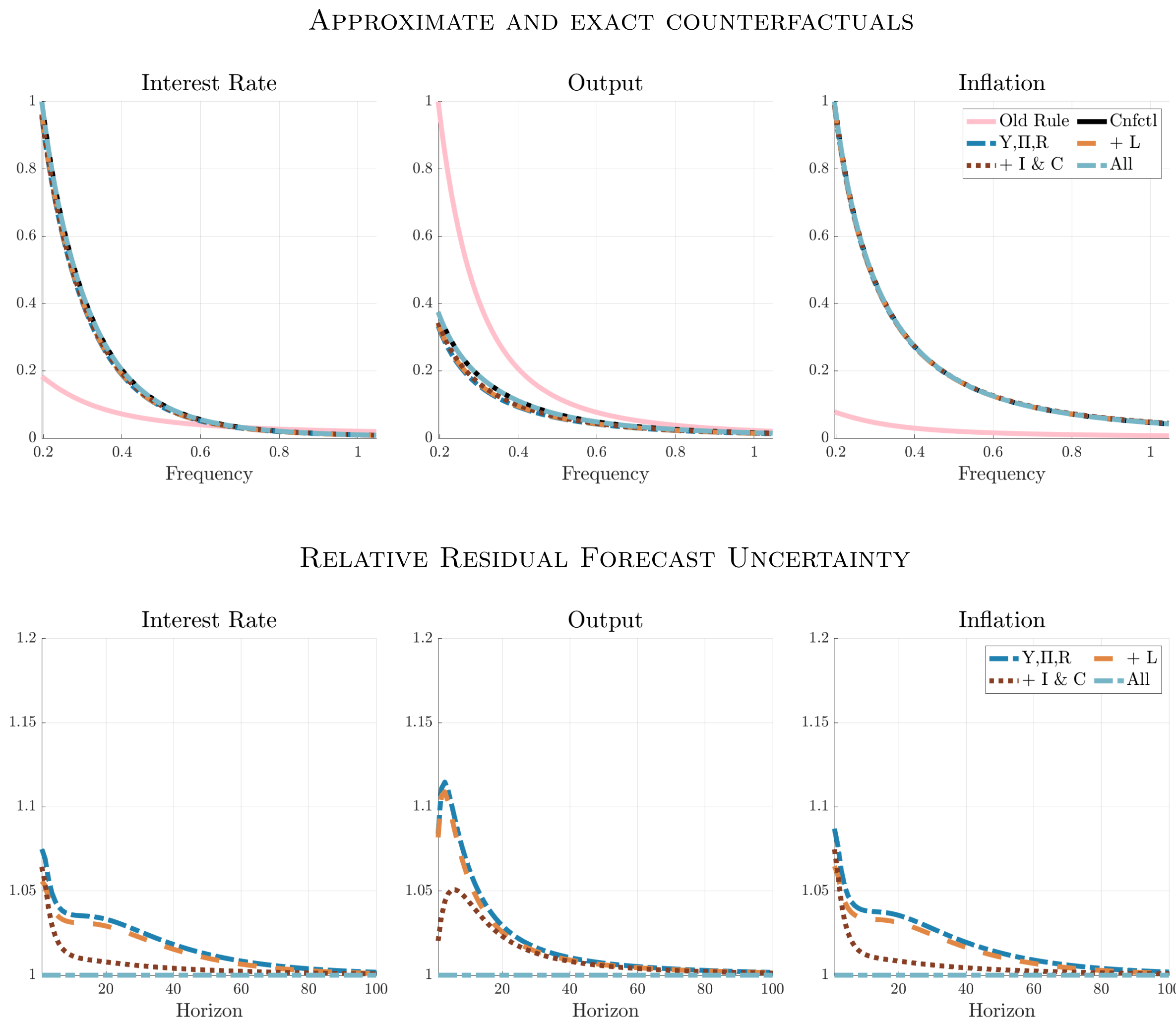

Figure 1: Top panel: business-cycle spectral densities for interest rates, output, and inflation under the old rule (solid pink) and under the counterfactual rule, true (solid dark blue) and predicted using Proposition 1 for different information sets (solid-dashed blue, dashed orange, dotted red, solid-dashed cyan). We normalize the peak spectral density to 1. 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.

reduced-form forecasts are accurate.

Simulation evidence. For our illustrations we consider a structural model—the medium-scale DSGE model of Smets and Wouters (2007)—as an artificial laboratory. We seek to recover the counterfactual second moments of output, inflation, and interest rates under an alternative monetary policy rule that puts a larger weight on output stabilization. To do so, we leverage Proposition 1, using the true matrix of policy shock causal effects Θ_ν, but then relying on information sets $\{y_{t-\ell}\}_{\ell=0}^{\infty}$ that are (potentially) insufficient to deliver invertibility. When invertibility fails, then this procedure will not exactly recover the true counterfactual; our question is just how inaccurate those predictions will end up being. For our explorations we will consider four 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. Among those four information sets, only the fourth one satisfies invertibility. Results are reported in Figure 1.

The headline takeaway is that even small information sets can deliver predicted policy counterfactuals that are almost indistinguishable from the true ones. The top panel summarizes second moments via spectral densities over business-cycle frequencies, with the pink lines corresponding to the prevailing monetary policy rule, while the other lines indicate the true (solid) and predicted (dashed) counterfactual spectral densities. We see that, for *all* information sets, the predicted counterfactuals are close to each other, and so to the truth. The explanation follows from our forecasting intuition given above, and is illustrated by the bottom panel. That panel shows residual forecast uncertainty for interest rates, output, and inflation at different horizons (x-axis), and for our different information sets (different lines), with the residual forecast uncertainty under the full information set normalized 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 instead strictly larger for smaller information sets. The differences, however, are moderate, with forecast variances that are only at most around 10 per cent larger than with the full information set. Even the small information sets can thus deliver accurate forecasts, and thereby also accurate counterfactuals. In other words, the key requirement is to have forecasts of output, inflation, and interest rates close to the accuracy of the full-information benchmark—and for that, the past history of those three series evidently suffices.

Since the Smets and Wouters model features seven shocks, the previous discussion substantiates our point of invertibility mattering only through forecasts: with three observables we are necessarily quite far from invertibility, and yet the reported counterfactuals are nevertheless accurate, simply because the formed forecasts are close to the full-information ones.

2.4 Takeaways and outlook

Our analysis in this section has added a simple yet practically important takeaway to the recent literature on sufficient statistics for policy evaluation (as in McKay and Wolf, 2023; Barnichon and Mesters, 2023): for all of the counterfactuals that we consider here, knowledge of primitive structural business-cycle shocks is actually not at all necessary—reduced-form forecasts that do not omit any important predictors are entirely sufficient. This observation will loom large when we compare our proposed approach to hitherto dominant full-information strategies for counterfactual policy evaluation.

Since forecasting is a relatively well-understood task, this insight frees the remainder of the paper to instead be concerned exclusively with the second sufficient statistic: the matrix of policy causal effects Θ_ν. We begin with empirical evidence on monetary policy propagation in Section 3, before turning to model-based causal effect extrapolation in Section 4—the two steps of our proposed "hybrid" approach.

3 Empirical evidence on monetary policy shocks

In this section we review what existing empirical work can tell us about the monetary policy shock causal effects Θ_ν that are required by our sufficient statistics results. We begin with some conceptual points on how to connect empirical analyses with our identification theory. We then use this lens to interpret the existing empirical work, and argue that it primarily teaches us about the effects of monetary shocks at the short end of the yield curve.

3.1 Interpreting empirical work

To interpret the estimand of empirical work on monetary shock identification, we return to our general formulation of policy rules in Section 2.2, re-stated here for convenience:

$$\mathcal{A}_x \boldsymbol{x} + \mathcal{A}_z \boldsymbol{z} + \boldsymbol{\nu} = \mathbf{0}. \tag{3'}$$

This general sequential formulation of the policy rule prescribes feedback from current and expected future endogenous macroeconomic outcomes $\boldsymbol{x}$ to current and expected future policy instrument paths $\boldsymbol{z}$, subject to shocks $\boldsymbol{\nu}$. In the monetary policy setting that we focus on here, the instrument is the short-term policy rate i, and a more familiar recursive formulation of a general monetary policy feedback rule takes the form (e.g., Woodford, 2003)

$$i_t \;=\; \underbrace{\sum_{\ell=-\infty}^{\infty} A_\ell \mathbb{E}_t\left[x_{t-\ell}\right]}_{\text{systematic feedback}} \;+\; \underbrace{\nu_{0,t} + \nu_{1,t-1} + \nu_{2,t-2} + \dots}_{\text{policy shocks}} \tag{8}$$

where the first part of (8) reflects the systematic feedback from current, lagged, and expected future macroeconomic outcomes to the policy instrument, and the second part is a general set of contemporaneous and news shocks to the policy rule, with the shock $\nu_{\ell,t-\ell}$ realizing at date $t-\ell$ and affecting policy at date t.[8] Empirical work on monetary policy transmission seeks to recover (instrumental variables for) the policy shocks $\nu_{\ell,t-\ell}$, thereby allowing the estimation of their dynamic causal effects through standard reduced-form methods like local projections or vector autoregressions. In general, a valid instrumental variable s_t in this setting is simply a series that correlates with (any combination of) the monetary policy shocks $\nu_{\ell,t-\ell}$, and not with any of the other current or lagged non-policy shocks e_t. Projection on this instrument series then recovers the corresponding weighted average treatment effect of the particular mix of policy shocks that correlate with s_t, by standard arguments (e.g., see Plagborg-Møller and Wolf, 2022, Appendix B.3).

To connect this interpretation of the estimand of empirical methods for policy propagation with our theoretical sufficient statistics, the key conceptual step is to note that, in the model environment of Section 2.1, policy affects the rest of the macro-economy exclusively through current and expected future values of the policy instrument, $\boldsymbol{z}$. Knowledge of the propagation of the policy shocks $\boldsymbol{\nu}$ is sufficient precisely because, and *only* because, it can be used to engineer any such instrument path. It follows that, when interpreting the output of empirical work on monetary policy shock propagation, it is immaterial whether any given identified shock or instrument is labeled as "contemporaneous" or as "forward guidance"; rather, all that matters is the current and expected future path of the policy rate that the shock induces, i.e., the change in the yield curve.[9] To summarize, an n_s-dimensional vector of policy shock instruments simply delivers an n_s-dimensional subset $\theta_\nu \subset \Theta_\nu$ of the full space of allocations implementable through policy, i.e., of our second sufficient statistic.

3.2 Monetary policy at the short end of the yield curve

The empirical literature on monetary policy shock propagation is voluminous, and was reviewed recently in Ramey (2016). In principle, the different shocks and instrumental variables

[8]See Wolf (2024, Appendix D) for a general discussion of the mapping between recursive formulations like (8) and sequential rules like (3').

[9]A different way of seeing this is to note the following: shocks that are labeled "contemporaneous" with respect to a policy rule featuring endogenous persistence would be labeled as having a "forward guidance" component when hitting a rule without such endogenous persistence. The labels are thus immaterial—all that matters are the induced policy instrument paths.

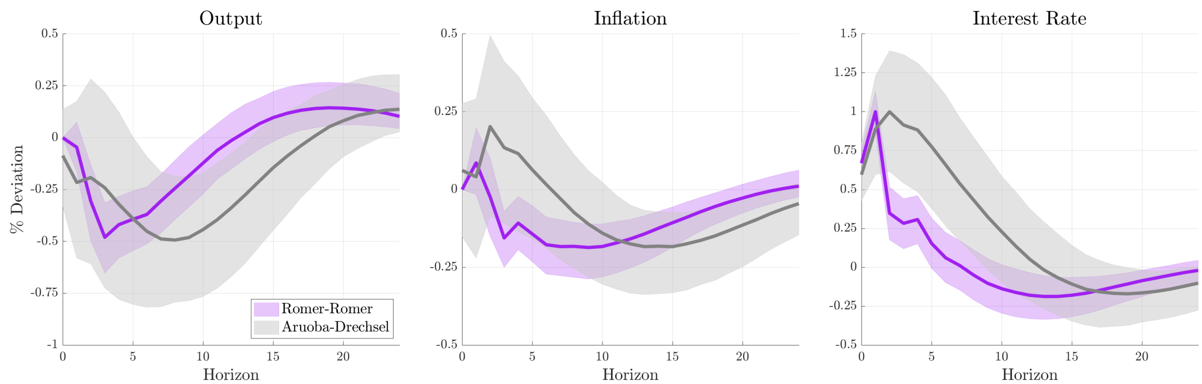

Figure 2: The purple and grey lines and shaded areas indicate impulse responses (with 16th and 84th percentile confidence bands) for the monetary policy shocks of Romer and Romer (2004) and Aruoba and Drechsel (2024). For estimation details please see Appendix C.1.

studied in that literature could lead to very heterogeneous monetary policy "treatments"—some monetary policy events may primarily revise the short end of the yield curve, while others may instead revise the medium or the long end, just as required to leverage our identification results. In practice, however, most of the historical events that are used to learn about the dynamic propagation of monetary policy shocks tend to primarily move the short to medium end of the yield curve. A representative example of this is Swanson (2024), which explicitly distinguishes between (purportedly more short-term) "federal funds rate" shocks as well as more long-term "forward guidance" shocks. In spite of their seemingly different nature, these two shocks actually both lead to broadly similar policy rate path revisions, with the latter only moderately more persistent.[10] And consistent with this observation, many of the other historically and recently popular monetary policy shock series—from Romer and Romer (2004) to the more recent Gertler and Karadi (2015), Jarociński and Karadi (2020), Miranda-Agrippino and Ricco (2023), or Aruoba and Drechsel (2024)—may differ somewhat in the precise timing of the induced nominal interest rate path revisions, but still all broadly move the short- to medium-end of the yield curve.

Given these takeaways, our summary of the empirical literature on monetary policy shock transmission is that it really only robustly pins down the effects of shocks to the short end of the yield curve. Specifically, we will take as our empirical reference point the dynamic causal effects displayed in Figure 2, estimated through projection on the classical monetary policy shock series of Romer and Romer (2004) as well as the recent refinement in Aruoba and

[10]Similar conclusions on the relatively small differences between interest rate paths induced by "contemporaneous" and "forward guidance" shocks are reported in Bundick and Smith (2020).

Drechsel (2024); estimation details are provided in Appendix C.1. We take these shocks as representative examples because they rely on relatively similar approaches to identification, yet somewhat differ in the induced interest rate time profiles, with the more recent series of Aruoba and Drechsel leading to a more delayed adjustment, consistent with greater policy gradualism in the more recent past. The two induced policy treatments, displayed in the right panel, lead to moderately persistent falls in the level of aggregate output (left panel) as well as a more delayed fall in the inflation rate (middle panel); as expected, since the Romer and Romer treatment is more transitory, the induced output and inflation responses are also more short-lived. Both qualitatively and quantitatively these patterns are in line with those estimated in other empirical work that uses different monetary policy shock series, including in particular the studies cited above.

To summarize, the empirical evidence has delivered a two-dimensional subset θ_ν of the full space of policy causal effects Θ_ν. We are thus not yet ready to implement our sufficient statistics identification result for the evaluation of *arbitrary* monetary policy counterfactuals: we only know the effects of shocks to expectations of policy over the subsequent year or two, and not of shocks to longer-term interest rate expectations. The next section addresses how we can leverage additional structural assumptions to learn about those missing causal effects of policy shocks further out on the yield curve.

4 Structural extrapolation to long-end shocks

In this section we use model structure to go from the estimated effects of monetary shocks at the short end of the yield curve, θ_ν, to the long end, thus completing the desired causal effect matrix Θ_ν. We first discuss how much model structure is needed for this purpose, contrasting our strategy with the informational requirements of conventional full-information approaches to policy evaluation. We then sketch the models of monetary policy transmission that we use, estimate them, and finally leverage them for extrapolation.

4.1 How much structure is actually needed?

Using the notation of Section 2.1, the dynamic causal effects of changes in policy on macroeconomic outcomes are pinned down by the tuple $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$—i.e., the part of the model structure that governs policy propagation are the model's non-policy relations, but *not* the structural shocks to those relations $\{\mathcal{H}_e e_0\}$, nor the policy rule $\{\mathcal{A}_x, \mathcal{A}_z\}$. The intuition is straightforward, and echoes insights from the analysis of static markets of demand and supply: to predict the effects of demand (say) changing to some hypothetical level, all we need to know are (i) initial outcomes and (ii) the slope of the supply function, but *not* the slope of demand, nor the actual shocks to demand or supply that generated the observed outcomes. In our dynamic macro setting, we similarly require initial outcomes (i.e., the autocovariance function) as well as the non-policy block (to deliver the causal effects of policy changes), but neither the policy rule nor the history of true policy and non-policy shocks.

These insights naturally suggest a partial-information, "impulse response-matching" approach (Rotemberg and Woodford, 1997; Christiano et al., 2005, 2010) to learning about the full matrix of monetary policy causal effects Θ_ν: first specify a *partial* model of monetary policy transmission (i.e., model structure that delivers a tuple $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$), then estimate it targeting the available empirical evidence on policy shocks at the short end of the yield curve (θ_ν, from Figure 2), and finally use this estimated structural model to learn about the missing long-end causal effects of monetary policy.[11] This strategy is a hybrid of empirical sufficient statistics and structural approaches in the exact same spirit as discussed in Chetty (2009, p. 455), just there in the context of public finance: theory is used to identify the relevant sufficient statistics, they are measured from data whenever possible, and the empirics are supplemented with model-based extrapolation when necessary.

VS. PURE SUFFICIENT STATISTICS APPROACHES. The proposed strategy nests but generalizes the sufficient statistics approach in McKay and Wolf (2023). If the empirically estimable sufficient statistics alone are already enough to evaluate the counterfactual of interest, then our approach here will by construction deliver the exact same conclusions—conclusions that are robust across a very wide *class* of models, with any further model-based extrapolation entirely irrelevant.[12] If the evidence does not suffice, then our strategy still remains applicable, now using the minimal amount of structure necessary to arrive at an answer.

VS. FULL-INFORMATION APPROACHES. By only leveraging exactly the parts of model structure that are essential for the task at hand, our proposed hybrid approach works with strictly weaker structural assumptions than the hitherto dominant full-information, likelihood-based

[11]In the standard impulse response-matching literature, the researcher writes down a policy rule, and then restricts attention to contemporaneous shocks to that rule. Commitment to a rule, however, is actually not at all necessary—one can simply find the best fit to the empirical impulse response targets within the overall *space* implementable by policy, sidestepping the need to commit to any rule. See Appendix B for details.

[12]This statement of course presupposes that the estimated model(s) can match the empirical evidence θ_ν. In our applications we make sure that this is the case.

strategies (as in Smets and Wouters, 2007; Justiniano et al., 2010). With such strategies, the researcher needs to correctly specify the *entirety* of the structural model of business-cycle fluctuations; perhaps most dubiously, she must correctly specify where exactly the underlying shocks should enter, how many there should be, and what stochastic processes they should follow. This obvious source for model misspecification has long attracted scrutiny, perhaps most notably in Chari et al. (2009), who argue that the misspecification of shocks invariably undermines the credibility of model-based counterfactual analyses. The full-information approach must similarly correctly specify the systematic rule followed by the policymaker—yet another Herculean task, given that historical policy conduct is unlikely to be well-described by any simple rule. Our approach avoids making such assumptions because they are simply not necessary; all the relevant information is already embedded in the reduced-form projections of Proposition 1. Structural assumptions are instead only made where they are needed, for policy causal effect extrapolation. We provide further discussion, touching also on some related concerns about weak identification, in Appendix B.3.

In addition to these robustness appeals, our proposed hybrid approach also has a transparency benefit: for any given policy counterfactual, the researcher can assess to what extent her conclusions are actually driven directly by the empirically measurable sufficient statistics, and what role the model-implied policy causal effect extrapolation is playing. Formally, to do so, she can simply compare the exact counterfactual implied by the full (model-extrapolated) policy matrix Θ_ν with an approximate counterfactual that is restricted to use only the empirical evidence on the policy's causal effects, θ_ν.[13] Decompositions of this sort are of course never available under the standard full-information approach. Overall, the robustness and transparency benefits that we describe here closely echo those of analogous "hybrid" strategies in other areas of economics, as perhaps most notably popular in public economics (see the literature review in Chetty, 2009).

Outlook. In the remainder of this section we use the impulse response-matching strategy sketched above to extrapolate monetary policy shock causal effects to the long end of the yield curve. We relegate implementation details to Appendix B; we there in particular elaborate on why, in contrast to existing implementations of impulse response-matching estimation (e.g., as in Christiano et al., 2010), we can actually avoid specifying a monetary policy rule.

[13] The approximate counterfactual can be constructed exactly as in McKay and Wolf (2023, Section 3), using the best linear approximation to the counterfactual of interest within the space of policy effects spanned by the available empirical evidence. See also Section 5.

4.2 Models of monetary policy transmission

For our impulse response-matching estimation and causal effect extrapolation, we will consider several standard models of monetary policy transmission. Our first model is a relatively 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 then enriches the consumer block to feature heterogeneous households with uninsurable income risk, e.g., 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. Finally, we will 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 monetary policy dynamic causal effect extrapolation to the long end of the yield curve.

The remainder of this section proceeds as follows. We will first sketch the representative-agent, rational-expectations model. We then explain how the heterogeneous-agent and behavioral models depart from this benchmark. Throughout we will only provide brief verbal descriptions, with details in Appendix C.2. We do so because all the models we consider are standard; our contribution is instead in how we *use* these models for extrapolation.

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 short-end empirical monetary shock evidence documented in Section 3.2. Following Christiano et al. (2005), 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 separable preferences for consumption and leisure, and allowing for habit formation. This agent chooses paths for consumption and assets to maximize lifetime utility. Labor supply is intermediated by labor unions (just as in Erceg et al., 2000), with households taking hours worked 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).

- *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; they sell their good to retailers who costlessly differentiate it, and set prices subject to Calvo frictions. Prices that are not re-optimized are fully indexed to lagged inflation (as in Christiano et al., 2005). The intermediate goods producers purchase capital goods from 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, and sell the capital good.

- *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. As discussed in Section 4.1 (and further in Appendix B.2), we do not need to specify any particular policy rule.

Stacking all model blocks except the behavior of the monetary authority, we then obtain $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$—i.e., the "non-policy" block (2). Solving the system for every possible path of monetary policy shocks and thus of nominal interest rates, we obtain the monetary policy causal effects, Θ_ν.[14] We estimate the model to ensure consistency between Θ_ν and the short-end estimated effects of Section 3.2; details on the set of estimated parameters and on the choice of priors are provided in Appendix C.3.

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

[14] Note that the fiscal rule is actually contained in the "non-policy" block (2). Our counterfactuals thus keep the fiscal rule fixed, and only change assumptions on monetary policy conduct.

delivers sluggish responses to changing aggregates while still maintaining large MPCs out of transitory income changes. The remainder of the model is unchanged.

Behavioral frictions. 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., 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; specifically, 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 expectation 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. Our behavioral models will feature $m = 0.65$, at the lower end of the range considered by Gabaix; we have decided to make this choice (rather than estimating m, like the other model parameters) because, as we will see, our empirically estimated monetary policy shock causal effects at the short end of the yield curve, θ_ν, are actually only very weakly informative about m.

4.3 Estimation and causal effect extrapolation

We now use the empirical evidence of Section 3.2 on the propagation of short-lived monetary shocks to estimate the four models of Section 4.2, and thus extrapolate to the missing causal effects at the long end of the yield curve. We first show that all four models can—pretty much equally well—match the targeted short-end evidence, so the data do not really discriminate between them. We then show what the implied long-end extrapolation looks like.

Model estimation results. We begin in Figure 3 by repeating the two empirical monetary policy shock estimation targets from Figure 2 and then overlaying 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 four of the models are able to match the two empirical estimation targets well: both interest rate impulses lead to a hump-shaped decline in output as well as a delayed

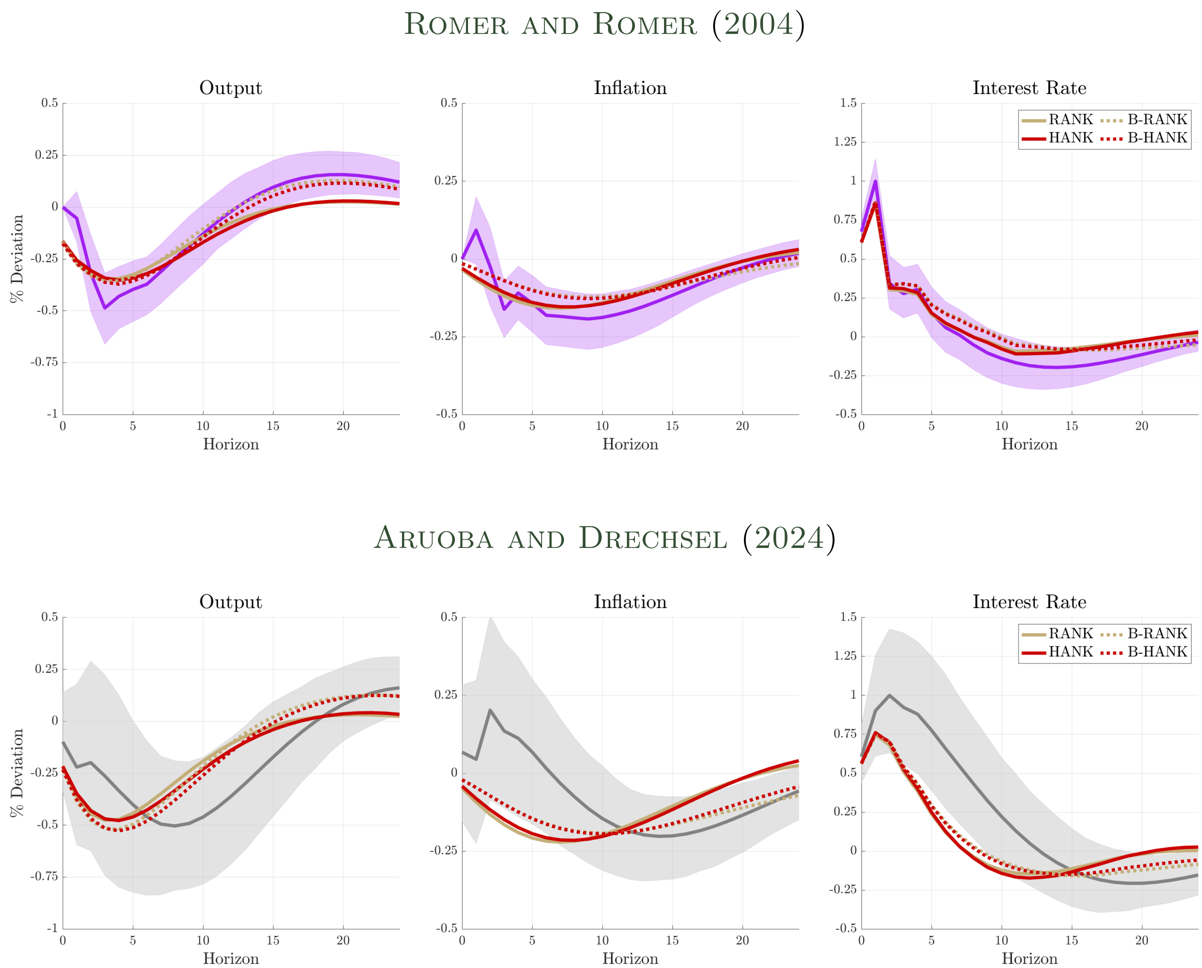

Figure 3: The purple and grey lines and areas indicate the two empirically estimated monetary shock impulse responses (see Section 3.2), with 16th and 84th percentile confidence bands. The remaining lines indicate the model-implied impulse responses at the estimated posterior modes. Beige: representative-agent consumer block. Red: heterogeneous-agent consumer block. Solid: no cognitive discounting. Dotted: cognitive discounting with $m = 0.65$.

decline in inflation, with the responses more back-loaded for the more delayed innovation of Aruoba and Drechsel (2024). This similar fit suggests that empirical evidence on monetary transmission at the short end of the yield curve does not strongly distinguish between our four models. To confirm this visual impression, we use the output of our impulse response-matching exercise to form posterior probabilities *across* models, assuming a uniform prior; detailed steps for how we do so are described in Appendix B.2. The results, displayed in Table 4.1, align with the visual impression: posterior probabilities have moved somewhat in the direction of the RANK models, but not overwhelmingly so, simply reflecting the fact

Model	Baseline	Behavioral	Total
RANK	0.4364	0.4459	0.8823
HANK	0.0693	0.0484	0.1177
Total	0.5057	0.4943	1.0000

Table 4.1: Posterior probabilities across the four models, assuming a uniform prior. The posterior model probabilities are computed as in (B.4).

that indeed all models provide very comparable fits.[15]

Monetary shock extrapolation. We now turn to the main purpose of our estimated models—extrapolation of monetary shocks to the long end of the yield curve. We will proceed in two steps: first comparing the baseline representative- and heterogeneous-agent models, and then asking how behavioral frictions change things. Results are displayed in Figures 4 and 5, which show the model-implied impulse responses to "forward guidance" shocks—i.e., nominal interest rate movements at the long end of the yield curve, much more delayed than our short-end empirical targets. Details on how we construct those particular delayed interest rate paths are provided in Appendix C.3.

- *RANK vs. HANK.* We consider a monetary policy intervention much farther out on the yield curve: a deviation from a standard monetary policy rule that is announced at $t = 0$ but occurs four years later. The right panel shows the response of nominal interest rates, while the left and middle panels display the effects of this monetary intervention on output and inflation in RANK (beige) as well as in HANK (red).

  The main takeaway is that the two models, which by construction closely agree on the effects of the two targeted interest rate paths, also at least 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 shown in Figure 4, but holds robustly for monetary shocks all along the yield curve. In other words, whether RANK or HANK are used to extrapolate from the empirically estimable short-end monetary policy shock causal effects to the missing effects at the long end does not really matter;

[15] Posterior model odds in DSGE estimation are usually very decisive, with odds greater than e^{20} being quite typical (Del Negro and Schorfheide, 2010). Here all models retain substantial posterior mass, reflecting the fact that they all provide comparable fits to our impulse response estimation targets.

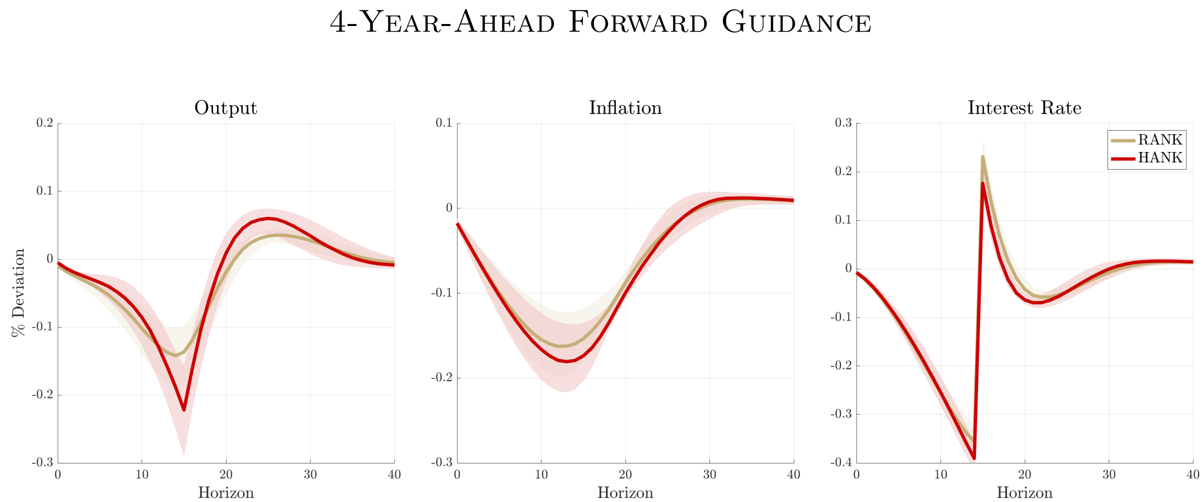

Figure 4: Policy causal effect extrapolation in the estimated RANK and HANK models. The figure shows output and inflation impulse responses to a forward guidance policy shock that leads to the interest rate movements depicted on the right. Solid line shows the posterior median. The shaded area shows the 16th and 84th percentile confidence band.

by our identification result in Proposition 1, it follows that the two estimated models will be approximately equivalent (in terms of macroeconomic outcomes) for *all possible* monetary policy counterfactuals. This discussion in particular reveals that differences in transmission channels—for RANK and HANK notably the contrast between direct vs. indirect propagation, e.g., as stressed by Kaplan et al. (2018)—may not matter at all for counterfactual aggregate outcomes.

- *Baseline vs. cognitive discounting.* Figure 5 repeats the same exercise, but now instead comparing the baseline (beige) and behavioral (blue) representative-agent models.

  We see that now there are more material differences. In particular, and just as expected given the additional cognitive discounting of the behavioral models, we see that the inflation responses in the behavioral model are more delayed, and the subsequent output overshooting is more pronounced. While Figure 5 shows this for the representative-agent models, we note that the same is also true for the heterogeneous-agent model variants. It follows that, for monetary policy counterfactuals that involve persistent interest rate changes, whether to extrapolate through models with or without behavioral frictions will matter greatly.

This completes our hybrid approach: we have used a combination of empirical estimation and model-based extrapolation to arrive at our second, more challenging "sufficient statistic":

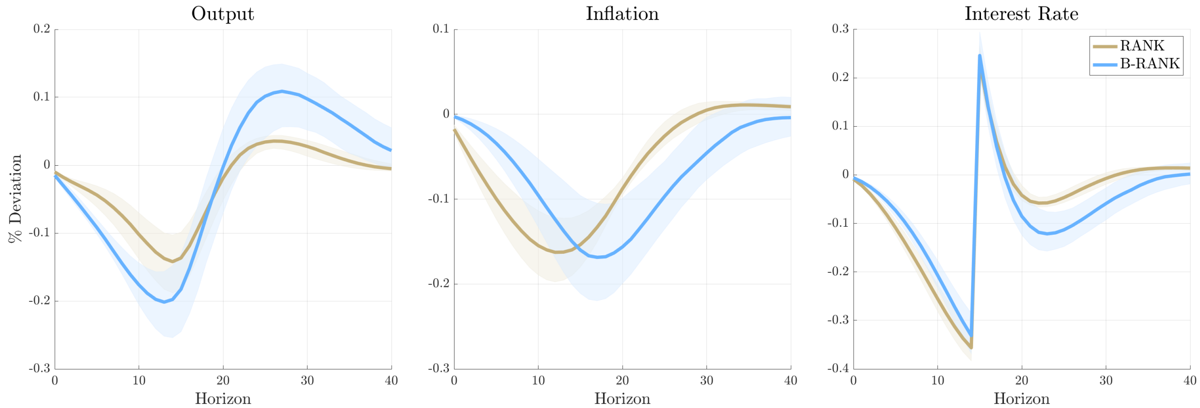

Figure 5: Policy causal effect extrapolation in the estimated RANK and behavioral-RANK (B-RANK) models. The figure shows output and inflation impulse responses to a forward guidance policy shock that leads to the interest rate movements depicted on the right. Solid line shows the posterior median. The shaded area shows the 16th and 84th percentile confidence band.

the effects of monetary policy shocks along the entire yield curve. We now have all ingredients in place to implement the identification results in Proposition 1.

5 Applications

We will leverage our hybrid approach for several applications to monetary policy evaluation. Section 5.1 describes our counterfactual assumptions on policy conduct. In Sections 5.2 to 5.4 we evaluate how such alternative policy design would have shaped the average business cycle as well as two particular historical episodes.

5.1 Policy experiment

For all three applications we will consider as our counterfactual monetary policy rule the one that minimizes the following standard central bank objective:

$$\mathcal{L} = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left\{ \lambda_\pi \pi_t^2 + \lambda_y y_t^2 + \lambda_i (i_t - i_{t-1})^2 \right\} \right]. \tag{9}$$

Formally, given the monetary policy causal effects Θ_ν, we set the counterfactual monetary policy rule coefficients $\{\tilde{\mathcal{A}}_x, \tilde{\mathcal{A}}_z\}$ in the rule (4) to yield the optimal implicit forecast targeting rule corresponding to the loss function (9), using the exact same expressions as in Proposition 2 of McKay and Wolf (2023).[16] We study this particular counterfactual exercise because it closely mirrors the popular strategy of flexible inflation targeting (see Svensson, 1997), and as such appears in much central bank communication (e.g., Federal Reserve Tealbook, 2016). Consistent with the discussion there we consider an equal-weights parameterization across the three objectives and over time, with $\lambda_\pi = \lambda_y = \lambda_i = 1$ and no discounting, i.e., $\beta = 1$.

5.2 Average business cycle

For our first application we ask how the average U.S. post-war business cycle would have differed had the Federal Reserve (always) followed the flexible inflation targeting monetary policy rule implied by (9). We communicate our main results by reporting the counterfactual volatilities of the output gap, inflation, and nominal interest rates.

Reduced-form projections. By the identification result in Proposition 1, we require the autocovariance function of the three core variables of interest, as well as other aggregates that are useful to predict them. One particularly simple and convenient way of recovering this object is to estimate a reduced-form VAR in a large vector of macroeconomic observables, and then translate this VAR to the implied Wold lag polynomial.

For our estimated reduced-form VAR we consider a set of 10 macroeconomic variables, following Angeletos et al. (2020). Differently from those authors, however, we will transform all of the included variables to stationarity (if necessary), as in Hamilton (2018); in particular, we treat the detrended real output series as a measure of the output gap. Our sample period stretches from 1960:Q1 – 2019:Q4—a long post-war, pre-covid sample. Further details on the VAR estimation and on how to translate the VAR to the Wold representation required to implement Proposition 1 are provided in Appendix D.1. We there also document that the forecasts implied by our reduced-form VAR are competitive with other forecasting strategies. Our counterfactuals are constructed at the reduced-form VAR point estimates.

Main results. Our headline finding is that the counterfactual policy would have achieved very materially lower output volatilities and somewhat more stable inflation, all at the cost of

[16] As discussed in Section 4.1, to transparently decompose the respective roles of direct empirical evidence and model-implied causal effect extrapolation, we will also report results for an approximate counterfactual that instead *only* uses the empirical evidence θ_ν. In words, this counterfactual minimizes the loss (9) within the empirically estimated space of (short-end) monetary policy shock causal effects.

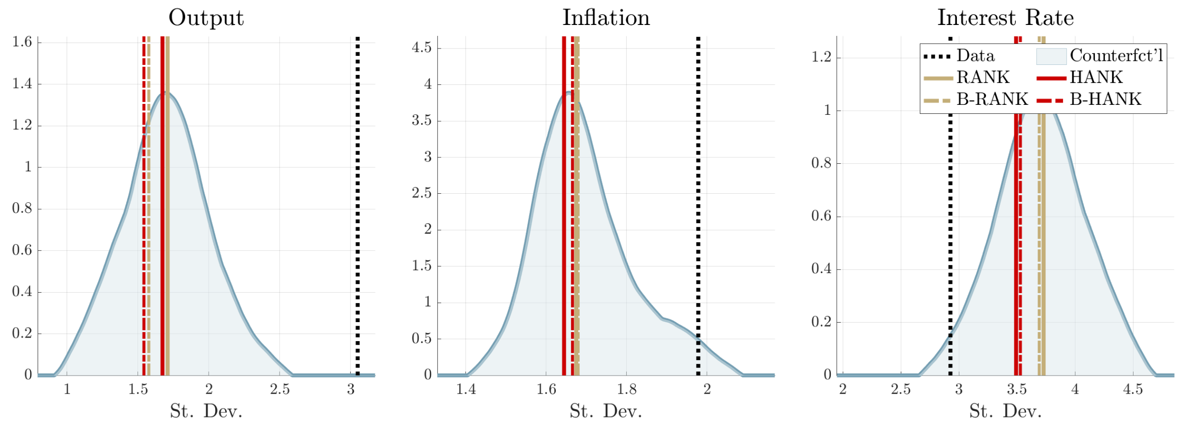

Figure 6: Counterfactual unconditional volatilities of output, inflation, and the federal funds rate, under the policy rule that minimizes (9). Black dashed: data point estimate under observed policy. Blue: posterior Kernel density of counterfactual volatilities drawing from posterior across all models and parameters. Beige: posterior mode of counterfactual using RANK models (baseline and behavioral). Red: posterior mode of counterfactual using HANK models (baseline and behavioral).

only moderately more volatile policy rates. Figure 6 shows actual (black) and counterfactual (colored) volatilities of output, inflation, and the federal funds rate. The black-dashed lines first of all depict the actual observed volatility.[17] The colored lines and shaded areas report counterfactual volatilities, which are constructed as follows. We begin by drawing from the posterior over monetary policy causal effects across the entire yield curve, i.e., Θ_ν, as implied either by each of our four estimated extrapolative models individually, or by considering all four of them jointly (with the weights of Table 4.1). For each draw we compute the volatilities under the contemplated counterfactual policy, following Proposition 1. The colored lines correspond to posterior modes for each of our four estimated models individually, while the blue shaded area is a smoothed Kernel density estimate of the entire posterior distribution when drawing jointly from all of the four models. The key finding is that, for output (and to a lesser extent for inflation), essentially all posterior mass—including in particular the four individual model modes—is concentrated far to the left of the in-sample volatility.[18]

Our second main result is that this possibility of materially lower output volatility is chiefly driven by the empirically estimated short-end causal effects of monetary policy, and less so by the model-based extrapolation to the long end of the yield curve. There are two

[17]Formally, these observed volatilities are constructed by translating the estimated reduced-form VAR into its implied Wold representation, and then from there computing the three volatilities.

[18]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 for pre-2007 results.

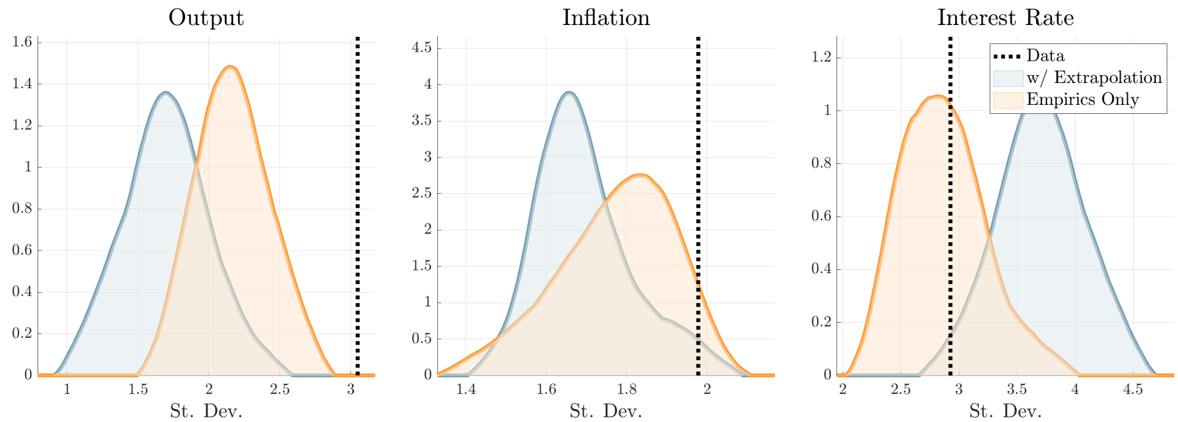

Figure 7: Counterfactual unconditional volatilities of output, inflation, and the federal funds rate, under the policy rule that minimizes (9). Black dashed: data point estimate under observed policy. Blue: posterior Kernel density of counterfactual volatilities drawing from posterior across all models and parameters. Orange: posterior Kernel density of counterfactual volatilities drawing from posterior of empirical causal effect estimates.

ways of seeing this. First, in Figure 6, the colored lines (which recall correspond to posterior modes of our four models estimated individually) are very close, even though we know that the behavioral and non-behavioral models extrapolate very differently. This suggests that the remaining posterior uncertainty (the light blue area) chiefly reflects uncertainty about the causal effects of transitory interest rate changes, rather than any across-model uncertainty in how to extrapolate those policy shock causal effects. Second, we repeat our counterfactual analysis using *only* the empirically estimated short-end monetary policy shock dynamic causal effects θ_ν, rather than the entire, model-extrapolated matrix Θ_ν; by construction, this approach of course foregoes the extrapolation altogether. Results are displayed in Figure 7, with the shaded blue areas exactly as in Figure 6, while the shaded orange areas show the restricted empirics-only counterfactual. We see that changes in monetary policy conduct at the short end of the yield curve do already suffice to attain material reductions in output volatility, revealing that the role of model extrapolation for this particular counterfactual is, as claimed, relatively limited.

Inspecting the mechanism. To see more clearly where those sharp results are actually coming from, we now zoom in further and study the counterfactual propagation of a particular linear combination of reduced-form Wold residuals that Angeletos et al. (2020) have labelled the "main business-cycle shock"—i.e., the reduced-form shock that "explains" the largest share of short-term volatility in real outcomes. The black-dashed lines in Figure 8

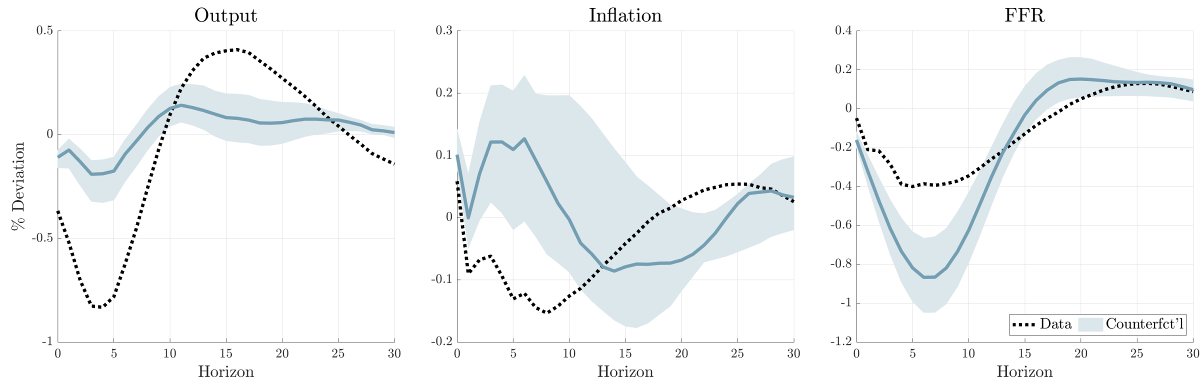

Figure 8: Counterfactual impulse responses of output, inflation, and the federal funds rate to the main business-cycle shock (see Angeletos et al., 2020), under the policy rule that minimizes (9). Black dashed: data point estimate under observed policy. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded), drawing from posterior across all models and parameters.

show the propagation of this shock under the in-sample monetary policy reaction: inflation drops just a little, output drops materially, and monetary policy somewhat leans against this contraction, by construction echoing the results reported by Angeletos et al.. Our contemplated counterfactual policy leans against this shock much more, stabilizing output at the cost of moderately higher inflation and larger nominal interest rate movements.[19]

The takeaway from Figure 8 is that our headline conclusions in Figure 6 are essentially driven by just two moments of the data: first, that "typical" output movements in the data are associated with moderate inflation movements and partial interest rate offset; and second, that interest rate cuts boost output, with only little effect on inflation. Combining those two empirical moments—which are already well-documented from much prior work (notably Ramey, 2016; Angeletos et al., 2020)—with our identification results in Section 2 immediately delivers the conclusions in Figure 6; in other words, and as also visible in Figure 7, the key for everything reported here are the empirically estimable short-end causal effects (together with reduced-form projections), with little incremental role for model-based extrapolation to the long end of the yield curve.[20] This first counterfactual thus illustrates our first main methodological takeaway—that substantively interesting monetary policy counterfactual results can already be established straight from empirical evidence, leveraging only the theory contained in the identification result of Proposition 1.

[19]This exercise is an informative diagnostic because it reduces the 10-dimensional Wold representation of the data to a 1-dimensional slicing that explains most output volatility, allowing easy graphical analysis.

[20]Model-based extrapolation simply allows the volatilities to be reduced even further by better tailoring the dynamic path of the interest rate response to the output fluctuations that it is designed to offset.

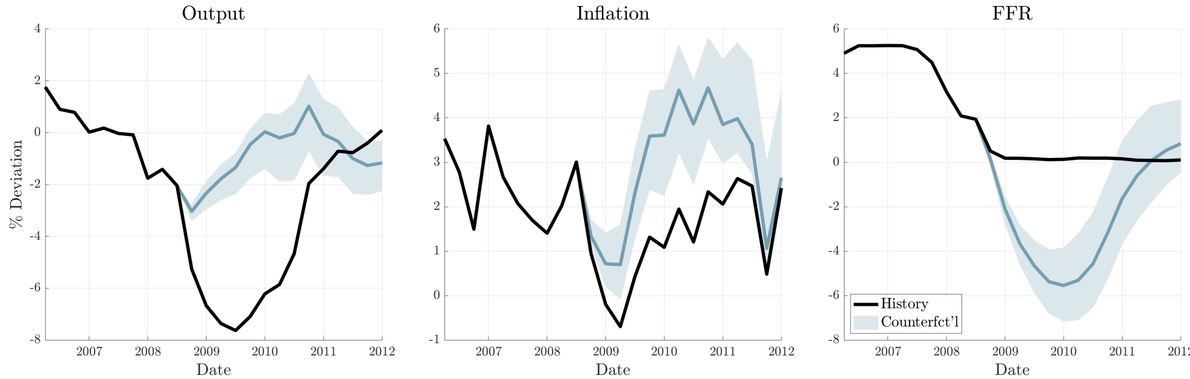

Figure 9: Counterfactual evolution of output, inflation, and the federal funds rate in the Great Recession, under the policy rule that minimizes (9) without any effective lower bound on rates. Black: data. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded), drawing from posterior across all models and parameters.

5.3 Great Recession

For our second application we evaluate how the economy would have evolved during the Great Recession if monetary policy had followed the inflation targeting framework described above, and *without* any effective lower bound on nominal interest rates.[21] Specifically, we will assume that the central bank follows this alternative, unconstrained rule from 2008:Q4 onwards, and does so throughout 2012:Q1—i.e., an example of a "historical evolution" counterfactual. A counterfactual of this sort is informative about the plausible costs of a binding lower bound constraint. We first discuss how we construct the required reduced-form forecasting inputs, and then present the main results.

Reduced-form projections. We here require forecasts of output, inflation, and nominal interest rates at each in-sample date, for the entire period under study. We construct these forecasts using the same 10-variable reduced-form VAR as in Section 5.2.

Main results. The headline finding is that, absent any effective binding lower bound on nominal interest rates, a policy that follows the rule of minimizing (9) would have involved a very aggressive rate cut, down to around -5 per cent. Such an (infeasible) interest rate cut would have materially reduced the output gap, at the cost of moderately elevated inflation.

[21] We note that our methodology remains applicable to model environments with a linear non-policy block (2) 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).

The results are summarized in Figure 9, which shows realized (black) as well as counterfactual (blue) paths of output, inflation, and interest rates; as before, the blue areas correspond to the posterior across all four of our models, with results for all individual models very similar. Our findings are informative about the broader policy response during the Great Recession. Given constraints on nominal interest rates, policymakers attempted to substitute through other stimulative measures, most notably unconventional monetary policy as well as fiscal stimulus. If we interpret (9) 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 500 basis points would have been necessary.

We also note that, just as in our first counterfactual, these takeaways are largely governed by the available evidence on short-end monetary policy shock propagation, and less so by the model-based causal effect extrapolation; see Appendix D.3 for further details. As before, the simple intuition is that the economy can already be meaningfully stabilized through changes in monetary policy conduct at the short end of the yield curve. This application thus provides a second illustration of our observation that, for monetary policy counterfactual analysis, the available empirical evidence alone often largely pins down the required "sufficient statistics," implying that the conclusions of our hybrid method will be governed by the data part, and not the extrapolation part.

5.4 Post-covid inflation

As the third and final application we evaluate monetary policy options at the height of the post-covid inflation. Specifically, looking from 2021:Q2 onwards, we ask what counterfactual conduct of monetary policy would have been expected to minimize the central bank objective (9). This is thus an example of the "conditional forecasts" counterfactual type.[22]

Reduced-form projections. We require forecasts of output, inflation, and nominal interest rates only for the date of interest, 2021:Q2. We construct these forecasts using exactly the same 10-variable VAR as in the previous two applications; differently from those, however, we extend the sample to the chosen forecast date.

Main results. Figure 10 shows the actual historical evolution of output, inflation, and the federal funds rate (black), their VAR-implied forecasts from 2021:Q2 onwards (grey-dashed),

[22] A similar counterfactual question is studied in Bocola et al. (2024). Differently from the present analysis, however, they rely on a standard full-information likelihood-based approach.

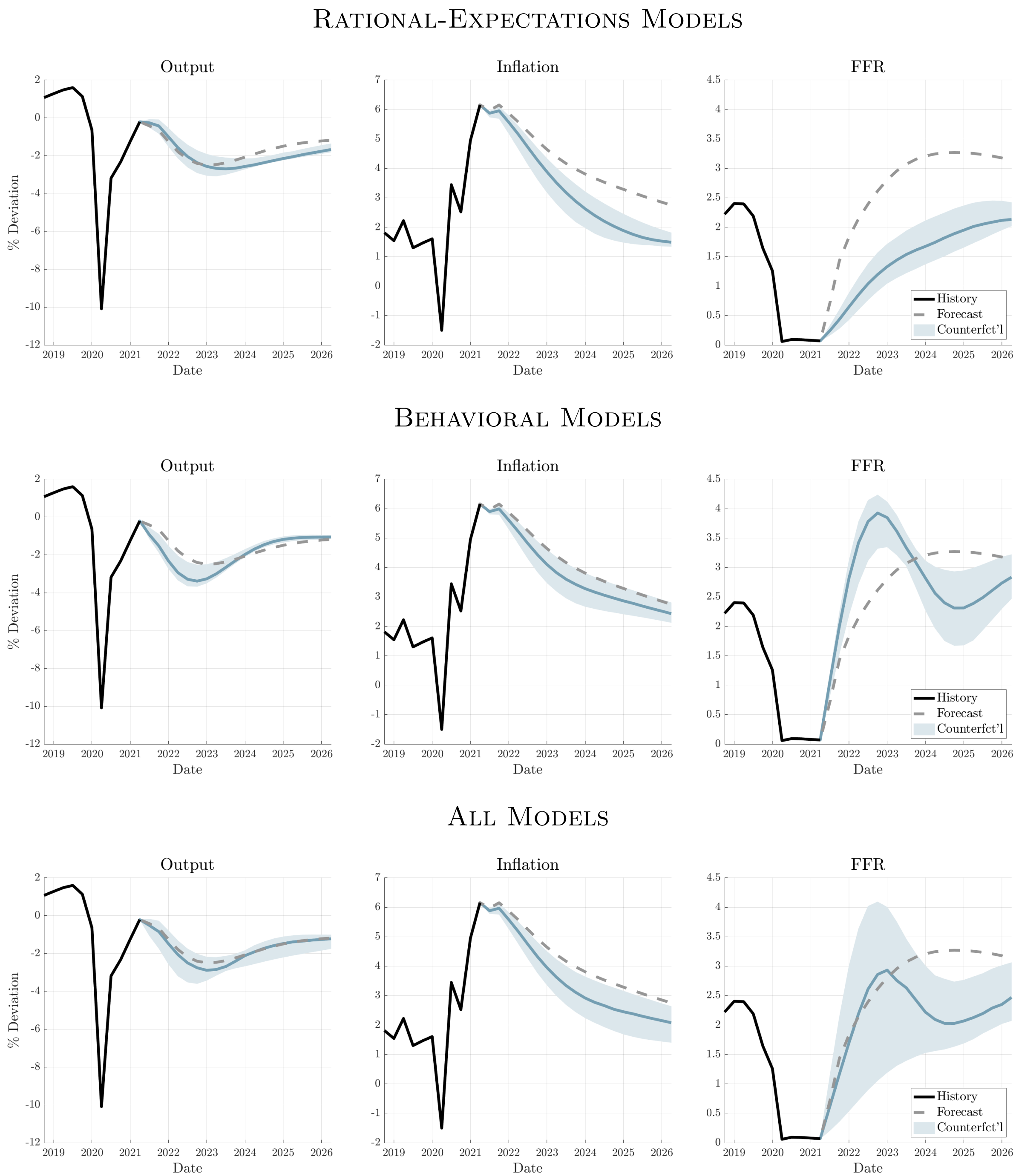

Figure 10: Counterfactual projections of output, inflation, and the federal funds rate in the post-covid inflationary episode (from 2021:Q2), under the policy rule that minimizes (9). Policy causal effects from rational-expectations models (top), behavioral models (middle), and all models jointly (bottom). Black: data. Grey: actual (VAR-implied) forecast. Blue: posterior median (solid) and 16th and 84th percentile bands (shaded).

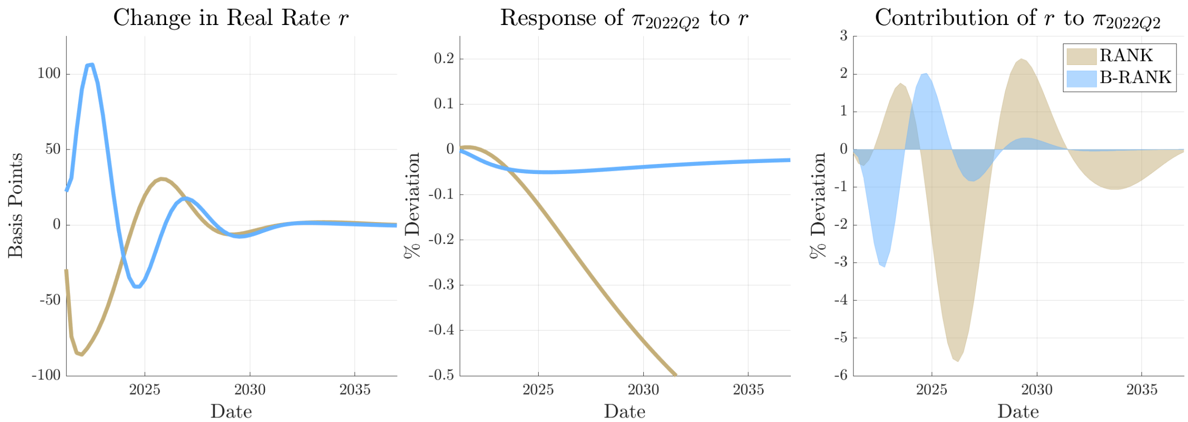

Figure 11: Behavior of real interest rates under the counterfactual monetary policy. Left panel: difference between the counterfactual and the forecast real interest rate path. Middle panel: change in inflation at 2022:Q2 in response to a policy-induced real interest rate change by horizon of the rate 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.

and their optimal counterfactual forecasts (blue). The panels distinguish the various models used to extrapolate monetary policy causal effects: first the two rational-expectations models (top), then the two behavioral models (middle), and finally the full set (bottom). Under the baseline forecast, inflation is expected to be persistently elevated, output is slightly depressed, and interest rates rise sharply. Our focus is now on how the optimal inflation targeting monetary policy moves the economy away from these baseline forecasts. Importantly, since inflation is expected to be persistently elevated, the effects of monetary policy further out on the yield curve—and thus model-implied causal effect extrapolation—are likely to matter greatly here, unlike our first two applications.

Consider first the top panel of Figure 10, which shows counterfactuals constructed with policy shock extrapolation via our two rational-expectations models. We see that policy here succeeds in reducing inflation sharply with only a small reduction in output. Furthermore, and counterintuitively, 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 depressed output 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 11 help 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 10, which constructs our policy counterfactuals using instead the behavioral models to extrapolate monetary policy effects. The counterfactual now looks very different: the federal funds rate here is hiked somewhat *more* aggressively than in the baseline forecast, thus bringing inflation down slightly, though at the cost of moderately lower output. Intuitively, the policymaker now cannot rely on 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 she chooses to respond to it through *higher* short-term real interest rates, simply because short-run inflation is more elevated than output is depressed. A visual illustration is provided with the blue lines in Figure 11: 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 relatively short-term expected real rate movements.

Finally, the bottom panel of Figure 10 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 response of interest rates. The four models are similar in their abilities to fit our estimation targets—i.e., the effects of a transitory monetary shock—yet they differ in their predictions for the effects of far-ahead rate changes. As this counterfactual features persistent interest rate changes, the disagreement across models translates to considerable uncertainty regarding the desired path of output and inflation.

Discussion. The application of this section is a clear example of a monetary policy counterfactual question for which the available empirical evidence on monetary shock propagation is insufficient. And not only is that evidence too limited to allow evaluation of the counterfactual of interest, it is in fact even insufficient to discriminate between different models of monetary transmission—here those with and without behavioral frictions—that extrapolate very differently to the missing long end of the yield curve. The concluding section will discuss implications of these observations for future empirical and theoretical work.

6 Conclusions

How much explicit model structure is needed to evaluate the effects of hypothetical changes in monetary policy conduct? How far can we get using empirical evidence? Our analysis in this paper suggests two main takeaways that we hope will shape future applied work.

1. Empirical evidence in conjunction with the identification results of the recent sufficient statistics literature (Barnichon and Mesters, 2023; McKay and Wolf, 2023) is often already enough to arrive at sharp conclusions. Intuitively, what matters most for many monetary policy counterfactuals are the effects of changes in policy conduct in the short run, say over the next few quarters. The effects on the macro-economy of such policy changes are already well-studied in existing empirical work.

2. The *only* ingredient for counterfactual monetary policy evaluation that we cannot directly get from the data are the effects of medium- to long-run changes in monetary policy. Such extrapolation of causal effects along the yield curve—and not providing a full structural account of all cyclical fluctuations—is the *sole* remaining purpose of model structure.

We have provided a new methodology—our hybrid approach—that reflects these insights, echoing earlier analogous advances in public economics (see Chetty, 2009) and bringing them to dynamic macroeconomic policy evaluation. The result are counterfactuals that are more transparent and empirically credible than those of the standard full-information approaches, with a clear separation of the roles played by empirical evidence on one side, and by structural assumptions on the other.

The two observations above have implications for future empirical and structural work. Empirically, a key task will be to estimate the dynamic causal effects of delayed or persistent changes in monetary policy. Doing so would expand even further the range of counterfactual questions that could be answered directly from empirical evidence. Theoretically, more attention should be paid to how different plausible models of monetary policy transmission extrapolate to the causal effects of far-ahead policy changes. In particular, discriminating between models with and without behavioral frictions appears much more important than the extent of household market incompleteness. In the meantime, to further ongoing applied work, we have made our empirically estimated and model-extrapolated monetary policy causal effects publicly available.[23] Future applied work on monetary policy evaluation can combine these effects with simple reduced-form forecasts to evaluate the likely implications of any given contemplated change in monetary policy conduct.

[23] All replication codes and in particular our monetary policy shock causal effect matrices Θ_ν are available here: `https://github.com/tcaravello/mp_modelcnfctls`.

References

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

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 (2024): "Identifying Monetary Policy Shocks: A Natural Language Approach," CEPR Discussion Paper No. DP17133.

Auclert, A. (2019): "Monetary policy and the redistribution channel," *American Economic Review*, 109, 2333–2367.

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.

Bocola, L., A. Dovis, K. Jørgensen, and R. Kirpalani (2024): "Bond market views of the fed," Tech. rep., National Bureau of Economic Research.

Brockwell, P. J. and R. A. Davis (1991): *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, 2nd ed.

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

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.

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

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.

——— (2010): "Bayesian econometrics," .

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.

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.

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.

Jarociński, M. and P. Karadi (2020): "Deconstructing monetary policy surprises—the role of information shocks," *American Economic Journal: Macroeconomics*, 12, 1–43.

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.

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

Lucas, Robert E., J. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4, 103–124.

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

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.

Miranda-Agrippino, S. and G. Ricco (2023): "Identification with external instruments in structural VARs," *Journal of Monetary Economics*.

Molavi, P. (2019): "Macroeconomics with learning and misspecification: A general theory and applications," *Unpublished manuscript*.

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.

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

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.

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

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

Stock, J. H. and M. W. Watson (2016): "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," in *Handbook of macroeconomics*, Elsevier, vol. 2, 415–525.

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

Swanson, E. (2024): "The Macroeconomic Effects of the Federal Reserve's Conventional and Unconventional Monetary Policies," Working Paper.

Werning, I. (2015): "Incomplete markets and aggregate demand," Tech. rep., National Bureau of Economic Research.

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.

——— (2024): "Interest rate cuts vs. stimulus payments: An equivalence result," Working Paper.

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

Online Appendix for:
Evaluating Monetary Policy Counterfactuals:
(When) Do We Need Structural Models?

This appendix contains supplemental material for the article "Evaluating Monetary Policy Counterfactuals: (When) Do We Need Structural Models?." We provide: (i) some additional theoretical results to complement the discussion in Section 2; (ii) implementation details and further comments on the robustness benefits of our hybrid strategy; (iii) further details for our empirical and model-based analysis of monetary shock propagation in Sections 3 and 4; and (iv) supplementary results to complement the applications in Section 5.

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

Contents

A Supplementary details on the identification result

This appendix provides several supplementary theoretical results. Appendix A.1 gives some missing details for the description of the general environment in Section 2, Appendix A.2 contains the proof of Proposition 1, Appendix A.3 elaborates on the role played by invertibility in our identification results, and finally Appendix A.4 discusses how the identification results apply in environments with behavioral frictions.

A.1 More on the general environment

This section provides some supplementary details on our general model set-up in Section 2.1. This is necessary for a precise definition of our (conditional) policy counterfactuals of interest, as well as for the proofs in Appendix A.2.

We first note that, under our assumptions on (1), the autocovariance function $\Gamma_y(\bullet)$ of macroeconomic observables y_t exists and by standard arguments is given as

$$\Gamma_y(\ell) = \sum_{m=0}^{\infty} \Theta_m \Theta'_{m+\ell}. \tag{A.1}$$

It is straightforward to use (A.1) to derive our first sufficient statistic, the Wold representation (6); see, e.g., Brockwell and Davis (1991).

We next provide some missing details for the definition of the conditional counterfactuals. Note that the counterfactual SVMA (5) embeds the assumption that the counterfactual rule (4) is actually followed *forever*. In our conditional counterfactuals, however, we will assume that the policymaker instead unexpectedly changes to the alternative rule (4) at some date t^*, having followed the original rule (3) up to $t^* - 1$. In that case we will have

$$\tilde{y}_t \;=\; \underbrace{\sum_{\ell=0}^{t-t^*} \tilde{\Theta}_\ell \varepsilon_{t-\ell}}_{\text{new shocks after } t^*} \;+\; \underbrace{\tilde{y}_t^*}_{\text{initial conditions}} \tag{A.2}$$

The first term in (A.2) 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 (4). 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

$$\mathcal{H}_w(\tilde{\boldsymbol{w}}^* - \boldsymbol{w}^*) + \mathcal{H}_x(\tilde{\boldsymbol{x}}^* - \boldsymbol{x}^*) + \mathcal{H}_z(\tilde{\boldsymbol{z}}^* - \boldsymbol{z}^*) \quad = \quad \mathbf{0}, \tag{A.3}$$

$$\tilde{\mathcal{A}}_x\tilde{\boldsymbol{x}}^* + \mathcal{A}_z\tilde{\boldsymbol{z}}^* \quad = \quad \mathbf{0}, \tag{A.4}$$

i.e., a system written in terms of forecast revisions.[24] This provides the required definitions of the additional terms appearing in our definitions of the conditional counterfactuals.

A.2 Proof of Proposition 1

Consider using the policy transmission map Θ_ν to predict the counterfactual propagation of the Wold innovations u_t under the counterfactual policy rule (4), 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

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

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

$$\tilde{A}_x\left(\Psi_{\bullet,x,j} + \Theta_{x,\nu}\tilde{\boldsymbol{\nu}}_j\right) + \tilde{A}_z\left(\Psi_{\bullet,z,j} + \Theta_{z,\nu}\tilde{\boldsymbol{\nu}}_j\right) = \mathbf{0}. \tag{A.5}$$

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

$$\tilde{y}_t = \sum_{\ell=0}^{\infty} \tilde{\Psi}_\ell u_{t-\ell}. \tag{A.6}$$

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

$$u_t = P\varepsilon_t,$$

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

$$\tilde{\Psi}(L) = \tilde{\Theta}(L)P'. \tag{A.7}$$

[24] Here, boldface denotes sequences from t^* onwards. (A.4) says the new, counterfactual policy rule holds. By (A.3), the revised forecasts remain consistent with all private-sector relationships.

We now recover each of the desired counterfactuals.

1. Consider using the counterfactual process (A.6) 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 (A.7), and the second follows since P is an orthogonal matrix.

2. Applying Proposition 1 of McKay and Wolf (2023) to the system (A.3) - (A.4) that defines initial conditions $\tilde{\boldsymbol{y}}^*$, we see that we can recover initial conditions at t^* as

$$\tilde{\boldsymbol{y}}^* = \boldsymbol{y}^* + \Theta_\nu \tilde{\boldsymbol{\nu}}^*, \tag{A.8}$$

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

$$\tilde{A}_x \left(\boldsymbol{x}^* + \Theta_{x,\nu}\tilde{\boldsymbol{\nu}}^*\right) + \tilde{A}_z \left(\boldsymbol{z}^* + \Theta_{z,\nu}\tilde{\boldsymbol{\nu}}^*\right) = \mathbf{0}. \tag{A.9}$$

   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 (6), 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 (A.6) and (A.8) to recover the conditional forecast $\mathbb{E}_t\left[\tilde{y}_{t+h}\right]$. We have

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

   (ii) Consider using (A.6) and (A.8) to recover the historical counterfactual $\tilde{y}_t$. We have

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

□

A.3 More on the role of invertibility

We begin by establishing, consistent with the intuition given throughout Section 2, that the sole purpose of invertibility is to generate full-information forecasts. Afterwards we provide some further discussion of what happens in the absence of invertibility. We also give some missing implementation details for the numerical explorations in Section 2.3.

Invertibility and forecasts. We provide a constructive argument showing that access to full-information forecasts is sufficient to recover our counterfactuals. For this it will prove convenient to reverse the order relative to the arguments in Proposition 1, beginning instead with counterfactuals for conditional episodes.

2. **Conditional episodes.**

   (i) *Conditional forecasts.* Recall that we need to construct $\tilde{y}^*_{t^*+h}$ and $\tilde{\Theta}_h \varepsilon_{t^*}$. Given the full-information forecasts $\mathbb{E}_{t^*-1}[y_{t+h}]$, $\tilde{y}^*_{t^*+h}$ can be constructed from Θ_ν exactly as in the proof of Proposition 1. Next note that $\Theta_h \varepsilon_{t^*}$ can be recovered as the *revision* in full-information forecasts

   $$\Theta_h \varepsilon_{t^*} = (\mathbb{E}_{t^*} - \mathbb{E}_{t^*-1})[y_{t^*+h}].$$

   We can then just as before use Θ_ν to turn those expectation revisions into $\tilde{\Theta}_h \varepsilon_{t^*}$, completing the argument.

   (ii) *Historical evolution.* We now need to construct $\tilde{y}^1_t$ as well $\sum_{\ell=0}^{t-t_1} \tilde{\Theta}_\ell \varepsilon_{t-\ell}$. As in the previous item, $\tilde{y}^1_t$ can still be computed directly from date-$t_1 - 1$ forecasts, as in the proof of Proposition 1. Next, for date t_1, we obtain $\Theta_h \varepsilon_{t_1}$ from forecast revisions as $(\mathbb{E}_{t_1} - \mathbb{E}_{t_1-1})[y_{t_1+h}]$, and then use Θ_ν to get the counterfactual $\tilde{\Theta}_h \varepsilon_{t_1}$, thus in particular giving $\tilde{y}_{t_1} = \tilde{\Theta}_0 \varepsilon_{t_1} + \tilde{y}^1_t$. Proceeding recursively, we for time $\check{t}$ obtain forecast revisions to get $\Theta_h \varepsilon_{\check{t}}$, and so as usual via Θ_ν recover $\tilde{\Theta}_h \varepsilon_{\check{t}}$. From here we then get the date-$\check{t}$ realized counterfactual outcome as

   $$\tilde{y}_{\check{t}} = \tilde{\Theta}_0 \varepsilon_{\check{t}} + \underbrace{\sum_{\ell=1}^{\check{t}-t_1} \tilde{\Theta}_\ell \varepsilon_{\check{t}-\ell} + \tilde{y}^1_t}_{\text{from previous steps}},$$

   completing the argument.

1. **Unconditional business cycles.** Proposition 1 presupposes knowledge of the autocovariance function of the observables y_t or, equivalently, access to an arbitrarily large sample of observations $\{y_t\}_{t=0}^{\infty}$. By the discussion in the previous item, knowledge of full-information forecasts suffices to instead construct an arbitrarily large counterfactual sample $\{\tilde{y}_t\}_{t=0}^{\infty}$, thus delivering the counterfactual autocovariance function $\tilde{\Gamma}(\ell)$.

From this discussion it follows that, conditional on full-information forecasts being observable, invertibility ceases to be necessary. Our empirical implementation of the VAR step is designed with this observation in mind.

PROPOSITION 1 WITHOUT INVERTIBILITY. Without invertibility, the orthogonalized reduced-form residuals u_t satisfy (e.g., see Wolf, 2020)

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

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 1 satisfies

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

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

The results in Section 2.3 furthermore reveal that it is not just our particular proof *strategy* that fails here—without invertibility, Wold-implied forecasts are generally not equal to full-information forecasts, and so the derived counterfactuals do not equal the truth (though they may be close, of course). Mathematically, the problem is that, while the true lag polynomial $\Theta(L)$ and the Wold lag polynomial $\Psi(L)$ generate the same autocovariance function, nothing guarantees that the counterfactual lag polynomials $\tilde{\Theta}(L)$ and $\tilde{\Psi}(L)$ will also generate the same second moments. It is only the assumption of invertibility—which ties the impulse responses in the lag polynomials $\Theta(L)$ and $\Psi(L)$ together in a particular way—that allows this argument to go through.

Numerical explorations without invertibility. Our laboratory data-generating process for the illustrations in Section 2.3 is the well-known structural model of Smets and Wouters (2007), but with one minor change—we assume that the monetary authority follows rules of the form

$$i_t = \rho_i i_{t-1} + (1 - \rho_i)(\phi_\pi \pi_t + \phi_y y_t), \tag{A.10}$$

which is slightly simpler than the headline specification considered by Smets and Wouters.

Specifically, we assume that the researcher observes data generated from the posterior mode parameterization of the Smets and Wouters model, but with the monetary policy rule taking the particular form (A.10) with $\phi_\pi = 1.5$ and $\rho_i = \phi_y = 0$. She then wishes to predict the counterfactual second-moment properties of interest if instead the monetary authority followed the rule (A.10) but with $\rho_i = 0.8$, $\phi_\pi = 1.5$ and $\phi_y = 0.5$. We have chosen these two particular policy rules because they imply quite starkly different second-moment properties, with the first one aggressively stabilizing inflation, while the second one smoothes interest rates and also stabilizes output. This allows us to most transparently illustrate our results about counterfactual accuracy under non-invertibility, as displayed in Figure 1.

A.4 Behavioral models

In this subsection we discuss to what extent our theoretical identification results are consistent with deviations from the usual full-information, rational-expectations (FIRE) benchmark. We first clarify what kinds of behavioral frictions are admissable (and which are not), and then explain why, to implement our methodology, researchers still always need to try to construct *full-information* forecasts, consistent with our main-text discussion.

Nested behavioral models. Every structural environment that can be written in the general form (2) - (3) is consistent with our identification results. Importantly, this contains models with behavioral frictions in which the deviation from FIRE is *independent of the policy rule*, in the sense that the behavioral friction is encoded in $\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z, \mathcal{H}_e$, and does not change as the policy rule changes.

More formally, begin by considering a model with FIRE, and consider the ith equation in its non-policy block, written as

$$\mathcal{H}^R_{i,w}\boldsymbol{w} + \mathcal{H}^R_{i,x}\boldsymbol{x} + \mathcal{H}^R_{i,z}\boldsymbol{z} + \mathcal{H}^R_{i,e}e_0 = \mathbf{0}. \tag{A.11}$$

A typical example of such a block would be the aggregate consumption function, mapping sequences of household income and asset returns into a path of consumption. Our theory is consistent with behavioral frictions in which the model equation (A.11) is replaced by an alternative of the form

$$\mathcal{H}^B_{i,w}\boldsymbol{w} + \mathcal{H}^B_{i,x}\boldsymbol{x} + \mathcal{H}^B_{i,z}\boldsymbol{z} + \mathcal{H}^B_{i,e}e_0 = \mathbf{0}. \tag{A.12}$$

where the matrices $\mathcal{H}^B_{i,\bullet}$ are a *policy rule-invariant transformation* of $\mathcal{H}^R_{i,\bullet}$:

$$\mathcal{H}^B_{i,\bullet}(\theta) = f(\mathcal{H}^R_{i,\bullet}, \theta),$$

with the parameter vector θ governing the behavioral friction. Examples of behavioral frictions that can be written in this general way include sticky information, sticky expectations, cognitive discounting, level-k thinking, and diagnostic expectations; see Auclert et al. (2021) for further details. Crucially, in all of these cases, agent behavior continues to be shaped by policy only through the current and expected future (full-information) paths of $\{\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{z}\}$, and so our identification results continue to apply.[25]

Implications for forecasting. The previous discussion reveals that, even if the underlying data-generating process features behavioral frictions (of the sort consistent with our identification result, of course), the forecasts that appear in (2) - (3) and thus our identification results are *full-information* forecasts. It follows that, when leveraging our identification result, researchers should aim to construct such full-information forecasts, and then note that their reported counterfactuals will be valid across a wide range of models with and without underlying behavioral frictions.

[25] A simple example that violates this restriction is the misspecified learning model of Molavi (2019). Here, a change in policy rule affects agent learning and thus alters the response to any given set of (full-information) expected future time paths, breaking the policy rule independence that we require. Mathematically, this is isomorphic to how incomplete information as in Lucas (1972) breaks our identification results.

B The hybrid strategy

Appendices B.1 and B.2 provide some further details for our proposed hybrid strategy: first on impulse response estimation, and then on model estimation. Appendix B.3 elaborates on some of the well-known fragilities of standard full-information likelihood-based approaches.

B.1 Impulse response estimation

The first step of the impulse response-matching approach is the estimation of the monetary shock impulse response targets. We assume access to n_ν distinct policy shocks, and that the subsequent model estimation targets the impulse responses of n_m outcome variables over H impulse response horizons to the identified shocks. We stack these impulse responses in the $n_\nu \times n_m \times H$ vector θ_ν.

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

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

As in much prior work on impulse response-matching, we propose to estimate θ_ν using standard Bayesian VAR methods. Estimation delivers draws $i = 1, 2, \ldots, N$ of the policy shock causal effect vector, denoted $\hat{\theta}_{i,\nu}$. 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

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

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

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

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 $\bar{H}_T = 8$.[26] 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.2 Model estimation

We next describe how we use the empirically estimated impulse response targets for structural model estimation. Relative to standard implementations of impulse response-matching, our approach differs in two respects. First, we allow for joint estimation of multiple models. This is important because in our applications we want to study whether the data can discriminate between different models of monetary policy transmission. Second, our approach does not require the researcher to specify a policy rule, thus further increasing the robustness of our approach. The remainder of this section provides implementation details, focusing in particular on these two novelties.

We consider the joint estimation of a list $\mathcal{M}$ of models of policy transmission, denoted by

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

$\mathcal{M}_j$ for $j = 1, 2, \ldots, M$. In the notation of Section 2.1, a "model" is a tuple $\{\mathcal{H}_w, \mathcal{H}_x, \mathcal{H}_z\}$. Each model then 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, as well as a prior probability $p(\mathcal{M}_j)$. We use impulse response-matching strategies to arrive at posterior distributions for parameters given models, and across models. We describe our approach in two steps. First, for a given model $\mathcal{M}_j$ and parameter vector ψ_j, we explain how to obtain $\theta_\nu(\psi_j, \mathcal{M}_j)$, the model-implied analogue of the empirically observed policy shock causal effect vector. This step is somewhat 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 instead entirely standard, so we will be brief.

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

1. Given $(\psi_j, \mathcal{M}_j)$, solve for impulse responses of the targeted outcome variables 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 McKay and Wolf (2023), the choice of that 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.[27]

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 $n = 1, \ldots, n_\nu$, 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 $\tilde{\nu} = [\tilde{\boldsymbol{\nu}}_1', \ldots, \tilde{\boldsymbol{\nu}}_{n_\nu}']'$. 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

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

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

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 $H+1, H+2, \ldots, T$ of vector $\tilde{\boldsymbol{\nu}}_n$.[28] 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).

3. With $\tilde{\nu}^*$ in 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.[29]

Posterior distribution & marginal likelihood. Given the above strategy to evaluate $\theta_\nu(\psi_j, \mathcal{M}_j)$, we can now estimate posteriors for models and model parameters using a standard limited-information Bayesian estimation approach. We can define an approximate likelihood of the "data," $\hat{\theta}_\nu$, as a function of ψ_j given $\mathcal{M}_j$:

$$p(\hat{\theta}_\nu \mid \psi_j, \mathcal{M}_j) \propto \exp\left[-0.5\left(\hat{\theta}_\nu - \theta_\nu(\psi_j, \mathcal{M}_j)\right)' V_{\theta_\nu}^{-1}\left(\hat{\theta}_\nu - \theta_\nu(\psi_j, \mathcal{M}_j)\right)\right]. \tag{B.3}$$

Combining the prior together with the likelihood (B.3), we obtain the posterior for ψ_j conditional on model $\mathcal{M}_j$ and given the policy shock causal effect data $\hat{\theta}_\nu$:

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

[28] 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.

[29] 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, consistent with the exact population result.

and where

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

is the marginal density of $\hat{\theta}_\nu$ given model $\mathcal{M}_j$. The final step is to recover posterior model probabilities—i.e., the posterior distribution across the model space $\mathcal{M}$. We have

$$p(\mathcal{M}_j \mid \hat{\theta}_\nu) = \frac{p(\hat{\theta}_\nu \mid \mathcal{M}_j) p(\mathcal{M}_j)}{\sum_{i=1}^{M} p(\hat{\theta}_\nu \mid \mathcal{M}_i) p(\mathcal{M}_i)}. \tag{B.4}$$

To actually compute these objects 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.[30] 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 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 our posterior distribution. We set $N_a = N_c = 100000$, $N_b = 50000$, $c_1 = 0.8$ and $c_2 = 0.6$ for all models. Our acceptance rates for all of the models considered range between 20 and 30 percent. We store $N_d = 1000$ draws, selected as 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).[31]

We have thus overall arrived at 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

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

In order to actually implement our applications, we need to store these large transmission maps for each draw from the posterior; i.e., we need to store 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, with $T_u = 200$.

[30]For our HANK models, we in this step use a standard deviation of 0.1 for the informational stickiness parameter (instead of 0.2, see Table C.2), to avoid getting too many draws outside of the parameter support.

[31]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.

B.3 Vulnerabilities of full-information structural approaches

We here provide some further details supplementing the discussion in Section 4.1, elaborating on some well-known vulnerabilities of the standard full-information approach.

Model misspecification and inference. Under standard full-information approaches to model estimation (like, e.g., Smets and Wouters, 2007), misspecification 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 misspecification in shock processes, as such misspecification is particularly likely in practice (Chari et al., 2009). Analogous arguments apply to misspecification in policy rules.

Suppose the true data-generating process is

$$y_t = \Theta^*(L)\xi_t \tag{B.5}$$

$$\xi_t = B^*(L)\varepsilon_t \tag{B.6}$$

where $\varepsilon_t \sim N(0, I)$. Relative to (1), the system (B.5) - (B.6) is written to explicitly separate the exogenous process (i.e., equation (B.6)) from the endogenous model propagation (i.e., equation (B.5)). 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 $\Psi^*(L) = \Theta^*(L)B^*(L)$.

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

$$y_t = \Theta_{\psi_1}(L)\xi_t \tag{B.7}$$

$$\xi_t = B_{\psi_2}(L)\varepsilon_t \tag{B.8}$$

where again $\varepsilon_t \sim N(0, I)$. We assume that there is no misspecification 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 misspecified; 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

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

Thus, when evaluated at $\psi^\dagger$ (and only then), the two processes (B.5) - (B.6) and (B.7) -

(B.8) 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., misspecification in the endogenous shock propagation part, including in particular the policy space Θ_ν. Since our proposed approach to policy evaluation does not require the researcher to take any stance on the shock process part $B(L)$, it is by design robust to such concerns.

A concrete illustration of this abstract discussion is provided by the model of Smets and Wouters. In that model, the exogenous shocks driving inflation already induce hump shapes (they follow ARMA(1,1)'s), and so other shocks—like monetary shocks—induce much weaker hump shapes than observed in the data; we thank Simon Gilchrist for making this point.

Weak identification. Standard full-information approaches to estimation of DSGE models are also often subject to concerns of weak identification (e.g., see Fernández-Villaverde et al., 2016). Our proposed limited-information approach is arguably less subject to this concern, simply because it only requires the researcher to partially specify the structural 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.[32]

[32] 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.

C Monetary policy causal effects

This appendix provides supplementary details on our analysis of monetary shock causal effects along the yield curve in Sections 3 and 4. We elaborate first on the empirical analysis in Appendix C.1, and then on model-based extrapolation in Appendices C.2 and C.3.

C.1 Empirical evidence

We provide further details on how we construct our empirical estimates of monetary policy shock transmission at the short end of the yield curve.

Data. We are interested in impulse responses of three outcome variables: the output gap, inflation, and the policy rate. These series are constructed as follows. Unless indicated otherwise, each series is transformed to stationarity following Hamilton (2018), and series names refer to FRED mnemonics.

- *Output gap.* We take log output per capita from FRED (`A939RX0Q048SBEA`). We interpret the stationarity-transformed series as a measure of the output gap.
- *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, and we consider a sample period from 1969:Q1–2006:Q4, stopping before the Great Recession to capture a relatively stable monetary regime. Our measure of monetary policy shock series are obtained from the replication files of Aruoba and Drechsel (2024) (for their shock) and Ramey (2016) (for the Romer and Romer shock). We aggregate by averaging the monthly series, and we set all missing values of the monetary shock IVs to zero, as in Känzig (2021).

Econometric implementation. We estimate a VAR in the two shock series together with our three outcome variables of interest. Following the recommendations of Plagborg-Møller and Wolf (2021), we order the Aruoba and Drechsel (2024) shock first in a recursive identification of our VAR, thus delivering invertibility-robust estimates of the desired dynamic causal effects. The Romer and Romer (2004) shock we instead order *after* output and inflation (but before the policy instrument), consistent with the original implementation in that paper. We include two lags, a linear time trend, and use a uniform-normal-inverse-Wishart distribution over the orthogonal reduced-form parameterization (as in Arias et al., 2018). Our estimation results are robust to these particular choices. This procedure yields draws of the monetary policy shock causal effect vector $\hat{\theta}_\nu$, which are then used to construct V_{θ_ν} following the steps outlined in Appendix B.1.

C.2 Structural models of monetary policy transmission

This section provides some supplementary details for our structural models of monetary transmission sketched in Section 4.2. 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 RANK

Households & unions. Households choose sequences of consumption c_t and assets a_t^H to maximize lifetime utility, given by

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

subject to a standard no-Ponzi condition as well as the budget constraint

$$c_t+a_t^H=w_t(1-\tau_t^\ell)\ell_t+d_t^H-\tau_t+\frac{1+r_{t-1}^n}{1+\pi_t}a_{t-1}^H, \tag{C.2}$$

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. We assume that $u(x)=\frac{x^{1-\gamma}}{1-\gamma}$ and $v(x)=\nu\frac{x^{1+\varphi}}{1+\varphi}$. 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—which summarizes the labor supply block—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:[33]

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

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.[34] 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:

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

[33]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.

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

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. 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 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:

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

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

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

and so, in log-deviations

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

or

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

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

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

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

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

Policy rule for computation. For our numerical analysis, we close the model with a determinacy-inducing Taylor rule, as discussed in Appendix B.2:

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

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 4 and 5, 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 HANK

The only two differences relative to the 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. 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.[35] 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). In order to compute the solution with informational rigidities, we follow Auclert et al. (2020): we first solve for the

[35] 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).

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 behavioral frictions

This subsection derives the price- and wage-NKPCs under cognitive discounting and price indexation. We derive the NPKCs under partial indexation and cognitive discounting, where ζ and ζ_w are the degrees of price indexation; $\zeta = \zeta_w = 1$ corresponds to the case considered in our main analysis.

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):

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

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:

$$\begin{aligned} \pi_t &= \frac{1-\theta_p}{\theta_p}\left(\tilde{P}_t^* - \tilde{P}_t\right) + \zeta\pi_{t-1} \\ (1-\zeta L)\pi_t &= \frac{1-\theta_p}{\theta_p}\left(\tilde{P}_t^* - \tilde{P}_t\right) \end{aligned} \tag{C.4}$$

Combining (C.3) and (C.4) we arrive at

$$(1-\zeta L)\pi_t = \frac{(1-\theta_p)(1-\bar{\beta}\theta_p)}{\theta_p}\mathbb{E}_t \sum_{\tau\geq t}^{\infty}(\bar{\beta}\theta_p m)^{\tau-t}\left[\tilde{\mu}_\tau + \sum_{s=t+1}^{\tau}(1-\zeta L)\pi_s\right]. \tag{C.5}$$

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 price-NKPC:

$$\pi_t-\pi_{t-1}=\kappa_p p_t^I+\beta^p\mathbb{E}_t\left[\pi_{t+1}-\pi_t\right] \tag{C.6}$$

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 the one used by the firm.[36]
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

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

[36]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.5). Operating exactly in the same way as before we obtain

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

With full wage indexation this gives the wage-NKPC used in our main analysis:

$$\pi_t^w - \pi_{t-1} = \kappa_w \chi_t + \beta^w \mathbb{E}_t \left[\pi_{t+1}^w - \pi_t \right] \tag{C.7}$$

C.3 Model calibration and estimation

This section provides details on the parameterization of our estimated models of monetary transmission. We proceed in two steps—first the calibration part, and then the estimation.

Calibration. For all four 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.[37]

We follow Wolf (2023) for the steady state calibration of the fiscal side. 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

[37] 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.

nominal assets is set to 27 percent of annual 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 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.[38]

Estimation. We estimate all models to ensure consistency with the empirical monetary policy shock impulse response targets $\hat{\theta}_\nu$. For the baseline RANK model we estimate five parameters: the strength of habits (h), the degrees of price as well as wage rigidity (θ_p and θ_w), the curvature of investment adjustment costs (κ), and the curvature of capacity utilization costs (ζ). For the baseline HANK model, the household information stickiness parameter (θ) replaces the degree of habit formation (h). Finally, for the behavioral models, 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 our data are only weakly informative about m, as is implicit in the results displayed in Table 4.1.

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

[38] 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 4 and 5.

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).

Model	Parameter	Prior			Posterior				
		Dist.	Mean	St. Dev	Mode	Mean	Median	5 percent	95 percent
RANK - RE	h	Beta	0.70	0.10	0.7656	0.7508	0.7585	0.6005	0.8769
	θ_p	Beta	0.67	0.20	0.9244	0.9094	0.9288	0.7820	0.9710
	θ_w	Beta	0.67	0.20	0.8795	0.7715	0.8353	0.3803	0.9623
	κ	Normal	5.00	1.50	5.4630	5.7889	5.7591	3.7038	7.9633
	ψ	Beta	0.50	0.15	0.3865	0.4547	0.4516	0.2167	0.7043
HANK - RE	θ	Beta	0.70	0.20	0.9812	0.9474	0.9589	0.8613	0.9946
	θ_p	Beta	0.67	0.20	0.9424	0.9136	0.9360	0.7733	0.9763
	θ_w	Beta	0.67	0.20	0.8534	0.7794	0.8340	0.4160	0.9635
	κ	Normal	5.00	1.50	5.8763	5.8997	5.8594	3.8234	8.1150
	ψ	Beta	0.50	0.15	0.4509	0.4387	0.4331	0.2085	0.6949
RANK - CD	h	Beta	0.70	0.10	0.7589	0.7522	0.7599	0.5971	0.8793
	θ_p	Beta	0.67	0.20	0.8596	0.8797	0.9092	0.7138	0.9619
	θ_w	Beta	0.67	0.20	0.9460	0.7608	0.8205	0.3760	0.9611
	κ	Normal	5.00	1.50	5.5686	5.8100	5.7563	3.7332	8.0054
	ψ	Beta	0.50	0.15	0.4945	0.4618	0.4600	0.2246	0.7078
HANK - CD	θ	Beta	0.70	0.20	0.9787	0.9483	0.9600	0.8623	0.9949
	θ_p	Beta	0.67	0.20	0.8499	0.8764	0.9036	0.7114	0.9667
	θ_w	Beta	0.67	0.20	0.9498	0.7964	0.8668	0.4251	0.9627
	κ	Normal	5.00	1.50	5.5916	5.9648	5.9323	3.8650	8.1611
	ψ	Beta	0.50	0.15	0.4994	0.4432	0.4396	0.2122	0.6889

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 contains supplementary results for our three monetary policy counterfactual applications in Section 5.

D.1 Reduced-form projections

We provide supplementary details on how we construct the reduced-form projections for our three applications. We elaborate on data construction and econometric implementation, and also compare the implied forecasts with other approaches.

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.

- *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. Note that the output gap, inflation, and federal funds rate series are all constructed exactly as in Appendix C.1 for our empirical analysis of monetary policy shock propagation.

Econometric implementation. We restrict attention to OLS point estimates. 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 second moments. For the forecast-based counterfactuals in Sections 5.3 and 5.4, we include two lags. Our use of a reduced-form VAR for this purpose agrees with the findings in Li et al. (2023) who show that VARs tend to dominate other estimation methods in terms of mean-squared error and thus for point estimation.

Comparison with alternative forecasts. We now perform two additional checks to demonstrate the good forecasting performance of our reduced-form VAR: we (i) check that the forecast accuracy is similar to that of the Survey of Professional Forecasters (SPF); and (ii) show that our 10-variable system contains nearly all of the information in the eight business cycle factors that Stock and Watson (2016) computed from a large set of macroeconomic and financial variables.

1. *Comparison with the SPF.* We assess forecast accuracy starting in 1981:Q3 (when the SPF forecast for the T-Bill rate becomes available) and ending in 2007:Q3 (before the onset of the Great Recession and the ZLB period).[39] Table D.1 shows the mean squared errors of the one- and four-quarter-ahead forecasts from our VAR and from the SPF, for our three main series of interest. Our VAR evidently performs well by this metric.

[39] In order to allow an apples-to-apples comparison with the SPF, we need to slightly modify our VAR. Specifically, we use raw GDP data (not per capita) and the T-Bill rate in place of the federal funds rate, as those are the variables that appear in the SPF. As in the baseline VAR analysis, we detrend all non-stationary series, but to compare to the SPF we add the trends back to the VAR-implied forecasts.

Variable	1-quarter ahead		4-quarter ahead	
	VAR	SPF	VAR	SPF
GDP	0.569	0.327	2.76	2.97
Inflation	0.357	0.934	0.646	1.85
T-Bill rate	0.461	0.740	1.66	3.17

Table D.1: Mean squared error of 1- and 4-quarter ahead forecasts.

	1-quarter ahead		4-quarter ahead	
	w/o f	w/ f	w/o f	w/ f
Output gap	0.918	0.935	0.648	0.774
Inflation	0.826	0.838	0.716	0.732
Interest rate	0.945	0.953	0.759	0.781

Table D.2: Assessing the incremental information content of the Stock-Watson factors: forecasting R^2 with and without inclusion of factors in the VAR.

2. *Information content of the Stock and Watson factors.* Stock and Watson (2016) estimate 8 factors that drive the bulk of the variation in a database of 207 quarterly time series on the U.S. macro-economy and financial markets. We now ask whether adding these factors to the information set of our VAR would lead to a substantial improvement in forecasting performance. Specifically, let the variables in our VAR be represented by the vector y_t and the 8 factors be represented by the vector f_t. For horizon $h \in \{1, 4\}$, we consider a regression of the form

$$y_{t+h} = B_0 y_t + B_1 y_{t-1} + B_f f_t,$$

and then assess the implications of setting $B_f = 0$. Table D.2 shows the results for our core observables. We see that, with the exception of the 4-quarter-ahead forecast of the output gap, the increase in R^2 from including the factors is quite small. We thus judge that including the Stock-Watson factors would not lead to materially different forecasts.

D.2 Average business cycle

We provide evidence 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 not driven by the Great Recession. To do so we in Figure D.1 construct counterfactual volatilities with reduced-form forecasts obtained on a sample that only stretches to 2007:Q1. We see that

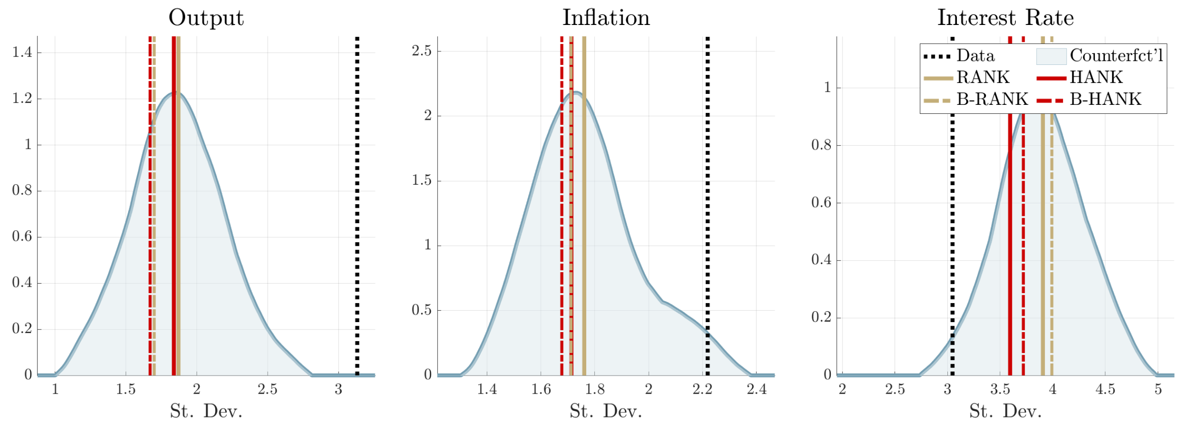

Figure D.1: Counterfactual early-sample (1960:Q1 – 2007:Q1) average volatilities of output, inflation, and the federal funds rate, under the policy rule that minimizes (9). Black dashed: data point estimate under observed policy. Blue: posterior Kernel density of counterfactual volatilities drawing from 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).

the picture is essentially unchanged relative to Figure 6: 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 same logic from our discussion in Section 5.2 continues to apply.

D.3 Great Recession

Figure D.2 constructs the Great Recession counterfactual using only our two empirically estimated monetary policy shock impulse responses. The resulting counterfactuals are broadly similar to our main results in Figure 9: the nominal interest rate is cut aggressively, leading to more stable output, at the cost of elevated inflation. Moving from the empirically estimated monetary policy causal effects to the entirety of the model-implied Θ_ν smoothes out the rate cut and helps somewhat better stabilize output; that being said, the differences are moderate, suggesting that the counterfactual policy does not rely much on model-implied extrapolation to the causal effects of interest rates forward guidance.

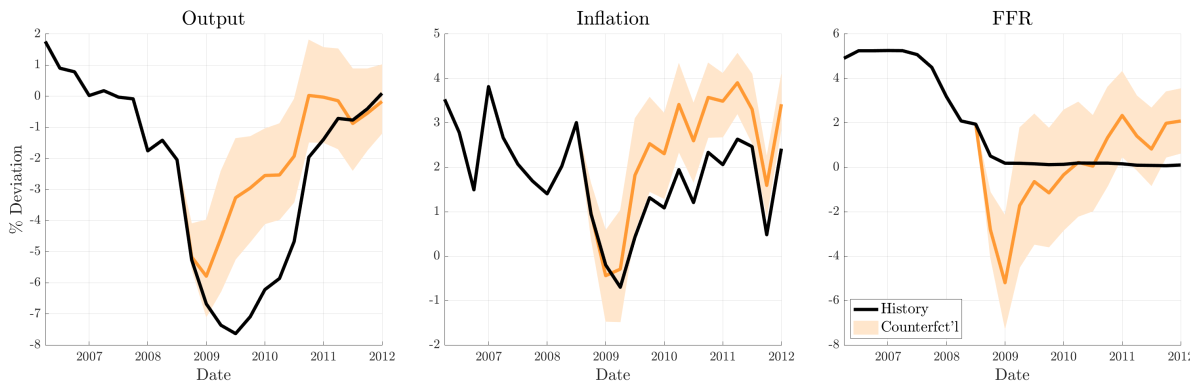

Figure D.2: Counterfactual evolution of inflation, output, and the federal funds rate in the Great Recession, under the policy that minimizes (9) without any effective lower bound on rates. Black: data. Orange: posterior median (solid) and 16th and 84th percentile bands (shaded), using only the two empirically estimated monetary policy shock IRFs.

D.4 Post-covid inflation

To complement the discussion in Section 5.4 we in Figure D.3 show counterfactual posterior modes for our four estimated models of monetary transmission separately (red and beige, solid and dashed). We see that the two behavioral and two rational-expectations counterfactuals are very similar, consistent with Figure 10. This further underscores our claims that behavioral frictions—and not market incompleteness, as in the HANK literature—are most important for monetary shock extrapolation to the long end of the yield curve.

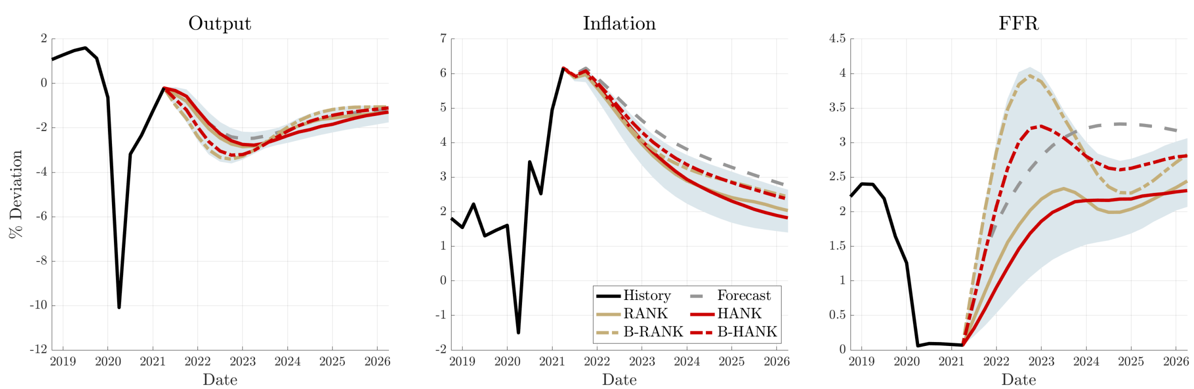

Figure D.3: Counterfactual projections of inflation, output, and the federal funds rate in the post-covid inflationary episode (from 2021:Q2), under the policy that minimizes (9). Black: data. Grey: actual forecast. Blue: 16th and 84th percentile bands drawing from posterior across all models and parameters. Beige: posterior mode of counterfactual using RANK models (baseline and behavioral). Red: posterior mode of counterfactual using HANK models (baseline and behavioral).