Thinking about the Economy, Deep or Shallow?*

Pierfrancesco Mei† Lingxuan Wu‡

Job Market Paper

November 6, 2024

Click here for latest version

Abstract

We propose a theory of *shallow thinking* to capture people's limited understanding of the long causal chains involved in the propagation of shocks. We cast general equilibrium as a system of causal relations in a directed cyclic graph. Estimation from our qualitative survey suggests that, on average, people think about only 2.6 steps of propagation, overlooking much of the graph and deviating significantly from rational expectations. Our theory implies that longer causal chains have diminishing influence on beliefs. Applying shallow thinking to a New Keynesian model with active monetary policy reconciles several bond market puzzles and yields macroeconomic consequences: (i) long-term interest rates underreact to cost-push shocks but overreact to monetary policy shocks; (ii) inflation expectations negatively predict bond excess returns; and (iii) the inflation response to cost-push shocks is stronger than under rational expectations and, contrary to that paradigm, increases with shock persistence. In a real business cycle model, relative to rational expectations, shallow thinking amplifies and prolongs output fluctuations from productivity shocks and predicts negative future stock excess returns.

*We thank Peter Andre, Adrien Bilal, John Campbell, Juanma Castro-Vincenzi, Gabriel Chodorow-Reich, Ben Enke, Xavier Gabaix, Thomas Graeber, Sam Hanson, Oleg Itskhoki, Baiyun Jing, Spencer Kwon, David Laibson, Chen Lian, Avi Lipton, Hugo Monnery, Matthew Rabin, Chris Roth, Karthik Sastry, Josh Schwartzstein, Dmitriy Sergeyev, Andrei Shleifer, Stefanie Stantcheva, Jeremy Stein, Ludwig Straub, Adi Sunderam, Alireza Tahbaz-Salehi, Chris Tonetti, Chris Wolf, and seminar participants at Harvard for their helpful comments. We thank Sam Cohen for research assistance and Roberto Colarieti for help with the survey design. The study received IRB approval from Harvard (IRB22-1403, IRB24-0959). We acknowledge financial support from the Michael S. Chae Macroeconomic Policy Fund and the Molly and Domenic Ferrante Fund, both awarded through Harvard. Wu is grateful for support from the Alfred P. Sloan Foundation Pre-Doctoral Fellowship in Behavioral Macroeconomics, awarded through the NBER.

†Harvard University, pierfrancescomei@g.harvard.edu, https://www.pierfrancescomei.com.

‡Harvard University, lingxuanwu@g.harvard.edu, https://www.lingxuanwu.me, *Corresponding Author*.

1 Introduction

In general equilibrium models of the economy, macroeconomic variables (such as inflation, interest rates, and firms' dividends) respond to shocks (such as monetary policy shocks) through complex *causal relations* of agents' responses and market outcomes. Expectations of these variables are crucial in macroeconomics and finance, influencing household consumption, firms' pricing, capital investment, asset pricing, and more. How well do actual economic agents understand these causal relations when forming their expectations? And how do these expectations, in turn, affect the economy's response to shocks?

The prevailing rational expectations hypothesis amounts to assuming that people understand all causal relations in the economy. However, growing evidence suggests that people do not understand the responses of macroeconomic policies to economic conditions (Cieslak, 2018; Bauer, Pflueger and Sunderam, 2024*a*) or the effects of shocks and policy changes on the economy (Andre et al., 2022; D'Acunto, Hoang and Weber, 2022; Coibion et al., 2023*b*). Research from behavioral economics and cognitive psychology further establishes that human causal reasoning is limited compared to the rational benchmark.[1]

In light of these insights, we propose a theory of *shallow thinking* to model people's limited understanding of the economy in general equilibrium.[2] We conceptualize general equilibrium as a system of causal relations in a directed cyclic graph, where loops embed general equilibrium feedbacks. These causal relations capture how one macroeconomic variable depends on others, driven by agents' responses (e.g., consumption responding to interest rates) and price determination in competitive markets (e.g., price changes due to supply or demand shifts). Motivated by the aforementioned evidence, we assume that people understand only short chains starting from a shock in the directed graph. We develop a survey to test and measure shallow thinking. We find that, on average, people think about only 2.6 steps, ignoring much of the economy. Consequently, causal relations more distant from shocks and longer feedback loops have less influence on beliefs.

[1] People underappreciate how new policies lead to new equilibriums in economic settings (Dal Bó, Dal Bó and Eyster, 2018). They struggle to understand complex causal relations for predictive tasks (Kendall and Oprea, 2024), and make predictions that are insufficiently sensitive to the strengths of causal relations (Rottman and Hastie, 2014). Further, they pay special attention to earlier nodes in causal chains (Ahn et al., 2000), and their knowledge of complex causal systems is sparse and shallow (Rozenblit and Keil, 2002).

[2] Recent work has relaxed the assumption of full-information rational expectations by removing common knowledge (Angeletos and Lian, 2018) or by modeling agents' limited strategic sophistication (García-Schmidt and Woodford, 2019; Farhi and Werning, 2019) or myopia (Gabaix, 2020), among other notable contributions. However, these studies still assume that agents understand general equilibrium. We will discuss our connection to these papers in more detail later.

We apply shallow thinking to the workhorse New Keynesian and real business cycle (RBC) models to uncover its consequences for macroeconomics and finance. These models feature multiple feedback loops that either amplify or offset shocks. By assigning less weight to longer loops, shallow thinking alters the sign or magnitude of the perceived net effect of general equilibrium feedbacks, which causes belief under- or overreaction in different occasions. It suggests belief underreaction of a variable to shocks affecting other variables, as agents underperceive shock propagation, and belief overreaction to shocks directly impacting the variable itself, in the presence of a long offset loop that is underappreciated. In the New Keynesian model, shallow thinking predicts that long-term interest rates underreact to cost-push shocks but overreact to monetary policy shocks. It also results in a larger inflation response to cost-push shocks than under rational expectations. In the RBC model, shallow thinking amplifies the economy's responses to productivity shocks and leads to a stock market boom and crash.

To begin with, in Section 2, we introduce shallow thinking by conceptualizing the textbook New Keynesian model with an active Taylor rule as a system of causal relations. We study transitory news shocks, namely period-1 shocks that are known in period 0, and later generalize to persistent shocks. Crucially, as agents observe all variables in each period, the period-1 equilibrium is a static general equilibrium, with no role for beliefs. However, the period-0 equilibrium depends on agents' beliefs about period-1 outcomes, as households and firms make forward-looking decisions. We introduce shallow thinking regarding period 1 and explore its consequences for the period-0 equilibrium.

Figure 1 depicts causal relations among macroeconomic variables in the period-1 economy using a directed cyclic graph. To draw this graph, we start from the basics of general equilibrium: agents' best responses and price determination in competitive markets. We cast each competitive market as a fictitious auctioneer who sets the price in response to changes in demand and supply, capturing Walras's idea of price adjustment via tâtonnement. With such auctioneers as fictitious agents, there is a unique representation of general equilibrium as a system of all agents' best response functions. We define a *causal relation* as the dependence of one variable on another (i.e., a partial derivative) driven by agents' responses. As reflected in Figure 1, each solid arrow represents a variable chosen by agents, pointing to the agents it influences, and an dashed arrow indicates a shock that changes some agents' behavior. The graph is cyclic, representing the equilibrium as a fixed point, appreciated by rational agents who take infinite steps starting from the shock.

In contrast to rational expectations, we hypothesize that individuals foresee only a

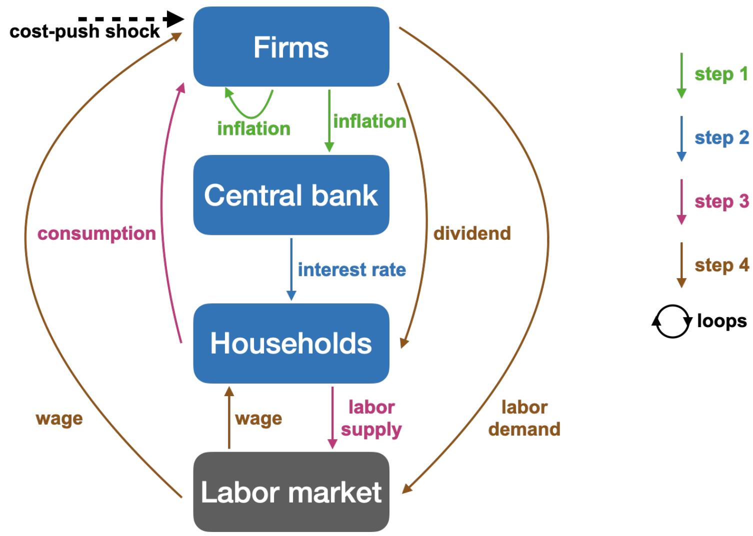


Figure 1: Period-1 New Keynesian economy as directed graph of causal relations

Notes: This figure depicts the causal relations among macroeconomic variables in the period-1 New Keynesian economy as a directed cyclic graph. Each node represents an agent type (firms, the central bank, households) or a competitive market (the labor market). For a competitive market, the price is interpreted as being set by a fictitious Walrasian auctioneer in response to supply and demand shifts, as explained in Section 2.1. We define a *causal relation* as the dependence of one variable on another (i.e., a partial derivative), arising from agents' best responses. Dashed arrows indicate shocks affecting some agents (e.g., cost-push shocks that trigger firms to raise prices in the absence of any changes in marginal cost), while solid arrows represent variables—decisions by either actual agents (e.g., inflation, interest rate, and consumption) or fictitious auctioneers (e.g., wage). Each arrow points to the agents (including fictitious ones) responding to it. Variables are color-coded to illustrate the step-by-step propagation of a cost-push shock.

finite number of steps in shock propagation on the graph, which we refer to as their *depth of thinking* d, and that d varies across the population. For example, we color-code Figure 1 to illustrate the propagation of a cost-push shock à la Clarida, Galí and Gertler (1999), i.e., a shock that triggers firms to raise prices in the absence of any changes in marginal cost. A depth-1 agent acknowledges only the most obvious implication: firms will raise their prices (causing inflation), while overlooking changes in all other variables. A depth-2 agent further appreciates that the central bank will raise the interest rate in response to higher inflation according to a standard Taylor rule, but fails to foresee additional implications. A depth-3 agent understands that a higher interest rate will discourage household consumption and incentivize labor supply. A depth-4 agent recognizes that changes in household behavior will affect the firms and the labor market. This iteration continues infinitely, and only a depth-∞ (i.e., rational) agent correctly assesses the strength of all loops and accurately forecasts the period-1 equilibrium. This iterative process captures the idea that more distant causal effects are harder to grasp.

Our theory makes a novel prediction about belief heterogeneity, enabling us to measure the distribution of d across the population. A shallow agent perceives changes only in a subset of variables that are close to the shocks in the directed graph, and this subset varies across shocks, as each shock propagates from a different part of the graph. Consequently, our theory predicts that, across variables and shocks, changes in more distant variables are perceived by fewer people, and this pattern informs the distribution of d.

We capture this rich belief heterogeneity with a parsimonious parametrization, assuming that the depth of thinking d follows a geometric distribution with continuation rate $\lambda \in [0, 1]$. This parameter λ is the only input required to apply our theory to macroeconomic models. A higher λ means that people think more deeply on average, with $\frac{1}{1-\lambda}$ representing the average depth of thinking, and $\lambda = 1$ nesting rational expectations (i.e., infinite depth). The average expectations, which drive the equilibrium in a large class of models (including those analyzed in this paper), are parsimoniously parametrized by λ. These average expectations are *as if* generated by a representative agent who knows all causal relations but dampens them by λ.

In Section 3, we develop a survey to test our theory and measure λ, investigating the general public's understanding of shock propagation. We ask respondents to reason through hypothetical scenarios involving classic macroeconomic shocks, such as oil shocks and monetary policy shocks. For each shock scenario, respondents provide directional forecasts for key macroeconomic variables, such as inflation and interest rates. We use directional responses from the empirical literature as the objectively correct answers. We then run a panel regression of correct directional forecasts on variable distance across respondents, variables, and shocks. We show that variable distance strongly predicts correct directional forecasts with a negative sign. This finding is robust to the inclusion of a rich set of fixed effects, which absorb potentially confounding sources of belief heterogeneity.

Our estimation strongly rejects the null hypothesis of $\lambda = 1$ and suggests $\lambda \approx 0.61$, implying an average depth of thinking ($\frac{1}{1-\lambda}$) of only about 2.6—far below the infinite depth assumed under rational expectations. This low value underscores the relevance of our emphasis on limited depth of thinking, as opposed to differences in perceived causal models, such as disagreement about model parameters or causal links. While such alternatives can generate heterogeneous beliefs, they struggle to explain why people misjudge directional responses in a way correlated with variable distance. Moreover, as people consider few steps, they overlook much of the economy, diminishing the role of differing perceptions about distant causal relations.[3] Later in the paper, we present formulas suggesting that the perceived strength of general equilibrium feedback loops declines exponentially with their length under shallow thinking, while other model parameters (such as the Taylor rule coefficient) and any misperceptions thereof only affect it proportionally, highlighting the importance of our emphasis.

In Section 4, we show that shallow thinking, deviating from rational expectations with a single parameter λ, reconciles several bond market puzzles that may seem unrelated or even contradictory. Bauer, Pflueger and Sunderam (2024*a*) show that long-term interest rates responded too little to inflation surprises before the March 2022 interest rate hike.[4] In contrast, a large body of literature suggests that long-term interest rates are excessively sensitive to monetary policy shocks (Cochrane and Piazzesi, 2002; Gürkaynak, Sack and Swanson, 2005; Hanson and Stein, 2015). Additionally, various macroeconomic variables predict bond excess returns, controlling for current yields (Ludvigson and Ng, 2009; Cooper and Priestley, 2009; Joslin, Priebsch and Singleton, 2014; Cieslak and Povala, 2015). Our theory suggests that the period-0 long-term interest rate *underreacts* to cost-push news shocks, because agents underperceive the Taylor rule—that the central bank will raise the short-term interest rate in response to inflation in period 1. Conversely, the period-0 long-term interest rate *overreacts* to monetary policy news shocks, as agents underappreciate a long offset loop: a positive monetary policy shock will lower inflation, prompting the central bank to slightly lower the short-term rate per the Taylor rule. With multiple shocks, current yields (driven by interest rate expectations) mainly reflect monetary policy shocks, while other macroeconomic variables load more on other shocks, providing additional predictive information on bond excess returns in a multivariate regression.

We then demonstrate an important macroeconomic consequence of shallow thinking: cost-push news shocks are more inflationary in period 0 than under rational expectations. In response to cost-push news shocks, interest rate expectations *underreact*, as established previously, while inflation expectations *overreact*. The overreaction arises from misperceptions of two feedback loops in general equilibrium—one amplifying the inflation response and one offsetting it. Inflation amplifies itself, as firms raise prices in response to higher aggregate inflation, creating a self-loop. It is offset by a longer loop involving monetary tightening in response to inflation, which lowers the real wage by influencing household

[3] For example, in response to cost-push shocks, a depth-2 agent perceives the response of the interest rate to higher inflation (i.e., the Taylor rule), but does not consider the responses of consumption or labor supply to the interest rate, nor the resulting change in wages.

[4] Cieslak (2018) finds that forecasters systematically underappreciate monetary easing during recessions.

labor supply. Shallow agents better understand the short amplification loop and thus perceive net amplification of inflation, though the true net effect is offset as the long offset loop is actually stronger. Because of the overestimation of future inflation and the underappreciation of the future interest rate hike and associated economic downturn, firms set higher prices and households consume more in period 0, leading to higher inflation.

In Section 5, we highlight another macroeconomic consequence of shallow thinking: more persistent cost-push shocks lead to higher inflation, by generalizing shallow thinking to persistent shocks in the New Keynesian economy. With persistent shocks, we focus on causal relations across variables, abstracting from the cross-horizon dimension.[5] We show that persistent cost-push shocks are more inflationary under shallow thinking than under rational expectations, as with transitory news shocks. Furthermore, *more* persistent cost-push shocks result in *higher* inflation under shallow thinking, whereas rational expectations predict lower inflation. This difference arises because a more persistent shock strengthens general equilibrium feedbacks, particularly boosting the long offset loop discussed previously. A rational agent recognizes that a more persistent cost-push shock is offset more, leading to lower inflation. In contrast, shallow agents anticipate higher future inflation via the short amplification loop, and their responses bring that about.

Last, in Section 6, we consider an RBC economy and demonstrate that shallow thinking amplifies the economy's responses to productivity shocks and results in a stock market boom and crash. In response to a persistent productivity shock, shallow agents underappreciate that the expansion in firms' labor demand will push up wages, thereby lowering the future return on capital. Consequently, firms invest more than in the rational expectations equilibrium. This overaccumulation of capital leads to a hump-shaped, persistent boom in output and an amplified response in labor hours. Additionally, agents underestimate dividends in the short term but overestimate them in the long term, resulting in an initially positive stock excess return that turns negative after a few quarters. This pattern is consistent with the classic Kindleberger (1978) narrative of crises, where innovations are followed by asset market booms and crashes.

In summary, shallow thinking captures people's limited understanding of the economy and alters its responses to shocks in ways that align with a range of empirical evidence.

[5] With persistent shocks, we define cross-variable causal relations as the *sequence-space Jacobians* à la Auclert et al. (2021), generalizing the partial derivatives in the case of transitory news shocks. To focus on the cross-variable aspect of causal reasoning, for example, regarding the dependence of consumption $\{c_t\}_{t\geq 0}$ on interest rates $\{i_t\}_{t\geq 0}$, we assume that if an agent understands how c_t depends on contemporaneous i_t, they also understand how c_t depends on future i_s.

1.1 Literature Review

At a high level, our theory enriches prior work by suggesting that, among multiple causal relations in general equilibrium, more distant ones are more dampened. Angeletos and Lian (2023*a*) review recent research that moves beyond full-information rational expectations and highlight a key commonality: in a stylized model in which general equilibrium operates through a single feedback effect, several prominent theories are equivalent in dampening that effect. Building on this insight, we show that workhorse macroeconomic models feature multiple causal relations and loops, some with opposing signs. By assigning less weight to longer feedback loops, shallow thinking alters the sign or magnitude of the perceived net effect of all general equilibrium feedbacks.[6]

Our theory closely relates to and broadens research on agents' limited strategic sophistication in macroeconomics and finance, with a consequential difference in modeling approach. This includes studies on macroeconomic policies, such as García-Schmidt and Woodford (2019), Farhi and Werning (2019), Iovino and Sergeyev (2023), and Bianchi-Vimercati, Eichenbaum and Guerreiro, 2024, as well as Greenwood and Hanson (2015) and Bastianello and Fontanier (2024) in finance.[7] These papers introduce models of level-k thinking (Nagel, 1995; Stahl and Wilson, 1994, 1995; Camerer, Ho and Chong, 2004) and competition neglect (Camerer and Lovallo, 1999) from the experimental and game-theoretical literature into general equilibrium. While they focus on households and firms, our theory generalizes this framework by including policymakers and Walrasian auctioneers as additional players, capturing agents' limited understanding of the macroeconomy via the technical apparatus of level-k thinking.[8,9] We show that the underappreciation of

[6]Moreover, prior work generates *horizon-dependent dampening* due to bounded rationality (Gabaix, 2020; Farhi and Werning, 2019) or information frictions (Angeletos and Lian, 2018; Angeletos and Huo, 2021). We focus on *cross-variable dampening* as a complementary aspect, and discuss the connection in Section 5.1.

[7]Compared to the simple q-theory model in Greenwood and Hanson (2015), which assumes an exogenous demand curve and fixed interest rates, we consider an RBC model that endogenizes interest rates, wages, demand, and misperceptions thereof.

[8]Level-k thinking models agents' reasoning about opponents' limited strategic sophistication in unfamiliar games. The cited papers aptly apply this to study unconventional macroeconomic policies (e.g., forward guidance), treating the New Keynesian model as a dynamic game among households and firms. Shallow thinking allows for the underappreciation of policy rules (such as monetary policy) and price determination, applicable to models not typically considered strategic interactions (such as the RBC model) as well as to conventional shocks (such as monetary policy shocks), all driven by a lack of knowledge.

[9]Moreover, in our survey, we measure strategic sophistication using the classic "guess 2/3 of the average" game and find *no* correlation with understanding of macroeconomic shocks (Appendix B.3). A reasonable interpretation is that shallow thinking reflects limited knowledge about the macroeconomy—a different aspect of bounded rationality than limited strategic sophistication.

policy rules and price determination is empirically relevant and consequential. Dampening multiple causal relations based on distance leads to a different perceived net effect from first collapsing them into a true net effect and then dampening that.

Our theory examines rationality in the absence of information frictions, and complements a large theoretical literature that preserves rationality while focusing on information frictions (Lucas, 1972; Gabaix and Laibson, 2001, Mankiw and Reis, 2002; Woodford, 2003*a*; Nimark, 2008; Angeletos and Lian, 2018; Angeletos and Huo, 2021),[10] rational inattention (Sims, 2003; Maćkowiak and Wiederholt, 2009; Molavi, 2019; Miao, Wu and Young, 2022) and learning (Evans and Honkapohja, 2001; Eusepi and Preston, 2018). Our survey shows that respondents may not correctly forecast the directional responses of macroeconomic variables, even with full information about the shocks, which supports our theory.

Our theory generates belief under- and overreaction in a manner endogenous to the causal relations involved. It predicts belief underreaction of a variable to shocks affecting other variables and belief overreaction to shocks that directly impact it, the latter occurring in the presence of a long offset. Hence it adds a layer of richness to theories of overreaction (Barberis, Shleifer and Vishny, 1998; Bordalo et al., 2020; Afrouzi et al., 2023; Azeredo da Silveira, Sung and Woodford, 2024), in a way that reconciles several bond market puzzles regarding under- and overreaction of long-term interest rates as previously discussed.

We provide a theory of heterogeneous mental models, contributing to the literature that uses surveys to measure people's mental models in specific scenarios (Stantcheva 2021, 2023*b*; Andre et al. 2022, 2024; Andre, Schirmer and Wohlfart 2024). Andre et al. (2022) show that people's macroeconomic forecasts in response to hypothetical shocks are highly heterogeneous in both level and direction. Our theory generates belief heterogeneity, and our survey confirms its prediction on the correctness of directional beliefs, bringing some order to their findings on belief heterogeneity. Prior to this paper, Wu (2023) presents evidence on people's imperfect mental models using existing survey forecasts. Compared to Wu (2023), this paper develops a survey that offers evidence on the heterogeneity of mental models, and derives consequences for macroeconomics and finance.

Our use of directed *cyclic* graphs connects to work using directed *acyclic* graphs (DAGs) for different purposes. Building on Auclert et al. (2021), who use DAGs to depict and solve macroeconomic models and break cycles only for technical convenience, we instead employ a cyclic graph to represent people's mental models and emphasize loops to capture

[10] Angeletos and La'O (2009) and Angeletos and Sastry (2021) remove common knowledge of the shock by introducing heterogeneous priors while preserving common knowledge of rationality.

equilibrium. We complement Spiegler (2016, 2020), who examines agents fully utilizing a perceived DAG with causal links that differ from the true DAG (e.g., reverse causality, ignoring confounders). We focus instead on the consequences of not thinking through all causal links, predicting that agents consider different links depending on the shock, as supported by our survey.[11] Further afield, Pearl (2009) and others in causal inference use DAGs for identification, which is not our aim.[12] Decades of macroeconomic research have been devoted to identifying the true model; we take established models as given, recognizing their inherent cyclic dependencies, akin to structural vector autoregression.

2 Shallow Thinking in a New Keynesian Economy

We set up the textbook New Keynesian model in Section 2.1 with transitory news shocks, i.e. period-1 shocks that are observed in period 0. We conceptualize the period-1 general equilibrium as a system of causal relations in a directed graph in Section 2.2. This causal representation of general equilibrium is our broader theoretical contribution that nests the New Keynesian model as well as other applications. Accordingly, we introduce the theory of shallow thinking and discuss our emphasis on the unique causal representation of general equilibrium as best responses in Section 2.3.

2.1 The New Keynesian Economy

We consider the New Keynesian model à la Woodford (2003*b*) and Galí (2015). The economy consists of three types of agents (firms, households, and a central bank) and a competitive labor market. We take a log-linear approximation around the steady state for simplicity and use lower-case letters for log-linear deviations.

We study period-1 shocks that are observed in period 0. Since the New Keynesian model is purely forward-looking, the economy returns to its steady state from period 2 onwards. Appendix C.1 provides the full details of the infinite-horizon model. Here, we focus on period 1, obtained by setting all log deviations from period 2 onwards to zero.

[11] Andre et al. (2024) suggest that people give narratives about the macroeconomy that can be represented by DAGs. However, our general equilibrium models feature cyclic dependencies. Our theory suggests that people may think with DAGs *because* they do not take enough steps to appreciate loops.

[12] For reviews, see Heckman and Pinto (2015) and Imbens (2020). Our treatment of equilibrium prices as determined by demand and supply shifts aligns with Imbens (2020)'s suggested representation in his Figure 12B, drawing on seminal work on simultaneous equations (Tinbergen, 1930; Haavelmo, 1943).

We conceptualize the period-1 general equilibrium as a system of causal relations as follows: (i) we interpret price determination in a competitive market as a fictitious Walrasian auctioneer setting a price in response to supply and demand shifts; (ii) we maintain the structural form of best responses of all agents (including Walrasian auctioneers), expressing their decisions as functions of decision-relevant variables. Importantly, this gives rise to a unique representation of general equilibrium as best responses.

Firms. There is a continuum of firms indexed by $j \in [0,1]$ that produce using labor to satisfy demand and set prices subject to Calvo rigidity. In period 1, firms choose labor demand, pay dividends, and reset prices if possible, taking as given the aggregate inflation rate π_1, real wage w_1, and aggregate demand c_1. Each firm produces a differentiated good, which collectively forms a constant-elasticity bundle that households consume, and charges a markup μ in the steady state.

All firms produce to satisfy demand using the same linear technology in labor, giving rise to the aggregate dividend and labor demand

$$\begin{aligned} div_1 &= c_1 - \frac{1}{\mu - 1} w_1 \\ n_1^d &= c_1 \end{aligned} \tag{1}$$

To anticipate our analysis of the labor market, we interpret labor demand n_1^d as a demand curve $n_1^d = \hat{n}_1^d + e_{n^d w} w_1$, which shifts by

$$\hat{n}_1^d = c_1 \tag{2}$$

and has an elasticity $e_{n^d w}$, which is 0 in this model, as firms only use labor in production.

A $(1-\theta)$ share of firms can reset their prices in period 1 to maximize dividends, and each chooses

$$p_{j1}^* = p_0 + (1 - \beta\theta)\left[w_1 + \sum_{k=0}^{\infty} (\beta\theta)^k \pi_1\right]$$

where β is the household time discount rate, the inverse of which equals the steady-state interest rate. Aggregate inflation results from the pricing behavior of the $(1-\theta)$ share of resetting firms as $\pi_1 = (1-\theta)\left(p_{j1}^* - p_0\right)$. Following the tradition at least since Clarida, Galí and Gertler (1999), we consider a cost-push shock ϵ_1^{π}, and thus inflation is

$$\pi_1 = \theta\kappa w_1 + (1-\theta)\,\pi_1 + \epsilon_1^{\pi} \tag{3}$$

with $\kappa \equiv \frac{(1-\theta)(1-\beta\theta)}{\theta}$ capturing the slope of the Phillips curve. Importantly, we do not move π_1 on the right-hand side to the left. We intentionally preserve the dependence of π_1 on itself, which encapsulates the within-period complementarity in individual price-setting, as each firm takes aggregate inflation as given.

Households. There is a continuum of households who live infinitely and maximize their lifetime utility, discounted by β, which is separable in consumption and labor supply. In period 1, households choose consumption and labor supply, taking as given the nominal interest rate i_1, the real wage w_1, and dividend div_1. Their optimal consumption and labor supply decisions are given by

$$\begin{aligned} c_1 &= -\sigma^{-1}\beta i_1 + \frac{(1-\beta)(\mu-1)\nu}{\sigma+\mu\nu} div_1 + \frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu} w_1 \\ n_1^s &= \nu^{-1}\beta i_1 - \frac{(1-\beta)(\mu-1)\sigma}{\sigma+\mu\nu} div_1 + \nu^{-1}\left[1-\sigma\frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu}\right] w_1 \end{aligned} \tag{4}$$

where σ^{-1} is the elasticity of intertemporal substitution, and ν^{-1} is the the Frisch elasticity of labor supply.

Similar to labor demand, we interpret labor supply n_1^s as a supply curve $n_1^s = \hat{n}_1^s + e_{n^s w} w_1$, which shifts by

$$\hat{n}_1^s = \nu^{-1}\beta i_1 - \frac{(1-\beta)(\mu-1)\sigma}{\sigma+\mu\nu} div_1 \tag{5}$$

and has an elasticity $e_{n^s w} = \nu^{-1}\left[1-\sigma\frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu}\right]$.

Central bank. The central bank follows a Taylor rule with a monetary policy shock ϵ_1^i,

$$i_1 = \phi\pi_1 + \epsilon_1^i \tag{6}$$

Labor market. Finally, to close the model, the wage is determined by equilibrating labor supply and demand $n_1^s = n_1^d$. We introduce a fictitious labor market auctioneer who sets

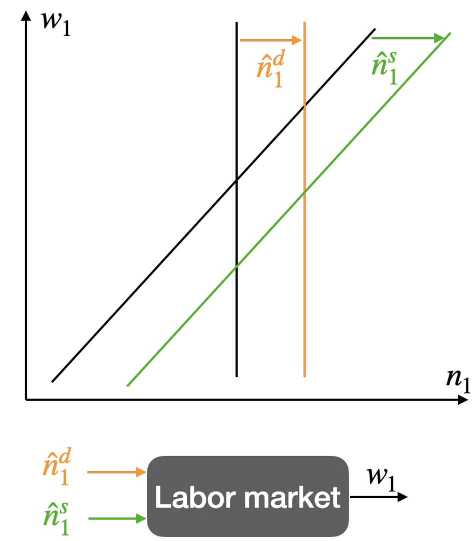


Figure 2: Competitive labor market

Notes: This figure depicts the labor market which gives rise to the real wage w_1 that balances supply shift $\hat{n}_1^s$ and demand shift $\hat{n}_1^d$. In the New Keynesian model, the labor demand curve is inelastic since firms only use labor in production.

the wage as the intersection of supply and demand curves,

$$\begin{aligned} \hat{n}_1^s + e_{n^s w} w_1 &= \hat{n}_1^d + e_{n^d w} w_1 \\ w_1 &= \left(e_{n^s w} - e_{n^d w}\right)^{-1} \left(\hat{n}_1^d - \hat{n}_1^s\right) \end{aligned} \tag{7}$$

This captures Walras's idea of tâtonnement, where the auctioneer raises the wage if there is excess demand for labor and reduces it otherwise, as illustrated in Figure 2.

Period-1 equilibrium. In period 1, agents observe all variables and best respond and the labor market clears. The equilibrium is characterized by (1-7).

2.2 General Equilibrium as a System of Causal Relations

We conceptualize the period-1 general equilibrium (GE) in the New Keynesian economy as a system of causal relations, and represent it in a directed graph. This conceptualization of GE as a causal system is our broader theoretical contribution, encompassing not only the New Keynesian model but also other potential applications.

The period-1 equilibrium in response to the cost-push shock ϵ_1^π and the monetary policy shock ϵ_1^i is fully characterized by (1-7). We collect all macroeconomic variables in a vector $V_1 \equiv \left(i_1, \pi_1, div_1, \hat{n}_1^d, c_1, \hat{n}_1^s, w_1\right)'$ and the two shocks correspondingly in $S_1 \equiv \left(\epsilon_1^i, \epsilon_1^\pi, 0, 0, 0, 0, 0\right)'$. While we focus on these two shocks in this paper, it is straightforward to incorporate additional shocks.

To capture the causal relations in the economy, we define M as a matrix of all partial derivatives among the macroeconomic variables in (1-7), detailed in (C8). Each element of M is a *partial derivative* that describes how one variable responds to another variable in period 1, which we define as a *causal relation*. A causal relation arises from either agents' best responses to decision-relevant variables (e.g., consumption c_1 responding to the interest rate i_1) or the determination of prices from supply and demand, i.e., the Walrasian auctioneer's response. We characterize the period-1 equilibrium as follows.

Proposition 1. (GE as a system of causal relations) *The period-1 New Keynesian economy is characterized by the fixed point to the system of causal relations among all* agents' actions and competitive prices, $V_1 \equiv \left(i_1, \pi_1, div_1, \hat{n}_1^d, c_1, \hat{n}_1^s, w_1\right)'$*, as*

$$\underbrace{V_1}_{\text{variables}} = \underbrace{M}_{\text{causal relations}} V_1 + \underbrace{S_1}_{\text{shocks}} \tag{8}$$

The equilibrium can be solved to yield

$$V_1 = (I - M)^{-1} S_1 = \sum_{n=1}^{\infty} M^{n-1} S_1 \tag{9}$$

which is a sum of all effects of varying distance, where each $M^{n-1}S_1$ term is an n-step effect of a shock on a variable via $n-1$ intermediate variables.

Each equation in system (8) describes how an outcome on the left depends on a set of causes on the right, where S_1 represents the direct (or *partial equilibrium*) effects of shocks and MV_1 captures the indirect (or *general equilibrium*) effects.[13] We visualize this causal system as a directed graph in Figure 1, formally supporting the intuition outlined in the introduction. Each node represents a type of agent (including the fictitious labor market auctioneer), and each arrow indicates a macroeconomic variable—either an agent's decision or the real wage determined in the competitive labor market.

[13] To incorporate shocks that directly affect multiple variables (e.g., a household preference shock impacting both consumption and labor supply), we can rewrite (8) as $V_1 = MV_1 + J^{direct} S_1$ in which J^{direct} captures the direct effects of each shock on a subset of variables and S_1 may differ in dimensionality from V_1.

Formula (9) expresses the solution to (8) as the sum of all effects of varying distance. The 1-step effect S_1 is the direct (or partial equilibrium) effect of shocks, while all subsequent terms represent indirect (or general equilibrium) effects of varying distance.

As these period-1 shocks are observed in period 0, if agents are rational, they will correctly forecast the period-1 equilibrium, i.e., $\mathbb{E}_0^{rational}[V_1] = V_1$. In this sense, the rational expectations hypothesis assumes that agents can take infinite steps in this graph to converge to the fixed point. Next, we introduce our theory of shallow thinking.

2.3 Shallow Thinking

Motivated by evidence from behavioral economics and psychology, we assume that agents foresee only a finite number of steps in shock propagation in the directed graph. We outline two key assumptions that shape agents' heterogeneous beliefs and lead to a parsimonious characterization of the average beliefs.

Assumption 1. Individuals vary in their finite *depth of thinking* $d \in \mathbb{N}^+$, with expectations

$$\mathbb{E}_0^d[V_1] \equiv \sum_{n=1}^{d} M^{n-1} S_1 \tag{10}$$

which implies an iterative formula for $d > 1$ as

$$\mathbb{E}_0^d[V_1] = M\mathbb{E}_0^{d-1}[V_1] + S_1 \tag{11}$$

Definition (10) formalizes the idea that a depth-d agent only understands the effects of shocks that take no more than d steps. The iterative formula (11) suggests that a depth-d agent can think one step further compared to a depth-$(d-1)$ agent.[14]

For example, in response to a cost-push shock ϵ_1^π (Figure 1), a depth-1 agent acknowledges only the most obvious implication: firms will raise their prices (causing inflation π_1). A depth-2 agent can further appreciate that the central bank will raise the interest rate i_1 in response to higher inflation. A depth-3 agent understands that a higher interest rate i_1 will discourage household consumption and incentivize labor supply. A depth-4 agent

[14] The iterative formula (11) can also be interpreted as level-k thinking in network games, as studied by Kneeland (2015) and Ballester, Rodriguez-Moral and Vorsatz (2024), where each player (households, firms, the central bank, and the labor market auctioneer in our model) best responds to a subset of other players. A depth-1 player believes that only the player directly affected by shocks will change their actions, thus expecting $\mathbb{E}_0^1[V_1] = S_1$. A depth-$d$ player, with $d > 1$, assumes all other players are of depth $(d-1)$.

recognizes that changes in household behavior will affect the firms and the labor market. This iteration continues infinitely, and only a depth-∞ (i.e., rational) agent correctly assesses the strength of all loops and accurately forecasts the period-1 equilibrium.

Remarks on representing GE and shallow thinking. By definition, shallow agents do not appreciate the fixed point, and their beliefs depend on the causal representation of general equilibrium. We emphasize a specific representation underlying shallow thinking. Once we cast competitive markets as fictitious Walrasian auctioneers that determine prices, there is one unique representation of general equilibrium that consists of the responses of all agents (including fictitious Walrasian auctioneers) to decision-relevant variables, as captured by (8) in Proposition 1. There are alternative representations that lead to the same fixed point as (8). However, all these alternatives implicitly mix in relations that *only hold in equilibrium,* and do not reflect best responses.

For example, a popular way of representing the textbook New Keynesian model is to use three variables—the interest rate i_1, inflation π_1, and consumption c_1 (i.e., output), as[15]

$$\begin{pmatrix} i_1 \\ \pi_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} 0 & \phi & 0 \\ 0 & 0 & \kappa(\sigma+\nu) \\ -\sigma & 0 & 0 \end{pmatrix} \begin{pmatrix} i_1 \\ \pi_1 \\ c_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1^i \\ \theta^{-1}\epsilon_1^\pi \\ 0 \end{pmatrix} \tag{12}$$

This representation expresses inflation π_1 as a function of consumption c_1, rather than in the best response form (3), by canceling the pricing complementarity term π_1 on the right-hand side with the left-hand side in (3) and implicitly substituting the equilibrium value of the wage, thereby failing to reflect firms' best responses. While (12) shares the same fixed point in terms of (i_1, π_1, c_1) as our representation (8), it is not a system of best responses and involves a different iterative process to arrive at the fixed point.

Belief heterogeneity and average beliefs. The iterative process of shallow thinking leads to a prediction about belief heterogeneity, which allows us to measure the distribution of d across the population in Section 3. Our theory predicts that agents with a low d only perceive changes in variables that are close to shocks in the directed graph. And the

[15] To derive this representation, we first solve for the other variables $V_1^{other} = \left(div_1, \hat{n}_1^d, \hat{n}_1^s, w_1\right)$ in (8) in terms of (i_1, π_1, c_1). Substituting V_1^{other} into the consumption function (4) yields $c_1 = -\sigma i_1$. Then, substituting V_1^{other} into the Phillips curve (3) and using $c_1 = -\sigma i_1$ to eliminate i_1, we obtain $\pi_1 = \kappa(\sigma+\nu)c_1 + \theta^{-1}\epsilon_1^\pi$, which expresses the inflation solely as a function of output. The Taylor rule (6) is kept unchanged.

set of variables that they perceive to change varies with shocks. By examining people's expectations of changes in different macroeconomic variables to various shocks, we can measure the distribution of d, which we formalize in Section 3.1.

We capture the rich belief heterogeneity arising from heterogeneous depths of thinking with a parsimonious parametric assumption.

Assumption 2. Individual depth of thinking d follows a geometric distribution over $\mathbb{N}^+$ with continuation rate $\lambda \in [0, 1]$, i.e.,

$$\mathbb{P}(d \geq n) = \lambda^{n-1}, \ \forall n \in \mathbb{N}^+ \tag{13}$$

We assume that everyone can take at least one step. A λ share of them take at least two, a λ^2 share take at least three, and so on. A higher shallow thinking parameter λ means that individuals are deeper on average, with $\frac{1}{1-\lambda}$ representing the average depth of thinking,[16] and $\lambda = 1$ nesting the rational expectations hypothesis (i.e., people having infinite depth of thinking). With this parametric assumption, we could aggregate heterogenous beliefs into average beliefs, which will drive the period-0 equilibrium as we analyze in Section 4.

Proposition 2. (Average beliefs) *The average beliefs* $\overline{\mathbb{E}}_0 [V_1] \equiv \sum_{n=1}^{\infty} \mathbb{P}(d = n) \cdot \mathbb{E}_0^n [V_1]$ *are sums of all effects of shocks of varying distance*

$$\overline{\mathbb{E}}_0 [V_1] = \sum_{n=1}^{\infty} \lambda^{n-1} M^{n-1} S_1 \tag{14}$$

Further, the average beliefs satisfy a fixed point

$$\overline{\mathbb{E}}_0 [V_1] = \underbrace{\lambda M}_{\text{average perceived causal relations}} \overline{\mathbb{E}}_0 [V_1] + S_1 \tag{15}$$

Equation (14) is comparable to (9) that expresses the equilibrium as a sum of all effects of varying distance, but with more distant effects dampened more, since fewer people appreciate them. Equation (15) further suggests that, the average beliefs are formed *as if* as a fixed point by a representative agent who knows all causal relations in M but underappreciates them by a factor λ, parallel to (8) that characterizes the equilibrium as a fixed point. The proof is simple, by summing all n-step effects with decaying weights.

[16]The expectation of a geometric distribution is $\mathbb{E}[d] \equiv \sum_{n=1}^{\infty} \mathbb{P}(d = n) \cdot n = \sum_{n=1}^{\infty} \mathbb{P}(d \geq n) = \frac{1}{1-\lambda}$.

Moreover, equation (15) coincides exactly with the formula of imperfect mental models in Wu (2023). That paper extracts an empirical moment based on this formula using existing forecasts data and rejects the null of $\lambda = 1$ (rational expectations). The nature of the Wu (2023) exercise is *quantitative*, as it compares forecasts to the true conditional expectations. In this paper, we provide *qualitative* evidence to support our theory based on heterogeneity in beliefs being directionally correct, elicited in a customized survey.

Remarks on assumptions. In summary, the idea of shallow thinking is that people understand only a limited number of steps in shock propagation, as captured by Assumption 1, which is the backbone of our theory. Assumption 2 serves as a convenient aggregator to generate average beliefs. Its nature is parametric rather than conceptual, akin to how Calvo pricing is a useful parametrization of nominal rigidity but not essential. With these two assumptions, one can generate heterogeneous and average beliefs in a macroeconomic model with a single additional parameter, λ.

If the model also satisfies an additional assumption introduced next, we can estimate λ with a panel regression using a survey that asks respondents to forecast the directional responses of different macroeconomic variables to various shocks.

3 Measuring Shallow Thinking

In this section, we test shallow thinking and estimate λ using a survey. Readers more interested in the consequences of shallow thinking may choose to skip this section.

We formalize the theoretical prediction that changes in variables more distant from shocks are perceived by fewer people in Section 3.1, which enables us to measure λ in Section 3.2. Before that, we provide a brief overview of our survey design, which examines people's understanding of shock propagation, with details in Appendix A.1.

Overview of survey design. We assess people's understanding of shock propagation by asking them to forecast the directional responses of key macroeconomic variables to hypothetical shocks. We conducted an online survey of 1,000 respondents in the US who are broadly representative of the US population.[17]

[17]Due to resource constraints, we run our survey on US households and apply the calibrated λ to both households and firms in our model. Notably, firm managers are largely uninformed about recent aggregate inflation or monetary policy and have inflation expectations that are far from anchored, similar

We study six classic macroeconomic shocks in three groups: oil price shocks (oil) and monetary policy shocks (MP) as group 1, government spending shocks (G) and personal income tax shocks (PIT) as group 2, and corporate income tax shocks (CIT) and transfer payment shocks (TP) as group 3. Half of the respondents are randomly assigned to group 1, which includes the two shocks analyzed in the model, while a quarter are assigned to each of groups 2 and 3 for additional evidence. For each shock, respondents provide directional forecasts of a set of macroeconomic variables, such as inflation and the interest rate. For each variable, respondents select "up," "down," "unchanged," or "I don't know" to indicate the expected change in response to the shock over the next 12 months.

Table 1 shows the baseline specification with eleven macroeconomic variables, their correct directional responses, and variable distance D_{vs}. We obtain the true direction of change for each variable-shock combination from the empirical literature, reviewed in Table B2, and determine accordingly whether each respondent is correct. This baseline specification includes variable-shock combinations whose responses are empirically and intuitively uncontroversial. The baseline variable distance D_{vs} is derived from our New Keynesian model, enriched with decreasing-returns production and a Taylor rule dependent on both inflation and unemployment. We consider various robustness checks for the selection of variable-shock combinations (v, s) and variable distance D_{vs} in Tables B3 and B4.

3.1 Empirical Content of Shallow Thinking

We formally establish the theoretical prediction that changes in more distant variables are understood by fewer people, with an additional assumption.

To set the stage, we index respondents by n, variables by v, and shocks by s in our survey. We capture the correctness of respondents' directional forecasts, based on the true directional changes from the empirical literature, using an indicator 1_{nvs} as follows.

Definition 1. We define *correct directional belief* 1_{nvs} as 1 if respondent n correctly forecasts the directional response of variable v to shock s, and 0 otherwise.

As discussed earlier, a shallow agent perceives changes only in variables that are close to shocks. When aggregated across the population, since people are of heterogeneous

to households (Candia, Coibion and Gorodnichenko, 2024; Coibion, Gorodnichenko and Kumar, 2018). Furthermore, professional forecasters also underappreciate the causal relations in the economy (Wu, 2023), including the response of monetary policy to economic conditions (Cieslak, 2018).

Table 1: Baseline version of variable distance and correct directions

	Group 1 (50%)		Group 2 (25%)		Group 3 (25%)	
	Oil $\uparrow$	MP $\uparrow$	G $\uparrow$	PIT $\uparrow$	CIT $\uparrow$	TP $\uparrow$
Output	3$\downarrow$	2$\downarrow$	1$\uparrow$	2$\downarrow$	3$\downarrow$	2$\uparrow$
Interest rate	2$\uparrow$	1$\uparrow$				
Price	1$\uparrow$	3$\downarrow$	2$\uparrow$	3$\downarrow$		
Unemployment	3$\uparrow$	2$\uparrow$	2$\downarrow$	2$\uparrow$	3$\uparrow$	3$\downarrow$
Labor hours	3$\downarrow$	2$\downarrow$	2$\uparrow$	2$\downarrow$	3$\downarrow$	3$\uparrow$
Durable consumption	3$\downarrow$	2$\downarrow$	2$\downarrow$	2$\downarrow$	3$\downarrow$	2$\uparrow$
Non-durable consumption	3$\downarrow$	2$\downarrow$	2$\downarrow$	2$\downarrow$	3$\downarrow$	2$\uparrow$
Dividend					2$\downarrow$	2$\uparrow$
Personal income tax			1$\uparrow$	1$\uparrow$		
Corporate income tax			1$\uparrow$		1$\uparrow$	
Government borrowing			1$\uparrow$	1$\downarrow$	1$\downarrow$	

Notes: Six shocks are oil price shock (oil), monetary policy shock (MP), government spending shock (G), personal income tax shock (PIT), corporate income tax shock (CIT), and transfer payment shock (TP). The latter four all concern the federal government. Each cell indicates the variable distance D_{vs} and the directional response (up or down) in the baseline specification. The directional responses are from the empirical literature reviewed in Table B2, and robustness versions of selection of variable-shock combinations and variable distance are in Tables B3 and B4.

depths of thinking, changes in variables further removed from shocks are understood by fewer people. This is a prediction about the correct directional belief 1_{nvs} at the population level, without the need to determine the depth of thinking d for each individual, which facilitates our empirical test.

In order to formally define the distance of a variable relative to a shock for our test, we introduce some additional notation. Notice that beliefs $\mathbb{E}^d[V_1]$ are linear in the shocks S_1, as determined in (10). With only a slight abuse of notation, we let $v \in V_1$ be a variable in our model and $s \in S_1$ be a shock. Thus, $\frac{\partial v}{\partial s}$ is the true sensitivity of variable v to shock s, whereas $\frac{\partial \mathbb{E}^d[v]}{\partial s}$ is the perceived sensitivity by a depth-d individual. And their signs indicate the true and perceived directional responses of variables to shocks, respectively.

Definition 2. We define the *variable distance* D_{vs} as the minimum d such that $\frac{\partial \mathbb{E}^d[v]}{\partial s}$ has the same sign as $\frac{\partial v}{\partial s}$, for each variable $v \in V_1$ and shock $s \in S_1$.

That is, variable distance D_{vs} corresponds to the depth of the shallowest individual who can correctly perceive the directional response of v to s. In our example with transitory cost-push and monetary policy shocks, D_{vs} equals the depth of the shallowest agent who perceives any change of v in response to s. That is, $\frac{\partial \mathbb{E}^d[v]}{\partial s}$ is 0 for all $d < D_{vs}$. Nonetheless, Definition 2 is more general when applied to persistent shocks and other models.

Assumption 3. Model parameters M are such that $\frac{\partial \mathbb{E}^d[v]}{\partial s}$ has the same sign as $\frac{\partial v}{\partial s}$ for all $d \geq D_{vs}$.

Assumption 3 holds true when more distant causal relations either amplify or offset the impact of the shock, once the correct direction is established, but do no overturn it. It is only useful for the next proposition that offers a reduced-form estimation of λ. This assumption is true in the New Keynesian model we study under a standard calibration. For instance, in response to the cost-push shock ϵ_1^π, the central bank will raise the interest rate i_1 to offset the shock, but does not lead to deflation. That is, the perceived inflation response, $\frac{\partial \mathbb{E}^d[\pi_1]}{\partial \epsilon_1^\pi}$, is positive for all $d \geq 1$. And since $\mathbb{E}^d[i_1] = \phi \mathbb{E}^{d-1}[\pi_1]$ from (11), the perceived interest rate response, $\frac{\partial \mathbb{E}^d[i_1]}{\partial \epsilon_1^\pi}$, is positive for all $d \geq 2$. Further, even if it is not true for all variable-shock combinations, as long as there exists a subset of such combinations with varying D_{vs}, our estimation can go through by focusing on this subset.

Under Assumption 3, any agent with a depth of thinking d greater than or equal to the variable distance D_{vs} will correctly forecast the directional change of variable v in response to shock s, while any agent with $d < D_{vs}$ will not. *Thus, the expectation of 1_{nvs} conditional on variable distance D_{vs} is the share of respondents with depth $d \geq D_{vs}$*, which is $\lambda^{D_{vs}-1}$ given Assumption 2, suggesting a reduced-form estimation of λ as follows.

Proposition 3. (Heterogeneity in correct directional beliefs) *The expectation of correct directional belief 1_{nvs}, conditional on variable distance D_{vs}, in the population is*

$$\mathbb{E}^{population}\left[1_{nvs} | D_{vs} = D\right] = \lambda^{D-1}, \ \forall D \in \mathbb{N}^+ \tag{16}$$

where $\mathbb{E}^{population}$ denotes the expectation in the population of survey respondents. Consequently,

1. *an ordinary least squares estimation of $1_{nvs} = \gamma D_{vs} + \alpha + \epsilon_{nvs}$ yields a negative slope γ;*
2. *a nonlinear least squares estimation of $1_{nvs} = b_1 \cdot b_2^{D_{vs}-1} + b_0 + \epsilon_{nvs}$ identifies λ with b_2.*

We consider both the nonlinear and the linear specifications. The null of $\gamma = 0$ and $b_2 = 1$ includes rational expectations and any other theories of beliefs that do not correlate with variable distance D_{vs}. Our estimation result in Section 3.2 will show that γ is negative and b_2 is lower than 1, both with high levels of statistical significance.

The nonlinear specification lets us estimate λ from a regression. Equation (16) suggests that, under our three assumptions, the conditional expectation of 1_{nvs}, which is the conditional probability of making correct directional forecasts, is exponentially decaying. Thus a nonlinear least-squares estimation of an exponential function can exactly recover λ. Our Assumption 3 crucially facilitates this estimation. As discussed earlier, if Assumption 3 does not hold for all possible combinations of variables and shocks, as long as one can find a subset of such combinations with varying D_{vs}, one can still estimate λ with the nonlinear regression on this subset. If even that is not possible, one can estimate λ by minimizing distance between the distribution of measured 1_{nvs} and the corresponding theory-implied distribution, as λ parametrizes the latter distribution.

A linear specification is valuable for two reasons: (i) it allows us to empirically control for fixed effects to purge confounding sources of belief heterogeneity, and we will show that our coefficient of interest γ is indeed robust to such controls; and (ii) it does not hinge on the parametric Assumption 2. A negative γ by itself indicates that some agents only perceive changes in variables close to shocks. When Assumption 2 does hold, the estimated slope γ is a weighted average of the local slopes of the nonlinear function λ^{D-1}.

3.2 Variable Distance Predicts Correct Directional Belief

We examine the predictability of correct directional belief 1_{nvs} by variable distance D_{vs} as prescribed by Proposition 3.

Figure 3 shows the expectation of correct directional belief 1_{nvs}, conditional on variable distance D_{vs}, together with the 99.9% confidence interval. The blue dot indicates the conditional expectation. The red diamond represents the conditional expectation, after controlling for individual-by-variable and individual-by-shock fixed effects δ_{nv}, δ_{ns}, the purpose of which we discuss soon.[18] The conditional expectation of 1_{nvs} declines drastically in variable distance D_{vs} in both cases.

Remarks on identification and relevance of shallow thinking. With Figure 3, we make three remarks. First, there is a significant decline in the share of correct directional belief 1_{nvs} from step 2 to step 3. Thus, the key parameter λ will not be solely identified by comparing step-1 variables with others. One could intuitively expect such a difference, since step-1 variables are directly shocked, while others are only indirectly affected in general equilibrium. This drop from step 2 to step 3 substantiates the key theoretical implication that among general equilibrium effects, some are better understood.

[18] The conditional expectations and confidence intervals, with and without fixed effects, are produced following Cattaneo et al. (2024).

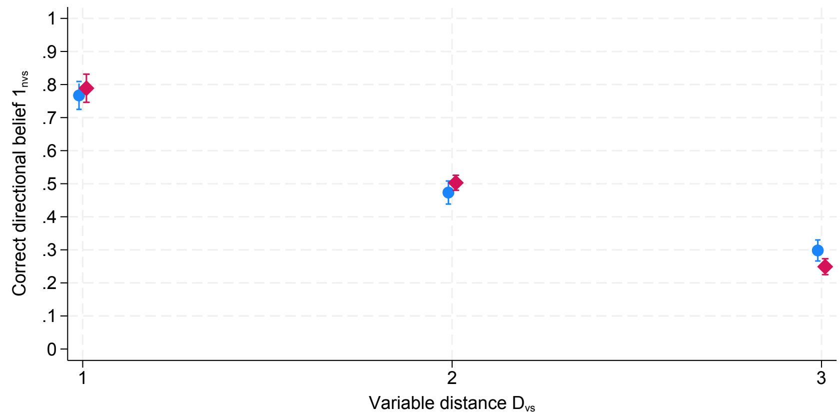


Figure 3: Expectation of correct directional belief 1_{nvs} conditional on variable distance D_{vs}

Notes: This figure shows the expectation of correct directional belief 1_{nvs}, conditional on variable distance D_{vs}, together with the 99.9% confidence interval. The blue dot indicates the conditional expectation. The red diamond represents the conditional expectation, controlling for individual-by-variable and individual-by-shock fixed effects δ_{nv}, δ_{ns}. Correct directional belief 1_{nvs} equals 1 if respondent n correctly forecasts the directional response of variable v to shock s, and 0 otherwise. Variable distance D_{vs} is derived from the New Keynesian model. Table 1 lists the correct directional responses and variable distance.

Second, the conditional expectation $\mathbb{E}^{population}[1_{nvs}|D_{vs} = 3]$, which is also the share of respondents with depth of thinking $d \geq 3$, at only around 30% suggests qualitative validity of our exercise. Since it takes only a few steps to reach all variables in the New Keynesian model, we cannot distinguish among very deep thinkers (e.g., $d = 99$ vs. $d = 100$). Had the decline in 1_{nvs} with variable distance been minor, we would be estimating the distribution of depth of thinking d from only the leftmost points in its range. However, the sharp decline—where the majority of respondents fail to take the third step—implies a population of fairly shallow agents, lending credibility to our estimation.

Last, regarding the relevance of our theory compared to alternatives, we focus on limited depth of thinking while assuming, for simplicity, that agents know the true causal links and model parameters. The steep decline in Figure 3 suggests this focus captures a key aspect of belief formation. If people could iterate their models infinitely many times but believed in different causal models, these models would have to be quite wrong—and wrong in a way that correlates with variable distance in the New Keynesian model to explain Figure 3. Alternatively, agents could iterate the same model infinitely but disagree on parameters (such as the slope of the Phillips curve). However, this would not clearly explain why they misjudge directional responses or why such errors correlate with variable distance. Moreover, our calibration suggests that more than half the population (specifically, $1-\lambda^2 \approx 0.63$ share) fail to consider the third step of propagation. For example, in response to cost-push shocks, they perceive changes only in inflation and interest rates, the latter via the Taylor rule, while ignoring all other causal relations. Hence, potentially differing perceptions about more distant causal relations are relevant only for a small fraction of the population, limiting their overall importance. Later in the paper, we present formulas suggesting that the perceived strength of general equilibrium feedback loops declines exponentially with their length under shallow thinking. In contrast, other model parameters (such as the Taylor rule coefficient) and any misperceptions thereof only affect it proportionally.

Estimation results. Table 2 presents various specifications of the ordinary least squares (OLS) regression

$$1_{nvs} = \gamma D_{vs} + \alpha + \delta_{nv} + \delta_{ns} + \epsilon_{nvs} \tag{17}$$

and the nonlinear least squares (NLS) regression

$$1_{nvs} = b_1 \cdot b_2^{D_{vs}-1} + b_0 + \epsilon_{nvs} \tag{18}$$

on correct directional belief 1_{nvs} on variable distance D_{vs} as prescribed by Proposition 3.

The coefficient γ from the linear regression tests for the theory-implied pattern that changes in more distant variables are understood by fewer people.[19] The null of $\gamma = 0$ includes rational expectations and any other theory of beliefs that does not correlate with distance D_{vs}. Further, if respondents are totally clueless about the economy and give random answers in our survey, that will not be reflected in γ. Thus, a negative γ not only implies that people make mistakes, but they do so in a distance-dependent way.

Column (1) uses variable distance D_{vs} as the only predictor and finds a statistically significant coefficient with a R^2 of 10%. Column (2) shows that individual fixed effects matter too, increasing the R^2 to 23%. That means some people are more likely to be correct than others, as our theory postulates. For all columns, standard errors are clustered at the individual level, since they may correlate across all answers submitted by an individual.

Column (3) shows that the slope estimate and its statistical significance are robust to

[19] Interestingly, in a very different context, using network data of relationships from Indian villages, Breza, Chandrasekhar and Tahbaz-Salehi (2018) show that the knowledge of whether certain pairs of households are linked declines steeply in the pair's network distance to the respondent. The distance in our context is distinct—it is a conceptual measure of how relatable a variable is to a shock under limited causal reasoning.

Table 2: Regression of correct directional belief 1_{nvs} against variable distance D_{vs}

	OLS					NLS
Correct directional belief 1_{nvs}	(1)	(2)	(3)	(4)	(5)	(6)
Variable distance D_{vs}	-0.22***	-0.24***	-0.24***	-0.27***		
	(0.01)	(0.01)	(0.01)	(0.01)		
$1_{D_{vs}=2}$					-0.29***	
					(0.02)	
$1_{D_{vs}\geq 3}$					-0.54***	
					(0.02)	
$b_2 - 1$						-0.39***
						(0.05)
Observations	10763	10763	10763	10763	10763	10763
R^2	0.10	0.23	0.30	0.63	0.63	0.11
Individual FE		Yes	Yes	Absorbed	Absorbed	
Variable FE			Yes	Absorbed	Absorbed	
Shock FE			Yes	Absorbed	Absorbed	
Individual-variable FE				Yes	Yes	
Individual-shock FE				Yes	Yes	

Standard errors in parentheses

Standard errors clustered at individual level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows the regression results of correct directional belief 1_{nvs} against variable distance D_{vs}, using the ordinary least squares (OLS) specification (17) and the nonlinear least squares (NLS) specification (18). Correct directional belief 1_{nvs} equals 1 if respondent n correctly forecasts the directional response of variable v to shock s and 0 otherwise. Variable distance D_{vs} is derived from a New Keynesian model. The OLS specification tests the null hypothesis that the slope is 0, controlling for individual-by-variable and individual-by-shock fixed effects δ_{nv}, δ_{ns}. The NLS specification identifies λ with b_2, with a null hypothesis of $b_2 = 1$. Hence, we show the estimate $b_2 - 1$ and the associated p-value.

the inclusion of variable and shock fixed effects. Controlling for these additional fixed effects addresses a concern that people may understand some variables or some shocks better in ways that happen to correlate with distance. Several theories of agents' optimal behavior with constraints on their perceptions, such as rational inattention (Sims, 2003), concern about robustness to model misspecification (Hansen and Sargent, 2007), and sparsity (Gabaix, 2014), share the idea that agents optimize their perceptions of variables or shocks most relevant for their decisions. Our findings show that, for the same variable, its change is understood more poorly when it is further away from a shock.[20]

[20]In Table B5, we run the same regression on subsamples of decision-relevant and decision-irrelevant variables separately. We find that, while the intercept on the decision-relevant subsample is indeed higher, indicating a higher unconditional probability of correct directional forecast, the slope is also higher, which suggests a faster decline in variable distance. Further, once we include individual-variable and individual-

Another concern is that some shocks (like monetary policy shocks) may take longer to transmit into the economy or some variables may respond more slowly, which may lead people to predict no changes over a fixed horizon. These possibilities are absorbed by shock and variable fixed effects too.[21]

Column (4) further controls for individual-by-variable or individual-by-shock fixed effects. They absorb potentially confounding sources of belief heterogeneity. For example, if a person believes in a post-pandemic quantity-constrained model of the economy, they will predict that prices respond to all shocks but quantities are fixed. Another person can instead believe in a price-constrained model. People may also differ in their perceptions of decision-relevant variables due to varying cognitive capacities, as in Sims (2003) and Gabaix (2014), or differing levels of concern about robustness, as in Hansen and Sargent (2007). Such heterogeneity is absorbed by individual-by-variable fixed effects. Similarly, individual-by-shock fixed effects account for the possibility that one person only understands monetary policy shocks whereas another only understands oil shocks.

Column (5) demonstrates that, relative to step-1 variables (that are directly shocked), step-2 variables are understood by fewer people, and step-3-and-above variables by even fewer, confirming what we observe from Figure 3.

Last, column (6) shows the nonlinear estimation and strongly rejects the null hypothesis of $b_2 = 1$. The estimation suggests that $\hat{\lambda} \approx 0.61$. That means, people on average think about only 2.6 steps, even under our parametric assumption that there is a distribution of people who could reason more than three steps. We use this value to apply our theory in workhorse macroeconomic models.

We conduct various robustness checks in Appendix B.2. We offer further evidence in Appendix B.3 to suggest that the limited depth of thinking vis-à-vis causal relations is an individual characteristic and reflects limited knowledge about the macroeconomy.

Calibration of shallow thinking parameter λ. We take the estimate from column (6), $\lambda = 0.61$, as our baseline calibration. This is identified from the declining pattern of correct directional belief 1_{nvs} against variable distance D_{vs}. Alternatively, one could infer λ from the average probability of correctly forecasting step-2 variables, $\mathbb{E}^{population}[1_{nvs}|D_{vs} = 2]$, based on (16). That is approximately 0.5, as shown in Figure 3. Under this alternative calibration, on average, people, understand only two steps.

shock fixed effects, the intercepts and slopes on the two subsamples become statistically indistinguishable.

[21]We also note that we purposefully select variables that have statistically significant impulse responses to shocks at the 12-month horizon for our specification.

The difference relates to a slight discrepancy between Assumption 2, which assumes that everyone correctly understands the directional responses of step-1 variables, and the empirical finding in Figure 3, which shows that they mostly, but not always, do. This discrepancy is quantitatively unimportant and may arise from respondents' occasional misunderstanding of, or inattention to, certain survey questions, as any noise in responses only lowers the indicator 1_{nvs}. We proceed with $\lambda = 0.61$ as our baseline calibration to err on the side of rationality, though we note that the consequences of shallow thinking are qualitatively similar and quantitatively stronger with a lower λ.

4 Consequences of Transitory News Shocks

We discuss belief under- and overreaction to shocks due to shallow thinking, and the consequences for asset prices and the macroeconomy in the New Keynesian model, in the case with transitory news shocks introduced in Section 2.

We make an important remark on the generality of analyzing such transitory news shocks: while we compare shallow thinking against rational expectations regarding news about period-1 shocks, the same comparison holds for persistent shocks that materialize in period 0 and last for 2 periods. That is simply because in the log-linearized economy, a 2-period persistent shock is equivalent to the sum of a period-0 shock and a period-1 shock that is observed in period 0. The economy's response to a period-0 shock is independent of agents' belief formation. Thus the comparison across theories of expectations regarding any shocks that last for 2 periods is solely driven by its news shock component.

We follow a standard quarterly calibration of the New Keynesian economy, with all parameters listed in Table 3.

4.1 Inflation Expectations and Long-Term Interest Rates

We study belief under- and overreaction in response to news about the cost-push shock ϵ_1^π and the monetary policy shock ϵ_1^i, and show that shallow thinking reconciles seemingly opposing empirical findings on the misreaction of long-term interest rates. We analyze the two shocks in sequence and present a synthesis afterward.

We assume that the yield of a 2-period bond $y_0^{(2)}$, i.e., the long-term yield, is determined

Table 3: Quarterly calibration of the New Keynesian economy

Parameter	Description	Value	Estimate/Target
Beliefs			
λ	Continuation rate of depth of thinking	0.61	Our survey evidence
Firms			
θ	Price stickiness	0.75	Average price duration of 1 year
κ	Phillips curve slope	0.086	$\kappa = \theta^{-1}(1-\theta)(1-\beta\theta)$
μ	Steady state markup	1.1	
Households			
β	Discount factor	0.99	Steady state annual $\bar{r} = 4\%$
σ^{-1}	Elasticity of intertemporal substitution (EIS)	1	
ν^{-1}	Frisch elasticity of labor supply	0.5	
Central bank			
ϕ	Taylor rule coefficient	1.5	

by the expectations hypothesis as[22]

$$y_0^{(2)} = \frac{i_0 + \overline{\mathbb{E}}_0 [i_1]}{2} \tag{19}$$

Cost-push news shocks. Proposition 4 characterizes the period-1 equilibrium in response to the cost-push shock ϵ_1^π and period-0 expectations thereof.

Proposition 4. (Period-1 cost-push shock) *The period-1 equilibrium in response to a cost-push shock ϵ_1^π features*

$$\pi_1 = \frac{1}{1 - \underbrace{(1-\theta)}_{\text{pricing complementarity}} + \underbrace{\phi\theta\kappa}_{\text{monetary policy loop}} \cdot \underbrace{\sigma^{-1}(\nu+\sigma)}_{\text{Keynesian cross}}} \epsilon_1^\pi \tag{20}$$

$$i_1 = \phi\pi_1 \tag{21}$$

whereas the period-0 average expectations upon observing the news about ϵ_1^π are

$$\overline{\mathbb{E}}_0 [\pi_1] = \frac{1}{1 - \lambda(1-\theta) + \lambda^4 \phi\theta\kappa K_1(\lambda)} \epsilon_1^\pi \tag{22}$$

$$\overline{\mathbb{E}}_0 [i_1] = \lambda\phi\overline{\mathbb{E}}_0 [\pi_1] \tag{23}$$

with $K_1(\lambda)$ increasing in λ under our calibration and $K_1(1) = \sigma^{-1}(\nu+\sigma)$.

[22]To microfound this in our model without aggregate risks, we assume that an intermediary prices the 2-period bond on behalf of all households by averaging their beliefs. We also assume that, like the 1-period bond, the 2-period bond is in zero supply, allowing us to discuss its pricing without impacting the economy.

To understand these results, we start with equilibrium inflation (20) and the average inflation expectation (22). The period-1 equilibrium is independent of agents' belief formation, as (20) is independent of λ. It coincides with rational expectations, i.e., (22) under $\lambda = 1$. Recall that both the equilibrium and the average expectations are the sum of all n-step effects, as established in (9) and (14), with more distant effects dampened more for expectations. We organize all these effects into three groups that involve different loops, color-coded in Figure 4, and inflation π_1 is directly involved in two of them.

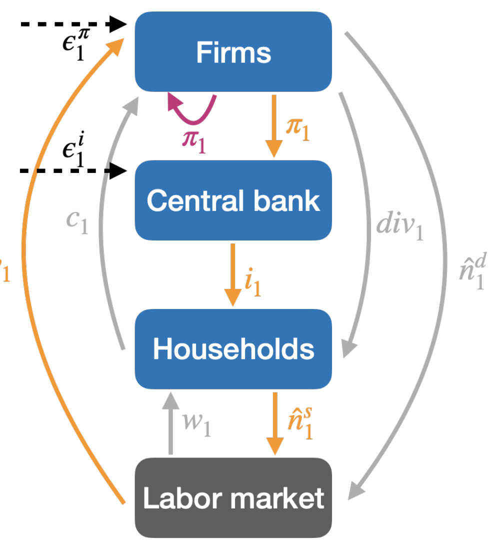


Figure 4: Causal relations of period-1 economy

Notes: This figure illustrates the three loops of causal relations in the period-1 economy in different colors, to accompany the discussion of Propositions 4 and 5. Three loops are the pricing complementarity self-loop (in purple), the monetary policy loop (in orange) and the Keynesian cross (in gray).

The first loop is a self-loop of *pricing complementarity*, shown in purple. A higher inflation π_1 incentivizes all firms to price higher, thus amplifying itself. This effect has a strength of $(1 - \theta)$. As we sum up the infinite series going through this loop, its strength appears in the denominator of π_1 in (20). It is dampened by λ in $\overline{\mathbb{E}}_0[\pi_1]$ in (22) since, with each loop, only a λ share of people remain.

The second one is a 4-step loop involving *monetary policy*, shown in orange. As inflation π_1 rises, the central bank raises the interest rate i_1, which encourages labor supply $\hat{n}_1^s$, lowering the real wage w_1. As a lower wage prompts firms to reduce prices, this offsets the inflation response. This loop with a strength of $\phi\theta\kappa$ takes 4 steps to close, meaning that whenever it loops once, only λ^4 share of the people perceive the next loop, resulting in a dampening of λ^4 in $\overline{\mathbb{E}}_0[\pi_1]$ in (22) relative to π_1 in (20).

The final loop concerns the *Keynesian cross*, shown in gray. As the central bank raises the interest rate i_1, it discourages household consumption c_1, leading firms to lower dividends div_1 and reduce labor demand $\hat{n}^d$, resulting in a lower wage w_1. Consequently, households want to consume even less, triggering additional adjustments by firms. This Keynesian cross strengthens any effect that impacts households, thus compounding the monetary policy loop. Once again, summing the infinite geometric series results in the strength of this loop $\sigma^{-1}(\nu + \sigma)$ appearing in the denominator of π_1 in (20), with its dampening for expectations $\overline{\mathbb{E}}_0[\pi_1]$ in (22) captured by $K_1(\lambda)$.

Overall, the average inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ in (22) is modified relative to the true inflation π_1 in (20), with different loops exponentially dampened by lengths. In contrast, other model parameters (such as the Taylor rule coefficient ϕ) and any misperceptions thereof only change the perceived strength proportionally. This distinction highlights the precise sense in which shallow thinking may be quantitatively more significant than other sources of heterogeneity in people's mental models, such as disagreement about model parameters or causal links.

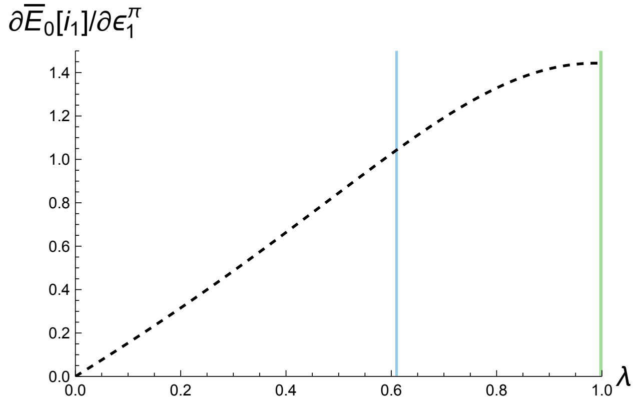
Once we establish the inflation response, the equilibrium interest rate response i_1 in (21) follows directly as the Taylor rule coefficient ϕ times inflation. The average interest expectation $\overline{\mathbb{E}}_0[i_1]$ in (23) is λ times ϕ times the inflation expectation, as it takes one more step for agents to appreciate the response of interest rate to inflation.

In the limit of $\lambda = 0$, all agents take only one step. They do not perceive any feedback on inflation, i.e., $\overline{\mathbb{E}}_0[\pi_1] = \epsilon_1^{\pi}$, and overlook changes in all other variables, e.g., $\overline{\mathbb{E}}_0[i_1] = 0$.

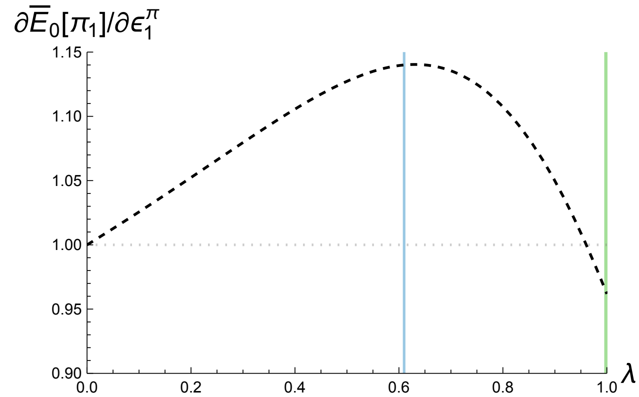
Figure 5 plots the interest rate expectation $\overline{\mathbb{E}}_0[i_1]$ and inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ as functions of λ in dashed black lines, in response to cost-push shock ϵ_1^{π}. In each graph, the blue vertical line indicates our calibrated λ, whereas the green vertical link corresponds to the rational expectations, which coincide with the true equilibrium responses i_1, π_1.

Panel 5a implies that $\overline{\mathbb{E}}_0[i_1]$ underreacts to the cost-push shock compared to the true response, because agents underappreciate the Taylor rule. In our economy, the interest rate expectation $\overline{\mathbb{E}}_0[i_1]$ is a forward rate and a component of the long-term yield $y_0^{(2)}$. Thus our theory implies that the long-term yield itself will underreact. This is in line with findings in Bauer, Pflueger and Sunderam (2024*a*) that long-term interest rates responded too little to inflation surprises before the March 2022 interest rate hike, compared to the rise in short-term interest rates that followed.[23]

[23] Bauer, Pflueger and Sunderam (2024*b*) and Bocola et al. (2024) provide estimates of the perceived monetary policy rule over longer sample periods.



(a) Interest rate expectation $\overline{\mathbb{E}}_0[i_1]$



(b) Inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$

Figure 5: Average beliefs in response to period-1 cost-push shocks ϵ_1^π

Notes: Panel (a) plots the average interest expectation $\overline{\mathbb{E}}_0[i_1]$ (relative to the size of the shock) as a function of the shallow thinking parameter λ, in response to news about a cost-push shock ϵ_1^π, and panel (b) plots the average inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$. The blue line indicates our calibration of λ, while the green line represents rational expectations ($\lambda = 1$).

Panel 5b suggests that the inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ is non-monotonic in λ. Further, our calibration suggests that the inflation expectation exceeds the size of the direct effect of one, whereas the true inflation response is below one. That is, shallow agents think that the cost-push shock will be amplified, though actually it will be dampened. The underlying reason is that, in determining inflation (20), there is a shorter loop that amplifies the cost-push shock and a much longer loop that offsets it. When agents are shallow, they understand the shorter loop relatively better than the longer loop. Thus, on net, they perceive amplification. That can be true even though the longer offset loop is actually stronger than the shorter amplification loop, leading to actual net offset.

Monetary policy news shocks. Proposition 5 characterizes the period-1 equilibrium in response to the monetary policy shock ϵ_1^i and period-0 expectations thereof.

Proposition 5. (Period-1 monetary policy shock) *The period-1 equilibrium in response to a monetary policy shock ϵ_1^i features*

$$i_1 = \left[1 + \frac{\phi\theta\kappa\sigma^{-1}(\nu+\sigma)}{1-(1-\theta)}\right]^{-1}\epsilon_1^i \tag{24}$$

$$\pi_1 = -\frac{\theta\kappa\sigma^{-1}(\nu+\sigma)}{1-(1-\theta)} i_1 \tag{25}$$

whereas the period-0 average expectations upon observing the news about ϵ_1^i are

$$\overline{\mathbb{E}}_0[i_1] = \left[1 + \frac{\lambda^4\phi\theta\kappa K_1(\lambda)}{1-\lambda(1-\theta)}\right]^{-1} \epsilon_1^i \tag{26}$$

$$\overline{\mathbb{E}}_0[\pi_1] = -\frac{\lambda^3\theta\kappa K_1(\lambda)}{1-\lambda(1-\theta)} \overline{\mathbb{E}}_0[i_1] \tag{27}$$

with $K_1(\lambda)$ increasing in λ and $K_1(1) = \sigma^{-1}(\nu+\sigma)$.

These results relate to those regarding the cost-push shocks in Proposition 4, but with a subtle and consequential difference. In this case, *all* general equilibrium effects offset the interest response to a monetary policy shock, differing from the inflation response to a cost-push shock analyzed previously, which involves both amplification and offset.

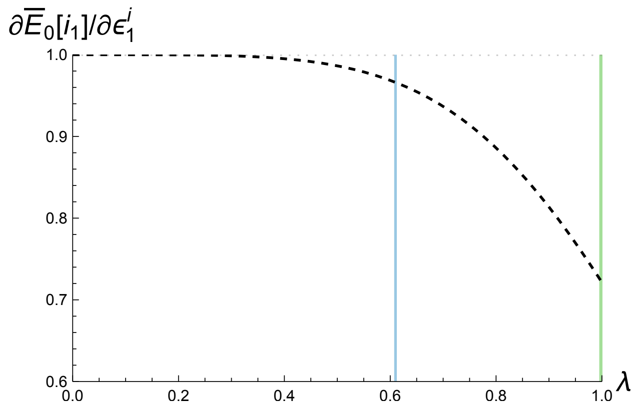
To appreciate that, we analyze the interest rate in (24), which is the sum of all n-step effects, as in (9). These effects belong to three different loops—pricing complementarity, the monetary policy loop, and the Keynesian cross—as previously established and displayed in Figure 4. Among the three loops, the interest rate response is directly involved in *only one*: the monetary policy loop. This loop offsets the interest rate response to a monetary policy shock in 4 steps: a higher interest rate i_1 encourages labor supply $\hat{n}_1^s$, which then lowers the real wage w_1, leading to lower inflation π_1 through firms' pricing decisions, ultimately prompting the central bank to lower the interest rate i_1 according to the Taylor rule. This 4-step monetary policy offset loop, with a strength $\phi\theta\kappa$, *compounds* with the other two loops—pricing complementarity and the Keynesian cross—which correspond to the $\frac{1}{1-(1-\theta)}$ and $\sigma^{-1}(\nu+\sigma)$ terms in (24). That occurs because pricing complementarity strengthens any effect on inflation, while the Keynesian cross reinforces any effect impacting households.

For the interest rate expectation $\overline{\mathbb{E}}_0[i_1]$ in (26), the 4-step monetary policy loop is dampened by λ^4, the pricing complementarity self-loop is dampened by λ, and the Keynesian cross is also dampened, captured by $K_1(\lambda)$, as in the prior case with cost-push shocks.

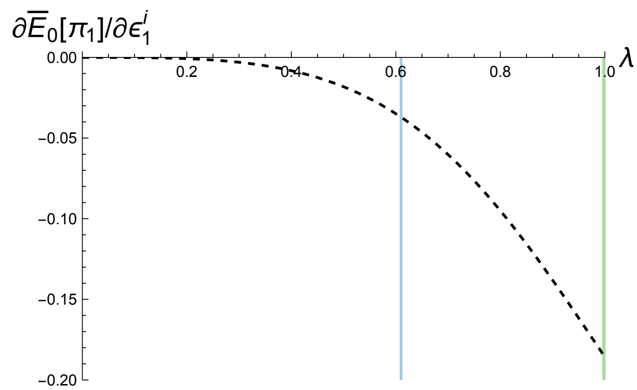
The equilibrium inflation response π_1 in (25) depends on the interest rate response i_1. It is compounded by the pricing complementarity $\frac{1}{1-(1-\theta)}$ and the Keynesian cross $\sigma^{-1}(\nu+\sigma)$, as any effect of the interest rate on inflation involves both firms and households. The inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ derives accordingly from the interest rate expectation $\overline{\mathbb{E}}_0[i_1]$, but is dampened by λ^3, since it takes 3 steps for the interest rate to affect the inflation.

In the limit of $\lambda = 0$, shallow agents perceive no general equilibrium effects, and thus $\overline{\mathbb{E}}_0 [i_1] = \epsilon_1^i, \overline{\mathbb{E}}_0 [\pi_1] = 0$.

Figure 6 plots the interest rate expectation $\overline{\mathbb{E}}_0 [i_1]$ and inflation expectation $\overline{\mathbb{E}}_0 [\pi_1]$ as functions of λ in dashed black lines, in response to monetary policy shock ϵ_1^i. As before, the blue vertical line indicates our calibrated λ, whereas the green vertical link corresponds to the rational expectations as well as the true equilibrium responses i_1, π_1.



(a) Interest rate expectation $\overline{\mathbb{E}}_0 [i_1]$



(b) Inflation expectation $\overline{\mathbb{E}}_0 [\pi_1]$

Figure 6: Average beliefs in response to period-1 monetary policy shocks ϵ_1^i

Notes: Panel (a) plots the average interest expectation $\overline{\mathbb{E}}_0 [i_1]$ (relative to the size of the shock) as a function of the shallow thinking parameter λ, in response to news about a monetary policy shock ϵ_1^i, and panel (b) plots the average inflation expectation $\overline{\mathbb{E}}_0 [\pi_1]$. The blue line indicates our calibration of λ, while the green line represents rational expectations ($\lambda = 1$).

Panel 6a indicates that $\overline{\mathbb{E}}_0 [i_1]$ overreacts to the monetary policy shock compared to the true response, thus suggesting that the period-0 long-term yield $y_0^{(2)}$ overreacts too. This aligns with a large body of literature on the excess sensitivity of long-term interest rates to monetary policy shocks, including Cochrane and Piazzesi (2002), Gürkaynak, Sack and Swanson (2005), and Hanson and Stein (2015), among others. The underlying mechanism is that, all general equilibrium effects offset the interest rate response to a monetary policy shock, and as they are dampened to varying degrees, shallow agents perceive less offset than the true overall effect. In contrast, panel 6b implies that $\overline{\mathbb{E}}_0 [\pi_1]$ underreacts, because agents underappreciate the effect of monetary policy on inflation.

Synthesis: strength and length of GE feedbacks jointly determine belief misreaction. So far, we have established a rich pattern of belief misreaction to news shocks, summarized in Table 4. For a variable of interest (such as inflation), we refer to shocks that directly affect it as *own shocks* (e.g., cost-push shocks) and shocks that directly affect other variables as *cross shocks* (e.g., monetary policy shocks). Under our calibration, the average expectations of a variable overreact to its own shocks and underreact to cross shocks.

Table 4: Belief over- and underreaction in the New Keynesian economy

News shock	Inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$	Interest rate expectation $\overline{\mathbb{E}}_0[i_1]$
Cost-push ϵ_1^{π}	Overreaction	Underreaction
Monetary policy ϵ_1^{i}	Underreaction	Overreaction

Notes: This table summarizes belief misreaction to different news shocks. The average expectations of variable v overreact (compared to the true equilibrium) to shock s if $\left|\frac{\partial \overline{\mathbb{E}}_0[v_1]}{\partial \epsilon_1^s}\right| > \left|\frac{\partial v_1}{\partial \epsilon_1^s}\right|$, and underreact otherwise.

We offer a synthesis of belief misreaction under shallow thinking. To start with, consider a simple model with a single general equilibrium (GE) loop. Beliefs about a variable underreact (or overreact) to its own shocks if the single GE loop amplifies (or offsets) the direct effect of these shocks, as shallow agents underappreciate the GE loop, in line with existing theories reviewed by Angeletos and Lian (2023*a*). Additionally, our theory suggests that beliefs about a variable underreact to cross shocks (e.g., inflation in response to cost-push shocks) because shallow agents underperceive shock propagation. This latter prediction extends to more complex models with multiple GE loops.

In models with multiple GE loops, such as our workhorse macroeconomic models, whether beliefs over- or underreact to own shocks depends jointly on the strength and length of these GE loops. For example, consider the inflation response to a cost-push shock ϵ_1^{π}. Under our calibration, shallow agents perceive the cost-push shock ϵ_1^{π} as amplified, even though it is actually offset, i.e., $\frac{\partial \overline{\mathbb{E}}_0[\pi_1]}{\partial \epsilon_1^{\pi}} > 1 > \frac{\partial \pi_1}{\partial \epsilon_1^{\pi}}$. Shallow thinking flips the sign of the perceived net GE effect, because shallow agents underappreciate the long, strong offset loop involving the monetary policy reaction. The order of operations is key, akin to Jensen's inequality, as GE feedbacks of varying length are dampened differently. The implication could differ if one ignored length by first collapsing multiple GE loops in the model into a single net effect (which, in this case, is offset) and then applying dampening.

As a thought experiment, we could consider varying the strength of the offsetting loop by adjusting the Taylor rule coefficient ϕ, as depicted in Figure 7. If, instead of $\phi = 1.5$,

ϕ takes on an intermediate value, e.g., $\phi = 1$, the long offset loop is too weak to turn the true net effect into offset. In that case, the true net effect is amplification, but shallow agents perceive even *more* amplification, i.e., $\frac{\partial \overline{\mathbb{E}}_0[\pi_1]}{\partial \epsilon_1^\pi} > \frac{\partial \pi_1}{\partial \epsilon_1^\pi} > 1$, as they overweigh the short amplification loop relative to the long offset loop. Again, if one naively collapsed multiple GE loops into a net amplification effect and dampened it, the perceived amplification would be less rather than more.

With an even lower ϕ, e.g., $\phi = 0$ in a case where the policy rate is constrained at the zero lower bound (ZLB) in response to a deflationary shock,[24] the long offset loop is weak or non-existent, leaving the short amplification loop as the dominant force. That leads us towards a simple case with only one GE loop that amplifies the direct effect, which we began this discussion with. As a result, shallow agents perceive less amplification on net than the true effect, i.e., $\frac{\partial \pi_1}{\partial \epsilon_1^\pi} > \frac{\partial \overline{\mathbb{E}}_0[\pi_1]}{\partial \epsilon_1^\pi} > 1$, nesting the simple one-loop case.

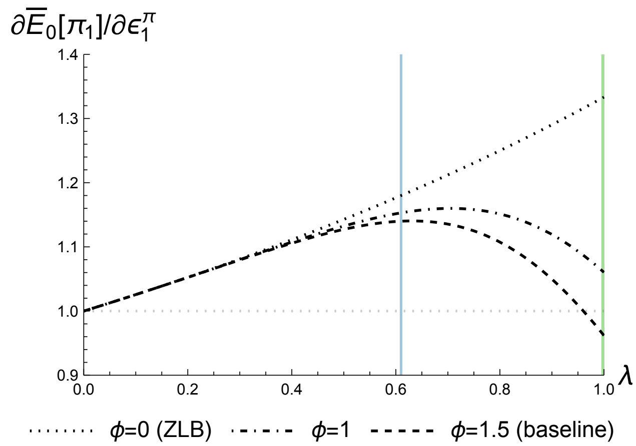


Figure 7: Inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ in response to cost-push shock ϵ_1^π (alternative ϕ)

Notes: This figure plots the average interest expectation $\overline{\mathbb{E}}_0[i_1]$ (relative to the size of the shock) as a function of the shallow thinking parameter λ, in response to news about a cost-push shock ϵ_1^π, under different values of the Taylor rule coefficient ϕ. The blue line indicates our calibration of λ, while the green line represents rational expectations ($\lambda = 1$).

[24] The Taylor principle requires $\phi > 1$ for determinacy of the textbook infinite-horizon New Keynesian model under rational expectations. In our 2-period setting, determinacy is not a concern. With persistent shocks, shallow thinking beliefs are uniquely defined by a formula similar to (10), which helps select an equilibrium. For further discussion on determinacy, which is not the focus of our paper, see Farhi and Werning (2019), Gabaix (2020) and Angeletos and Lian (2023*b*).

4.2 Predictability of Bond Excess Returns

In this subsection, we consider an economy hit by two news shocks, the cost-push shock ϵ_1^π and the monetary policy shock ϵ_1^i, and study the bond excess returns. We show shallow thinking implies that bond excess returns can be predicted by macroeconomic variables, controlling for current yields, as established empirically by Ludvigson and Ng (2009), Cooper and Priestley (2009), Joslin, Priebsch and Singleton (2014), Cieslak and Povala (2015), and Cieslak (2018), among others.

We consider the excess return of holding a 2-period (i.e., long-term) bond from period 0 to period 1, relative to holding a 1-period (i.e., short-term) bond,[25]

$$xr_{0\to 1}^{(2)} \equiv \underbrace{-i_1 + 2y_0^{(2)}}_{\text{return of long-term bond}} - \underbrace{i_0}_{\text{return of short-term bond}} = \overline{\mathbb{E}}_0\left[i_1\right] - i_1 \tag{28}$$

where the equality follows from (19). When the period-0 expectation of period-1 interest rate exceeds its actual value, the long-term bond is undervalued in period 0 and will appreciate in period 1, leading to a positive excess return, and vice versa.

In particular, we study the predictability of bond excess return $xr_{0\to 1}^{(2)}$ by the average inflation expectation $\overline{\mathbb{E}}_0\left[\pi_1\right]$, as noted by Joslin, Priebsch and Singleton (2014) and Cieslak (2018), controlling for the forward rate $\overline{\mathbb{E}}_0\left[i_1\right]$,[26]

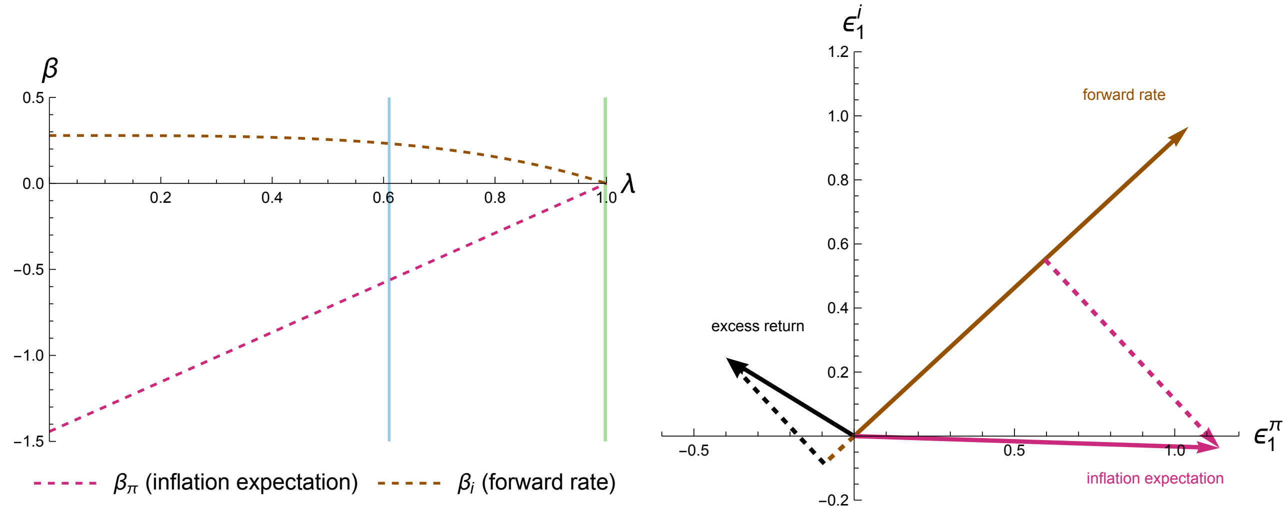
$$\underbrace{xr_{0\to 1}^{(2)}}_{\text{bond excess return}} = \beta_\pi \underbrace{\overline{\mathbb{E}}_0\left[\pi_1\right]}_{\text{inflation expectation}} + \beta_i \underbrace{\overline{\mathbb{E}}_0\left[i_1\right]}_{\text{forward rate}} + \alpha + \epsilon_{0\to 1} \tag{29}$$

Figure 8a illustrates the theory-implied coefficients β_π and β_i as functions of λ, in brown and purple respectively. As long as $\lambda < 1$, our theory predicts a negative β_π, which is what Joslin, Priebsch and Singleton (2014) and Cieslak (2018) find when using inflation expectations or other macroeconomic variables to predict bond excess returns.

To understand the mechanism, we start by examining the limits of $\lambda = 0$ and 1 and build intuition with shock loadings illustrated in Figure 8b. In the limit of $\lambda = 0$, we have $\overline{\mathbb{E}}_0\left[\pi_1\right] = \epsilon_1^\pi, \overline{\mathbb{E}}_0\left[i_1\right] = \epsilon_1^i$ from Propositions 4 and 5. As the interest rate expectation

[25] The return of the 2-period bond is the difference between its period-1 log price deviation, $-i_1$, and its period-0 log price derivation, $-2y_0^{(2)}$.

[26] This regression, with expectations as predictors, is equivalent to a bivariate Coibion and Gorodnichenko (2015) regression. The coefficients reflect the misappreciation of causal relations between variables, which Wu (2023) systematically analyzes both theoretically and empirically across many variable pairs.



(a) Coefficients β_π and β_i

(b) Loadings on news shocks (calibrated λ)

Figure 8: Predictability of bond excess returns, with transitory news shocks

Notes: Panel (a) plots the theory-implied coefficients β_π and β_i from the predictive regression of bond excess returns $xr^{(2)}_{0\to1}$ in (29) on the inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ and forward rate $\overline{\mathbb{E}}_0[i_1]$, as functions of the shallow thinking parameter λ. The period-0 economy is impacted by news about period-1 cost-push shocks ϵ_1^π and monetary policy shocks ϵ_1^i. The blue vertical line corresponds to our calibration of λ, while the green line represents rational expectations ($\lambda = 1$).
Panel (b) illustrates the loadings of bond excess return $xr^{(2)}_{0\to1}$, inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ and forward rate $\overline{\mathbb{E}}_0[i_1]$ on the cost-push news shock ϵ_1^π (x-axis) and the monetary policy news shock ϵ_1^i (y-axis), as solid vectors. The vectors are determined under our calibration of λ, but their quadrant placements are generically true when $\lambda \in (0,1)$. Figures C5a and C5b show the cases with $\lambda = 0, 1$. The two dashed vectors show the residuals of bond excess return $xr^{(2)}_{0\to1}$ and inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ projected onto the forward rate $\overline{\mathbb{E}}_0[i_1]$. Their opposite directions imply a negative β_π, following the Frisch-Waugh-Lovell theorem.

underreacts to the cost-push shock ϵ_1^π and overreacts to the monetary policy shock ϵ_1^i, the excess return $xr^{(2)}_{0\to1}$ loads negatively on the former and positively on the latter, as determined by (28). That is, it lies in the second quadrant (plotted in Figure C5a). Because the forward rate and the inflation expectation are unit vectors along the x- and y-axes in this case, the coefficient β_π is negative and β_i is positive.

In the limit of rational expectations ($\lambda = 1$), $\overline{\mathbb{E}}_0[\pi_1]$ loads positively on the cost-push shock ϵ_1^π and negatively on the monetary policy shock ϵ_1^i, placing it in the fourth quadrant (plotted in Figure C5b). $\overline{\mathbb{E}}_0[i_1]$ loads positively on both shocks, placing it in the first quadrant. However, in this case, and only in this case, the excess return $xr^{(2)}_{0\to1}$ is 0, as there is no expectational error. Thus, both coefficients β_π and β_i are 0.

Generically, the inflation expectation $\overline{\mathbb{E}}_0[\pi_1]$ lies in the fourth quadrant, and the forward rate $\overline{\mathbb{E}}_0 [i_1]$ is in the first quadrant, as long as $\lambda > 0$ (i.e., some understanding of shock propagation). The excess return $xr^{(2)}_{0\to1}$ is placed in the second quadrant if $\lambda < 1$ (i.e., not fully understanding shock propagation). Figure 8b illustrates these generic placements. It also shows the residual of $xr^{(2)}_{0\to1}$ projected onto $\overline{\mathbb{E}}_0 [i_1]$, denoted $xr^{(2)}_{0\to1}|\overline{\mathbb{E}}_0 [i_1]$, and the residual $\overline{\mathbb{E}}_0 [\pi_1] |\overline{\mathbb{E}}_0 [i_1]$, both shown as dashed vectors. These two residuals point to opposite directions due to the quadrant placements of the three vectors involved. According to the Frisch-Waugh-Lovell theorem, the coefficient β_π from the bivariate regression (29) equals the univariate regression coefficient of the residual $xr^{(2)}_{0\to1}|\overline{\mathbb{E}}_0 [i_1]$ on the residual $\overline{\mathbb{E}}_0 [\pi_1] |\overline{\mathbb{E}}_0 [i_1]$. Opposite directions of these residuals imply a negative coefficient β_π.

In this subsection, we explain an important finding from Joslin, Priebsch and Singleton (2014) and Cieslak (2018) that inflation expectations negatively predict bond excess returns. We illustrate this in an economy with two shocks: one to the interest rate and another macroeconomic shock. In reality, the economy is impacted by multiple shocks. Mathematically, in a linear model with N shocks, N independent predictors would span all shocks. However, in practice, the entire yield curve is well captured by the first three principal components. Our theory suggests that other macroeconomic variables may contain additional information about non-monetary-policy macroeconomic shocks, and can therefore predict bond excess returns.

Shallow thinking reconciles several bond market puzzles. Taking stock of findings here and in the previous subsection, shallow thinking offers a unified explanation for several bond market puzzles that seem unrelated or even contradictory. These include the underreaction of long-term interest rates to changes in economic conditions or inflation surprises (Cieslak, 2018; Bauer, Pflueger and Sunderam, 2024*a*), their overreaction to monetary policy shocks (Cochrane and Piazzesi, 2002; Gürkaynak, Sack and Swanson, 2005; Hanson and Stein, 2015), and the predictability of bond excess returns by macroeconomic variables after controlling for current yields (Ludvigson and Ng, 2009; Cooper and Priestley, 2009; Joslin, Priebsch and Singleton, 2014; Cieslak and Povala, 2015).

The first two puzzles concern responses conditional on shocks, while the last involves a multivariate regression that is unconditional on shocks. When unconditional on shocks, whether a univariate regression of the long-term interest rate on the short-term rate suggests over- or underreaction depends on the relative importance of shocks, according to our theory. Indeed, Hanson, Lucca and Wright (2021) show that long-term interest rates are overly sensitive to changes in the short-term rates and predictably revert after 2000, suggesting overreaction. However, this pattern does not hold prior to 2000. Through the lens of our theory, one possibility is that there were more supply shocks (such as oil shocks) before 2000, contributing to underreaction in the mix and obscuring overreaction to monetary policy shocks.[27]

4.3 Macroeconomic Effects of Cost-Push News Shocks

We examine the macroeconomic effects of shallow thinking in period 0, in response to transitory news shocks. In particular, we establish that news about period-1 cost-push shocks are more inflationary and less contractionary than the rational expectations prediction.

Period-0 equilibrium. In period 0, the shocks have not materialized, but firms and households are forward-looking. The period-0 equilibrium arises from firms' and households' optimal behavior, given their perfect observation of all period-0 variables and their beliefs about period-1 outcomes, along with the central bank's Taylor rule and the labor market clearing condition.

The period-0 equilibrium consists of seven variables $\left\{i_0, \pi_0, div_0, n_0^d, c_0, n_0^s, w_0\right\}$, similar to the period-1 equilibrium. Among these seven variables, three of them (the interest rate i_0, dividend div_0 and labor demand n_0^d) depend only on the contemporaneous values of the other variables. These contemporaneous causal relations are the same as their period-1 counterparts (1, 2, 6). Three variables (the inflation π_0, consumption c_0 and labor supply n_0^s) depend on agents' beliefs about period-1 outcomes, since firms's pricing decisions and households' consumption and labor supply decisions are forward-looking, detailed next. Last, the wage w_0 arises from the labor market clearing condition $n_0^s = n_0^d$.

The period-0 inflation π_0 satisfies

$$\pi_0 = \theta\kappa\left(w_0 + \beta\theta\overline{\mathbb{E}}_0\left[w_1\right]\right) + (1-\theta)\left(\pi_0 + \beta\theta\overline{\mathbb{E}}_0\left[\pi_1\right]\right) \tag{30}$$

which increases in expectations of both future real wage w_1 and inflation π_1, as firms want to front run a higher future marginal cost.

[27] We show the comparative statics of this univariate regression coefficient with respect to the relative importance of monetary policy shocks in Appendix C.4, in the case with persistence shocks, to which we generalize shallow thinking in Section 5.

The period-0 consumption c_0 and labor supply n_0^s follow

$$c_0 = -\sigma^{-1}\beta\left(i_0 - \overline{\mathbb{E}}_0\left[\pi_1\right] + \beta\overline{\mathbb{E}}_0\left[i_1\right]\right) + \frac{(1-\beta)(\mu-1)\nu}{\sigma+\mu\nu}\left(div_0 + \overline{\mathbb{E}}_0\left[div_1\right]\right) + \frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu}\left(w_0 + \overline{\mathbb{E}}_0\left[w_1\right]\right) \tag{31}$$

$$n_0^s = \nu^{-1}\beta\left(i_0 - \overline{\mathbb{E}}_0\left[\pi_1\right] + \beta\overline{\mathbb{E}}_0\left[i_1\right]\right) - \frac{(1-\beta)(\mu-1)\sigma}{\sigma+\mu\nu}\left(div_0 + \overline{\mathbb{E}}_0\left[div_1\right]\right) + \nu^{-1}\left[1 - \sigma\frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu}\right]w_0 - \frac{(1-\beta)(1+\nu)\nu^{-1}\sigma}{\sigma+\mu\nu}\overline{\mathbb{E}}_0\left[w_1\right] \tag{32}$$

Households react to the future interest rate, dividend, and wage, as well as the future inflation, as a higher inflation lowers the real interest rate from period 0 to period 1.

In determining the period-0 equilibrium, the only decision-relevant beliefs are about inflation π_1, wage w_1, interest rate i_1 and dividend div_1. Figure 9a illustrates the determination of period-0 equilibrium, where these beliefs act like shocks to firms and households.[28]

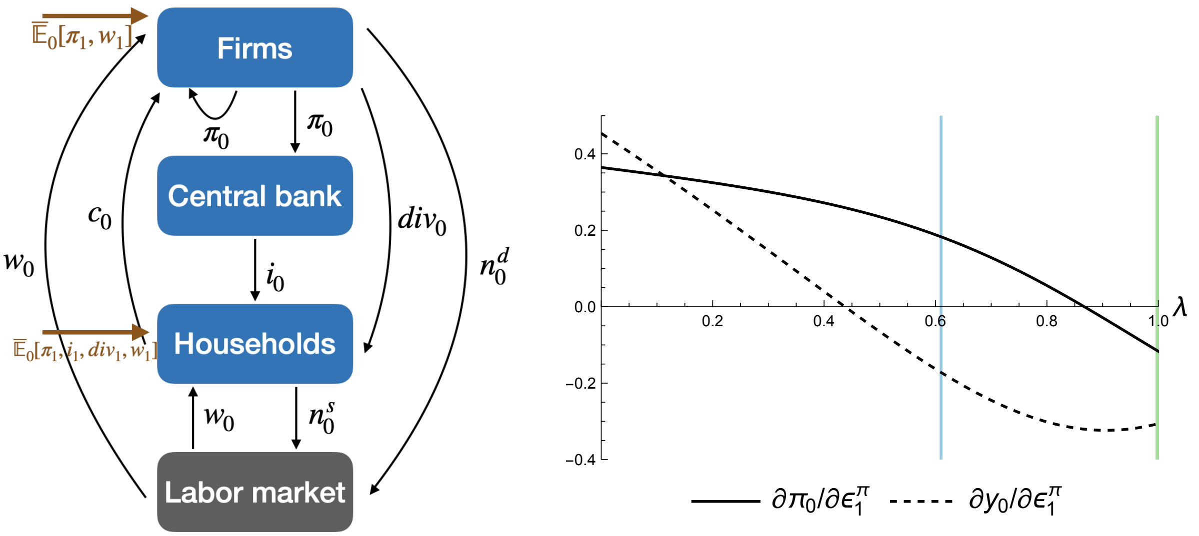
In particular, the inflation expectation $\overline{\mathbb{E}}_0\left[\pi_1\right]$ serves two roles here: it acts like a cost-push shock for firms as they want to front run future inflation, and it functions as a demand shock for households since a higher inflation lowers the real interest rate. This observation will prove useful when we discuss the effects of cost-push shocks next.

Cost-push news shocks. Figure 9b depicts the period-0 inflation π_0 and output y_0 (which is simply the consumption c_0), in response to news about a period-1 cost-push shock ϵ_1^π. Unlike the period-1 equilibrium, which is independent of λ, the period-0 equilibrium does depend on λ. The blue line indicates the equilibrium under our calibration of λ, whereas the green line stands for the rational expectations equilibrium (REE).

A cost-push news shock is *inflationary* in period 0 under the calibrated shallow thinking parameter λ, but *deflationary* under rational expectations, seen from the solid line. The latter finding is because rational agents understand that the central bank will raise interest rate i_1 in response to higher inflation π_1, causing a contraction of the period-1 economy. Anticipating that, households cut back on their consumption today, resulting in a contraction in period 0 as well, represented by the dashed line.

In contrast, in the limit of extremely shallow agents ($\lambda = 0$), they only perceive a change

[28] In Figure 9a that determines the equilibrium given beliefs, we equalize the actual labor demand and supply n_0^d, n_0^s as the market clearing condition. We only introduce fictitious Walrasian auctioneers for the purpose of characterizing beliefs. In the equilibrium given any beliefs, markets always clear.



(a) Period-0 equilibrium

(b) Period-0 inflation π_0 and output y_0, in response to cost-push news shocks ϵ_1^π

Figure 9: Period-0 equilibrium in response to news shocks

Notes: Panel (a) demonstrates the period-0 equilibrium in response to transitory news shocks, which are observed in period 0 but only impact the economy in period 1, driven by firms' and households' beliefs about period-1 outcomes.
Panel (b) plots the period-0 inflation π_0 and output y_0 (which equals consumption c_0) in response to news about a cost-push shock ϵ_1^π, normalized relative to the size of the shock, as functions of the shallow thinking parameter λ. The solid line stands for inflation π_0 and the dashed line represents output y_0. The blue line indicates the equilibrium under our calibration of λ, while the green line represents the rational expectations equilibrium with $\lambda = 1$.

in future inflation, $\overline{\mathbb{E}}_0[\pi_1] = \epsilon_1^\pi$, but not in any other variable. As we have established with (30-32), inflation expectations act like a cost-push shock for firms and a demand shock for households. As a result, firms set higher prices and households consume more and work less, leading to period-0 inflation and output expansion.

With the calibrated λ, agents somewhat, but not fully, expect an economic downturn in period 1. In that case, period-1 cost-push shocks lead to less inflation and a smaller output contraction in period 0 compared to the rational expectations equilibrium.

From transitory news shocks to persistent shocks. So far, we have focused on transitory news shocks, but this is much more informative than it may seem. As we have noted at the beginning of this section, comparing theories of beliefs under transitory news shocks is equivalent to comparing them under two-period shocks, since the period-0 equilibrium response to contemporaneous shocks is independent of belief formation.

Furthermore, analyzing transitory news shocks can help build intuition for persistent shocks. Under a first-order approximation, a persistent cost-push shock observed in period 0 is a sum of a current shock and a series of future shocks. Shallow thinking implies that a future cost-push shock is inflationary in period 0, whereas rational expectations predict otherwise. This suggests that a persistent cost-push shock is more inflationary under shallow thinking. Additionally, a *more* persistent cost-push shock is a sum of more future shocks. Thus, shallow thinking predicts that a *more* persistent cost-push shock leads to *higher* inflation, contrary to the rational expectations prediction.

Next, we will generalize shallow thinking to persistent shocks to substantiate these two claims. The generalization will address one missing piece from the intuitive analysis above: in responding to a persistent shock, agents must think about how all decision-relevant variables in any future period t_1 depend on shocks in further periods t_2.

5 Consequences of Persistent Shocks

We generalize to persistent shocks in the New Keynesian economy and show that the insights gained from transitory news shocks still hold, while new lessons emerge. In particular, more persistent cost-push shocks lead to higher inflation under shallow thinking, in contrast to the lower inflation predicted under rational expectations.

We introduce shallow thinking of a dynamic general equilibrium in Section 5.1, by focusing on the cross-variable causal relations and abstracting away from the cross-horizon dimension. This will allow us to formally address belief misreaction to persistent shocks and their effects in Section 5.2. Readers who are more interested in the consequences and less interested in the technical details could skip the first subsection.

We consider persistent shocks $\{\epsilon_t\}_{t\geq 0}$ that are observed at time 0^-, and assume that agents form expectations $\mathbb{E}[\cdot]$ once and for all at time 0^-, i.e., $\mathbb{E}_t[v_s] \equiv \mathbb{E}[v_s]$ for $s > t$. This assumption simplifies the analysis, as it nests the rational expectations benchmark and is reasonable for our theory for the following two reasons.

First, an important reason for expectation updates over time is that agents gradually learn about shocks. We focus on the rationality of beliefs in the absence of any information frictions, and our survey design mimics this environment.

Second, expectations can update as agents learn how aggregate variables respond to shocks *through repeated experiences*, similar to how economists study shock propagation using time series data. We do not explicitly model such learning processes, but we note that in measuring shallow thinking through a survey, we consider conventional shocks, such as oil shocks and monetary policy shocks. These are age-old shocks, so our estimation already incorporates knowledge that people have gained over time. This contrasts with unconventional policy shocks (such as forward guidance), to which people had no prior exposure until recently. That said, how agents form their mental models through learning and how this interacts with information frictions are promising topics for future research.

5.1 Dynamic General Equilibrium and Shallow Thinking

We conceptualize a dynamic general equilibrium (GE) as a system of causal relations, which can be similarly represented in a directed graph, generalizing Sections 2.2 and 2.3. This conceptualization of a dynamic GE will be useful for both the New Keynesian economy and the RBC economy, which we analyze later in Section 6.

The infinite-horizon New Keynesian model is characterized by sequences of seven variables $\left\{i_t, \pi_t, div_t, n_t^d, c_t, n_t^s, w_t\right\}_{t\geq 0}$. We will refer to this collection of variables as $\mathcal{V}$. The first six variables are agents' actions, which we collect as $\mathcal{V}^{action}$, whereas the last is a price formed in the competitive labor market. Note that we start with the labor supply and demand n_t^s, n_t^d in order to reinterpret them as supply and demand curves shortly. In that reinterpretation, with only slight abuse of notation, we will use $\mathcal{V}$ and $\mathcal{V}^{action}$ to denote variables of interests with n^s, n^d replaced by $\hat{n}^s, \hat{n}^d$.

With persistent shocks, we distinguish between two sets of equations that jointly determine the equilibrium outcomes, similar to García-Schmidt and Woodford (2019). The first set corresponds to relations among economic variables that arise from optimal decisions of economic agents, given the current realizations and their expected future values of variables that directly affect them. We call these temporary equilibrium relations. Based on these temporary equilibrium relations, we will define the second set of equations that characterizes expectations, which generalizes Assumption 1.

In terms of the temporary equilibrium relations, three variables i_t, div_t, n_t^d only depend on contemporaneous values of other variables, laid out in (C1, C2, C6). Three variables π_t, c_t, n_t^s are forward-looking. Inflation π_t is a linear function in w_t, π_t and $\left\{\overline{\mathbb{E}}_t\left[w_{t+k}\right], \overline{\mathbb{E}}_t\left[\pi_{t+k}\right]\right\}_{k\geq 1}$, and consumption and labor supply c_t, n_t^s are linear functions in i_t, div_t, w_t and $\left\{\overline{\mathbb{E}}_t\left[i_{t+k}\right], \overline{\mathbb{E}}_t\left[\pi_{t+k}\right], \overline{\mathbb{E}}_t\left[div_{t+k}\right], \overline{\mathbb{E}}_t\left[w_{t+k}\right]\right\}_{k\geq 1}$, detailed in (C3-C5). The last variable w_t arises from equilibrating labor supply n_t^s and demand n_t^d. These seven equations completely characterize the equilibrium, given beliefs.

REE as causal relations in sequence space. To determine shallow thinking beliefs, we first characterize the rational expectations equilibrium (REE) and causal relations thereof. For REE, by replacing each expectation $\overline{\mathbb{E}}_t[v_\tau]$ with the true outcome v_τ under rational expectations, the six variables in $\mathcal{V}^{action}$ that agents choose can be represented in the sequence space, following Auclert, Rognlie and Straub (2024) and Auclert et al. (2021), as

$$\mathbf{v}^{REE} = \sum_{\mathbf{u}\in\mathcal{V}} \mathbf{J}_{\mathbf{vu}}\mathbf{u}^{REE} + \boldsymbol{\epsilon}^{v},\ \forall \mathbf{v} \in \mathcal{V}^{action} \tag{33}$$

where

$$(\mathbf{J}_{\mathbf{vu}})_{ts} \equiv \begin{cases} \frac{\partial v_t}{\partial u_s} & s \leq t \\ \frac{\partial v_t}{\partial \overline{\mathbb{E}}_t[u_s]} & s > t \end{cases} \tag{34}$$

is the Jacobian of the sequence of one variable $\mathbf{v} \equiv (\{v_t\}_{t\geq 0})'$ with respect to the sequence of another variable $\mathbf{u} \equiv (\{u_t\}_{t\geq 0})'$, and $\boldsymbol{\epsilon}^{v} \equiv \left(\left\{\epsilon_t^{v}\right\}_{t\geq 0}\right)'$ denotes the sequence of a structural shock. The Jacobians $\mathbf{J}_{\mathbf{vu}}$ are upper triangular matrices in the New Keynesian model, which is purely forward-looking, but the formulation (33) is general to accommodate models with state variables that depend on the past, such as the RBC model we study later.

For labor supply and demand $\mathbf{n}^s, \mathbf{n}^d$, by separating their dependence on the wage $\mathbf{w}$ from the rest, we interpret them as supply and demand curves in the sequence space

$$\mathbf{v}^{REE} = \mathbf{J}_{\mathbf{vw}}\mathbf{w}^{REE} + \hat{\mathbf{v}}^{REE},\ \mathbf{v} \in \left\{\mathbf{n}^s, \mathbf{n}^d\right\}$$

with elasticities $\mathbf{J}_{\mathbf{n}^s\mathbf{w}}, \mathbf{J}_{\mathbf{n}^d\mathbf{w}}$ and shifts $\hat{\mathbf{n}}^{s,REE}, \hat{\mathbf{n}}^{d,REE}$ defined as

$$\hat{\mathbf{v}}^{REE} = \sum_{\mathbf{u}\in\mathcal{V}\setminus\{\mathbf{w}\}} \mathbf{J}_{\mathbf{vu}}\mathbf{u}^{REE} + \boldsymbol{\epsilon}^{v},\ \mathbf{v} \in \left\{\mathbf{n}^s, \mathbf{n}^d\right\} \tag{35}$$

Demand elasticity $\mathbf{J}_{\mathbf{n}^d\mathbf{w}}$ is a matrix of zeros in this model since firms only use labor inputs, but, more generally (such as in the RBC model), it does not have to be. Supply elasticity $\mathbf{J}_{\mathbf{n}^s\mathbf{w}}$ is an upper triangular matrix, as households' labor supply responds to future wages.

By equalizing supply and demand $\mathbf{n}^{s,REE} = \mathbf{n}^{d,REE}$, we can interpret the wage $\mathbf{w}^{REE}$ as resulting from supply and demand shifts $\hat{\mathbf{n}}^{s,REE}, \hat{\mathbf{n}}^{d,REE}$

$$\mathbf{w}^{REE} = \left(\mathbf{J}_{\mathbf{n}^s\mathbf{w}} - \mathbf{J}_{\mathbf{n}^d\mathbf{w}}\right)^{-1}\left(\hat{\mathbf{n}}^{d,REE} - \hat{\mathbf{n}}^{s,REE}\right) \tag{36}$$

This rule of price determination generalizes its counterpart in the period-1 economy (7) and incorporates the response of time-t wage on time-s demand and supply shifts. To specify beliefs, we reason with $\hat{\mathbf{n}}^{s,REE}, \hat{\mathbf{n}}^{d,REE}$ as in (35) instead of $\mathbf{n}^{s,REE}, \mathbf{n}^{d,REE}$ in (33).

Taking stock of the rational expectations equilibrium, (33) describes all agents' actions $\mathcal{V}^{action}$ and (36) characterizes the wage from the competitive labor market. We stack the sequences of outcomes to form a long vector $\mathbf{V} \equiv (\{\mathbf{v}'\}_{v\in\mathcal{V}})'$, sequences of shocks as $\mathbf{S} \equiv \left(\left\{\boldsymbol{\epsilon}^{v'}\right\}_{v\in\mathcal{V}}\right)'$, and correspondingly stack the Jacobians as a giant matrix $\mathbf{M} \equiv \left(\{\mathbf{J}_{\mathbf{vu}}\}_{v,u\in\mathcal{V}}\right)$. $\mathbf{M}$ captures the causal relations in the dynamic general equilibrium.

Proposition 6. (Dynamic GE as a system of causal relations in the sequence space) *The rational expectations equilibrium in the New Keynesian economy is characterized by the fixed point to the system of causal relations in the sequence space among all* agents' actions and competitive prices, $\mathbf{V} \equiv \left(\mathbf{i}', \boldsymbol{\pi}', \mathbf{div}', \hat{\mathbf{n}}^{d'}, \mathbf{c}', \hat{\mathbf{n}}^{s'}, \mathbf{w}'\right)'$*, as*

$$\underbrace{\mathbf{V}^{REE}}_{\text{sequence of variables}} = \underbrace{\mathbf{M}}_{\text{sequence-space causal relations}} \mathbf{V}^{REE} + \underbrace{\mathbf{S}}_{\text{sequence of shocks}} \tag{37}$$

The rational expectations equilibrium can be solved to yield

$$\mathbf{V}^{REE} = (\mathbf{I} - \mathbf{M})^{-1}\mathbf{S} = \sum_{n=1}^{\infty} \mathbf{M}^{n-1}\mathbf{S} \tag{38}$$

which is a sum of all effects of varying distance, where each $\mathbf{M}^{n-1}\mathbf{S}$ *term is an n-step effect of the sequence of a shock on the sequence of a variable via sequences of* $n-1$ *intermediate variables.*

This proposition generalizes Proposition 1 in the case of period-1 general equilibrium to dynamic general equilibrium with persistent shocks.

Shallow thinking. With Proposition 6, we generalize Assumption 1 as follows to capture shallow thinking about dynamic general equilibrium.

Assumption 1′. Individuals vary in their finite *depth of thinking* $d \in \mathbb{N}^+$, with expectations

$$\mathbb{E}^d[\mathbf{V}] \equiv \sum_{n=1}^{d} \mathbf{M}^{n-1}\mathbf{S} \tag{39}$$

which implies an iterative formula for $d > 1$ as

$$\mathbb{E}^d[\mathbf{V}] = \mathbf{M}\mathbb{E}^{d-1}[\mathbf{V}] + \mathbf{S} \tag{40}$$

Parallel to Assumption 1, (39) and (40) embed the idea that a depth-d agent only understands effects that take no more than d steps and they think one step further than a depth-$(d-1)$ agent. Each step is from the sequence of one variable to the sequence of another variable.

Remarks on cross-variable causal relations vs. cross-horizon causal relations. As briefly mentioned in the introduction, we focus on the causal relations across variables and abstract away from the cross-horizon dimension. We assume that if agents understand how consumption c_t depends on the contemporaneous interest rate i_t, they also understand how c_t depends on future i_s. This reflects in (40), where the dependence of $\mathbb{E}^d[c_t]$ on $\mathbb{E}^{d-1}[\mathbf{i}]$ is mediated via the causal relations $\mathbf{M}$, which collects all the Jacobians $\mathbf{J}_{\mathbf{vu}}$ in (34) that characterize the rational expectations equilibrium. This assumption simplifies the analysis and generates dampening of cross-variable relations in beliefs, complementing horizon-dependent dampening in Angeletos and Lian (2018), Farhi and Werning (2019) and Gabaix (2020). One could further generalize our theory to introduce additional dampening across periods, by modifying $\mathbf{M}$ in (39) and (40).[29]

Assumptions 1′ and 2 jointly lead to the following characterizations of the average beliefs, generalizing Proposition 2.

Proposition 7. (Average beliefs of dynamic GE) *The average beliefs $\overline{\mathbb{E}}[\mathbf{V}]$ are sums of all effects of shocks of varying distance*

$$\overline{\mathbb{E}}[\mathbf{V}] = \sum_{n=1}^{\infty} \lambda^{n-1}\mathbf{M}^{n-1}\mathbf{S} \tag{41}$$

Further, the average beliefs satisfy a fixed point

$$\overline{\mathbb{E}}[\mathbf{V}] = \lambda\mathbf{M}\overline{\mathbb{E}}[\mathbf{V}] + \mathbf{S} \tag{42}$$

[29]For example, by replacing $(\mathbf{J}_{\mathbf{vu}})_{ts}$ with $m^{s-t}(\mathbf{J}_{\mathbf{vu}})_{ts}$ for $s > t$ with a factor $m < 1$ and adjusting $\mathbf{M}$ accordingly in (39) and (40), we can allow for the possibility that agents underappreciate the dependence of one variable on future variables relative to its dependence on contemporaneous variables, à la Gabaix (2020).

Equilibrium under shallow thinking. Given the average beliefs $\overline{\mathbb{E}}[\cdot]$, the equilibrium $\left\{i_t, \pi_t, div_t, n_t^d, c_t, n_t^s, w_t\right\}_{t\geq 0}$ is characterized period by period, by firms' and households' optimal behavior given their perfect observation of time-t variables and expectations about future variables, as in (C1-C5), along with the Taylor rule (C6) and the labor market clearing condition $n_t^s = n_t^d$ (C7).

5.2 Macroeconomic Effects of Persistent Cost-Push Shocks

We consider responses of beliefs and equilibrium to a persistent cost-push shock $\epsilon_t^\pi = \rho^t \epsilon^\pi$. In particular, we highlight that a more persistent (higher ρ) cost-push shock leads to higher inflation under shallow thinking, contrary to the prediction of rational expectations.

REE and shallow thinking beliefs. In response to such an exponentially decaying shock, both beliefs and equilibrium outcomes decay exponentially at the same rate. Proposition 8 establishes the rational expectations equilibrium (REE) and shallow thinking beliefs, characterizing their time-t values relative to the time-t size of the shock ϵ_t^π.

Proposition 8. (Persistent cost-push shock) *The rational-expectations equilibrium (REE) response to a persistent cost-push shock $\epsilon_t^\pi = \rho^t \epsilon^\pi$ features*

$$\pi_t^{REE} = \left[1 - \frac{(1-\theta) + \left(\rho - \phi\right)\theta\kappa\frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}}{1-\beta\theta\rho}\right]^{-1} \epsilon_t^\pi \tag{43}$$

$$i_t^{REE} = \phi \pi_t^{REE} \tag{44}$$

whereas the average expectations under shallow thinking are

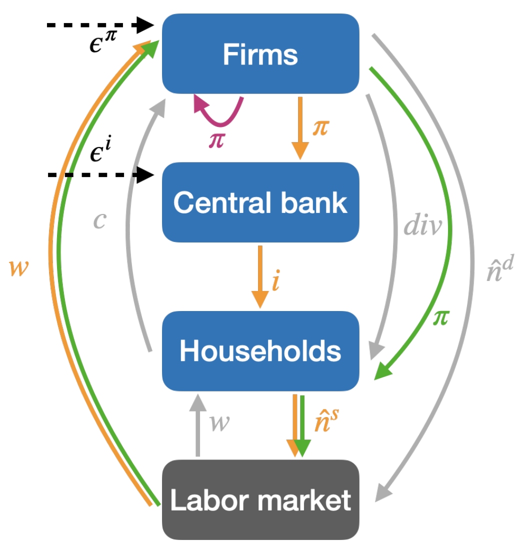
$$\overline{\mathbb{E}}[\pi_t] = \left[1 - \frac{\lambda(1-\theta) + \left(\lambda^3\rho - \lambda^4\phi\right)\theta\kappa K(\lambda,\rho)}{1-\beta\theta\rho}\right]^{-1} \epsilon_t^\pi \tag{45}$$

$$\overline{\mathbb{E}}[i_t] = \lambda\phi\overline{\mathbb{E}}[\pi_t] \tag{46}$$

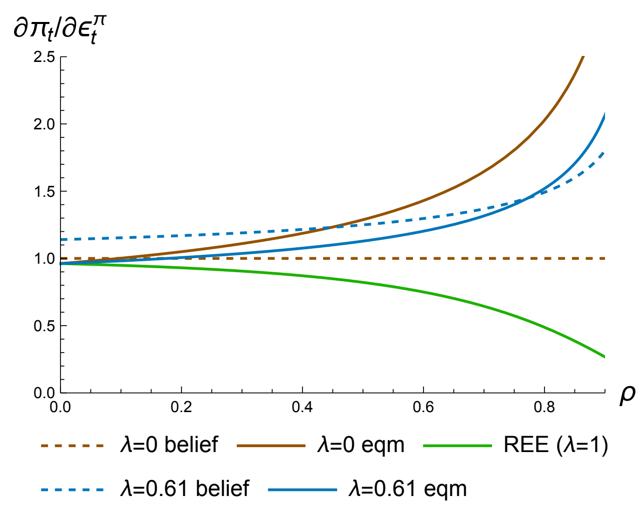
with $K(\lambda,\rho)$ increasing in λ and ρ under our calibration and $K(1,\rho) = \frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}$.

This proposition shows how exactly the persistence of shock ρ matters, nesting Proposition 4 with $\rho = 0$. A positive ρ gives rise to new terms and modifies existing terms, which we dissect in order by analyzing REE inflation (43) and shallow expectations (45).

Regarding new terms, four general equilibrium loops across sequences of variables are involved, instead of three, displayed in Figure 10a. Relative to the case with period-1 shocks (Figure 4), persistent shocks give rise to new causal relations—the responses of household consumption and labor supply to inflation π, plotted in green. This results in the fourth loop, the *inflation-on-households loop* (in green), in addition to pricing complementarity (in purple), the monetary policy loop (in orange), and the Keynesian cross (in gray). This inflation-on-household loop takes 3 steps to close and amplifies the inflation response to cost-push shocks, since a higher inflation π simultaneously encourages consumption c and discourages labor supply $\hat{n}^s$, which leads to a higher wage w, feeding into inflation π. As a result, this loop is dampened by λ^3 in expectations (45). Further, the strength of this loop is proportional to ρ, meaning that it only exits when $\rho > 0$ and is stronger when ρ is higher. That occurs because only future inflation impacts household behavior by changing the real interest rate, and the expected inflation decays with rate ρ.



(a) Causal relations in the sequence space



(b) Inflation π_t in response to cost-push shock ϵ_t^π

Figure 10: New Keynesian economy with persistent shocks

Notes: Panel (a) illustrates the four loops of causal relations across sequences of variables in the New Keynesian economy in different colors, to accompany the discussion of Proposition 8. Four loops are the pricing complementarity self-loop (in purple), the inflation-on-households loop (in green), the monetary policy loop (in orange) and the Keynesian cross (in gray).
Panel (b) plots the equilibrium inflation π_t and the average expectation thereof (relative to the size of the shock) in response to a persistent cost-push shock $\epsilon_t^\pi = \rho^t \epsilon^\pi$, as functions of the persistence ρ, under different values of λ. The dashed lines indicate expectations, and the solid lines represent the equilibrium. Different colors stand for different λ, with $\lambda = 0$ in brown, our calibration of λ in blue, and $\lambda = 1$ (rational expectations) in green. In the case with rational expectations, the beliefs coincide with the equilibrium.

Concerning the strength of these terms, a positive ρ does two things, in addition to activating the inflation-on-households loop. First, it strengthens all loops by $\frac{1}{1-\beta\theta\rho}$, because firms' pricing decisions are forward-looking and respond to a sum of future changes discounted by $\beta\theta$. Second, it further boosts the Keynesian cross via $K(\lambda, \rho)$. This additional boost occurs because households' decisions are forward-looking as well. Since the Keynesian cross reflects the feedback between firms and households, the persistence of shocks is compounded. As ρ increases, all loops get stronger, but the inflation-on-households and monetary policy loops become even stronger relative to the pricing complementarity self-loop, which has important consequences, as we will discuss next.

Equilibrium under shallow thinking. Figure 10b plots the average inflation expectation $\overline{\mathbb{E}}[\pi_t]$ and the equilibrium inflation π_t, both relative to the time-t size of the shock, as functions of persistence ρ. The dashed lines indicate expectations, and the solid lines represent the equilibrium. Different colors stand for different values of λ, with $\lambda = 0$ in brown, our calibrated λ in blue, and $\lambda = 1$ (rational expectations) in green. In the case with rational expectations, the beliefs coincide with the equilibrium.

We note from Figure 10b that two insights obtained with transitory news shocks extend here, and a new lesson emerges regarding persistence. First, inflation expectations exceed their equilibrium values under shallow thinking when shocks are not too persistent, as the dashed blue line lies above the solid blue line when ρ is small.[30] Second, cost-push shocks are more inflationary under shallow thinking than under rational expectations, since the solid blue line lies above the solid green line. These two findings generalize the previous results from transitory news shocks. Last, a persistent cost-push shock leads to higher inflation under shallow thinking, but lower inflation under rational expectations, since the solid blue line increases in ρ while the solid green line decreases in ρ.

To understand this new lesson regarding persistence, we start with the limit case of $\lambda = 0$ and $\lambda = 1$ (rational expectations). In the limit of $\lambda = 0$, inflation expectations always have the same size as the shock, with the dashed brown being flat. Inflation expectations change agents' behavior by acting like a cost-push shock for firms and a demand-shock for households. As their decisions are forward-looking and depend on discounted sums of future disturbances, a more persistent shock leads to larger changes in their behavior in any period, thus resulting in higher inflation.

[30] We also find that the average interest rate expectation is higher than its equilibrium value in response to monetary policy shocks, but is lower in response to cost-push shocks, regardless of ρ (Figure C4).

In contrast, in the limit of $\lambda = 1$ (rational expectations), as analyzed based on Proposition 8, a higher ρ strengthens all general equilibrium effects and, further, boosts the inflation-on-households and monetary policy loops relative to the pricing complementarity self-loop. The monetary policy loop offsets the inflation response, whereas the other two amplify it. As the monetary policy loop is the strongest among them and rational agents appreciate that, a higher ρ strengthens the offset and leads to a lower inflation.

Under our calibrated λ, shallow agents' inflation expectations increase in ρ. The mechanism is that, while the monetary policy offset loop is objectively the strongest, it is also the longest and gets dampened the most in expectations. Thus, shallow agents believe that a more persistent shock leads to more amplification, as they better understand the shorter amplification loops than the monetary policy offset loop. As a result, the equilibrium inflation is higher when ρ is higher. That increasing relationship is less drastic than in the extreme case of $\lambda = 0$, as shallow agents partially understand the monetary policy reaction and its effects on the economy.[31]

6 Consequences in an RBC Economy

Last, we consider an RBC economy, and show that shallow thinking amplifies its responses to persistent productivity shocks and produces a stock market boom and crash.[32]

We outline the model and shallow thinking thereof in Section 6.1, with details in Appendix D, and present the effects of productivity shocks on the macroeconomy and asset prices in Section 6.2.

6.1 Shallow Thinking in an RBC Economy

The RBC economy consists of two types of agents (firms and households) and two competitive markets (the goods market and the labor market). We study a first-order approximation around the steady state. Differing from the convention in the New Keynesian model, here we use capital letters for variables in levels, so that different components of GDP are in the same units.

[31] Figure C3 suggests that for values of λ around or below our calibrated value, a persistent cost-push shock always leads to higher inflation compared to a purely transitory shock ($\rho = 0$), whereas for values of λ near 1, the opposite holds.

[32] As we analytically applied shallow thinking to the New Keynesian economy, studying an RBC economy is also useful for illustrating its applicability to a broader class of models that can be solved using the sequence-space Jacobian developed by Auclert et al. (2021). We present the procedure in Appendix E.

As in the prior analysis of the New Keynesian economy, we assume that agents observe the sequence of shocks at time 0^- and form their expectations $\overline{\mathbb{E}}[\cdot]$ once and for all at time 0^-, i.e., $\overline{\mathbb{E}}_t[v_s] \equiv \overline{\mathbb{E}}[v_s]$ for $s > t$.

Firms. There is a continuum of firms that produce using capital K_t and labor N_t^d with a production function $Y_t = Z_t \left(\frac{K_t}{\alpha}\right)^{\alpha} \left(\frac{N_t^d}{1-\gamma}\right)^{1-\gamma}$. Firms own capital and make investment I_t to increase the capital stock in the next period, $K_{t+1} = (1-\delta) K_t + I_t$. In addition, firms are subject to capital adjustment cost $\Psi(I_t, K_t) = \frac{\psi}{2}\left(\frac{I_t}{K_t} - \delta\right)^2 K_t$, which gives rise to an investment-q relation as in Hayashi (1982), and pay dividends $DIV_t = Y_t - W_t N_t^d - I_t - \Psi(I_t, K_t)$. Firms maximize their values, i.e., the sum of dividends discounted by the gross interest rate. We use $1 + r_t$ to denote the gross interest rate from period t to $t+1$ and assume that it is known in period t, as the return of a 1-period bond in zero supply.

In each period, firms choose investment I_t, output Y_t and labor demand N_t^d, taking as given the prevailing wage W_t, interest rate r_t and productivity Z_t, as well as their expectations of the future values of these variables $\left\{\overline{\mathbb{E}}[W_{t+k}], \overline{\mathbb{E}}[r_{t+k}], Z_{t+k}\right\}_{k\geq 1}$.

Households. There is a continuum of households who live infinitely and maximize their lifetime utility, discounted by β, which is separable in consumption and labor supply. In each period, households choose consumption C_t and labor supply N_t^s, taking as given the prevailing wage W_t, interest rate r_t and dividend DIV_t, as well as their expectations of the future values of these variables $\left\{\overline{\mathbb{E}}[W_{t+k}], \overline{\mathbb{E}}[r_{t+k}], \overline{\mathbb{E}}[DIV_{t+k}]\right\}_{k\geq 1}$. The household side of the RBC economy is the same as that of the New Keynesian economy, except that they invest in a real bond as opposed to a nominal bond.

Goods and labor markets. The interest rate $\{r_t\}_{t\geq 0}$ and wage $\{w_t\}_{t\geq 0}$ arise to clear the goods and labor market, $Y_t = I_t + \Psi(I_t, K_t) + C_t, N_t^d = N_t^s$.

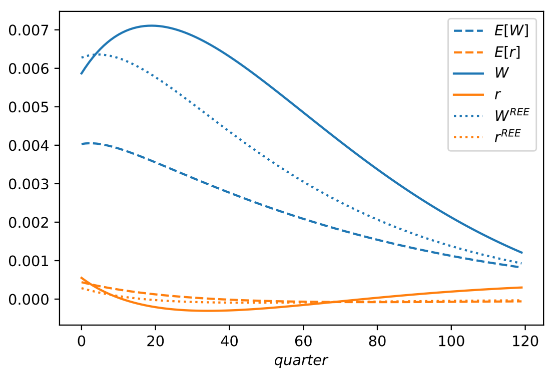
Shallow thinking. In terms of belief formation, the system of causal relations consists of firms and households' best responses, as well as the determination of the interest rate and wage by two fictitious auctioneers for the goods and labor markets, respectively. The interest rate depends on the shifts in firms' output and investment and in households' consumption ($\hat{Y}, \hat{I}$ and $\hat{C}$), each of which is a function of decision-relevant variables for firms and households other than the interest rate. Similarly, the wage is determined by the shifts in firms' labor demand and households' labor supply ($\hat{N}^d, \hat{N}^s$), both of which depend on decision-relevant variables other than the wage. Figure D1 illustrates these causal relations, which pin down the average expectations $\overline{\mathbb{E}}[\cdot]$.

Equilibrium. Given the average expectations, the equilibrium is determined by agents' best responses and market clearing conditions, period by period.

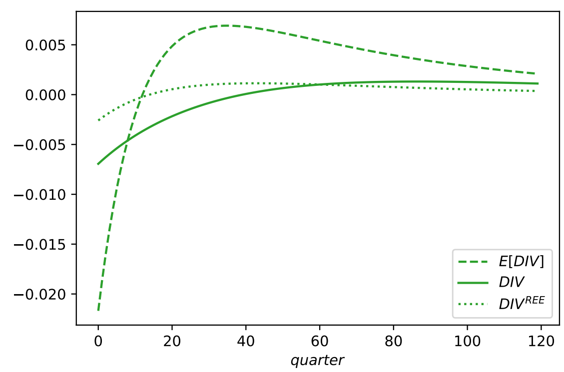
We adopt a quarterly calibration of the RBC economy, with all parameters listed in Table D1. We assume slight decreasing returns to scale in production ($\alpha < \gamma$) and the existence of a small fringe of rational agents, which we discuss in greater detail in Appendix D.

6.2 Effects of Productivity Shocks on Macroeconomy and Asset Prices

We study the effects of a persistent productivity shock, with $\rho = 0.979$ following King and Rebelo (1999), on the macroeconomy and asset prices. In particular, we show that shallow thinking amplifies the economy's responses and leads to a stock market boom and crash.



(a) Wage W and interest rate r

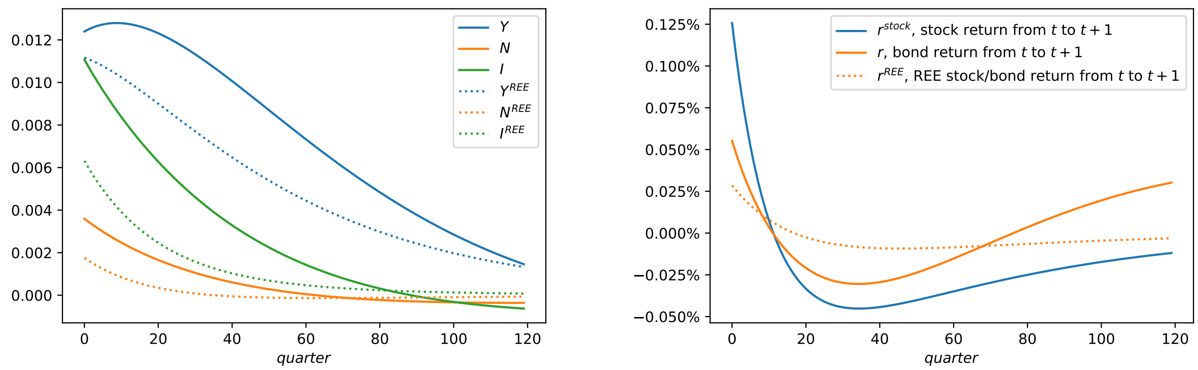


(b) Dividend DIV

Figure 11: Decision-relevant beliefs in response to productivity shock Z_t

Notes: Two panels plot the shallow thinking beliefs (in dashed lines), the shallow thinking equilibrium outcomes (in solid lines) and the rational expectations equilibrium outcomes (in dotted lines) of wage W, interest rate r, and dividend DIV, in response to a persistent productivity shock.

Figure 11 illustrates the beliefs about the interest rate r, wage W, and dividend DIV, which are relevant for firms' and households' decisions in dashed lines. In comparison, the solid lines indicate their equilibrium values, and the dotted lines indicate the rational expectations equilibrium (REE). Panel 11a suggests that shallow agents underappreciate the wage response. That occurs because agents believe that the firms will produce more by hiring and investing more in response to a productivity shock, but fail to recognize that firms' behavior will push up wages in the economy. Panel 11b shows that shallow agents perceive a much more volatile stream of dividends compared to the equilibrium values or the REE. This is because they believe firms will invest more early on by cutting dividends and pay more dividends later. However, in equilibrium, facing elevated wages and changes in interest rates, firms will not invest as much as shallow agents believe.



(a) Output Y, labor hours N and investment I

(b) Stock return r^{stock} and bond return r

Figure 12: Equilibrium outcomes in response to productivity shock Z_t

Notes: Two panels plot the shallow thinking equilibrium outcomes (in solid lines) and the rational expectations equilibrium outcomes (in dotted lines) of output Y, labor hours N, investment I, stock return r^{stock} and bond return r (which is the interest rate), in response to a persistent productivity shock. In the rational expectations equilibrium, the stock and bond returns coincide.

Figure 12 shows the responses of macroeconomic variables and asset returns in solid lines, compared to the rational expectations equilibrium in dotted lines. Panel 12a demonstrates that the output response (in blue) under shallow thinking is hump-shaped and more persistent. The responses of investment and labor hours (in green and orange) are almost twice as large.

Upon observing the productivity shock, the stock market experiences a positive revaluation at time 0^-. Panel 12b suggests that, after the revaluation, the stock excess return relative to bond from t to $t+1$, $xr_t^{stock} = r_t^{stock} - r_t$, is initially positive but turns negative after a few quarters. This occurs because agents underestimate the dividends in the short term but overestimate them in the long term, as shown in Figure 11b. This pattern of positive stock excess returns turning negative is consistent with the classic Kindleberger (1978) narrative of crises and many real-world episodes, in which technological or financial innovations lead to booms and crashes in asset markets.

Table 5: Moments of RBC impulse responses

	(1) REE	(2) ST	(3) ST firms	(4) ST households
Shock (same for all columns)				
Half life of shock	33	33	33	33
Shock size $\mathrm{d}Z_0/\overline{Z}$	1.0%	1.0%	1.0%	1.0%
Macroeconomy				
Half life of Y_t	49	69	71	66
Quarter of peak output Y_t (hump shape)	0	9	3	11
Peak output $\mathrm{d}Y_t/\overline{Y}$	1.12%	1.28%	1.22%	1.98%
Peak hours $\mathrm{d}N_t/\overline{N}$	0.18%	0.36%	0.32%	1.15%
Peak investment $\mathrm{d}I_t/\overline{I}$	3.54%	6.2%	5.66%	17.68%
Peak consumption $\mathrm{d}C_t/\overline{C}$	0.8%	0.83%	0.79%	1.33%
Asset prices				
Peak stock price $\mathrm{d}P_t/\overline{P}$	0.85%	2.55%	2.59%	2.59%
Quarter of excess return xr_t^{stock} turning negative	$xr_t^{stock} = 0$ always	12	9	$xr_t^{stock} < 0$ always

Notes: Four columns correspond to the rational expectations equilibrium (REE), the shallow thinking equilibrium (ST), a model in which firms are shallow but households are rational, and a model in which households are shallow but firms are rational. In the latter two cases which feature belief disagreement, to illustrate the stock market dynamics, we simply assume that the stock is priced by the shallow agents.

Table 5 summarizes the key moments of the impulse responses to a persistent productivity shock, with the rational expectations equilibrium in column (1) and shallow thinking equilibrium in (2). Quantitatively, shallow thinking leads to amplified and more persistent responses in labor hours, investment and the stock price relative to the rational expectations equilibrium. Qualitatively, it produces hump-shaped responses in consumption and output, as well as a stock market boom and crash, which are absent under rational expectations. Columns (3) and (4) further examine cases where only firms or households exhibit shallow thinking, suggesting that both parties' beliefs are important.

7 Conclusion

This paper develops a theory of *shallow thinking* as the structure of belief formation, supports its empirical content using a customized survey, and illustrates its consequences for macroeconomics and finance.

The key implication of our theory is that more distant causal relations and longer feedback loops have less influence on beliefs. Our estimation suggests that, on average, people think about only 2.6 steps of shock propagation—far below infinity assumed by rational expectations. They ignore much of the economy and underappreciate feedback loops. While our primary contribution is to develop a psychologically grounded model of expectations for macroeconomic analysis, our study of the understanding of macroeconomy also advances the causal reasoning literature (Waldmann, 2017). This literature typically presents participants with simple examples in short experiments, whereas we examine a real-world domain involving many variables and long-term data accumulation.

Our theory leads to a rich set of consequences. In a New Keynesian model, long-term nominal rates underreact to cost-push shocks but overreact to monetary shocks, as agents underappreciate shock propagation and offsetting loops. This insight reconciles multiple bond market puzzles. Additionally, cost-push shocks are more inflationary than under rational expectations, and more persistent cost-push shocks lead to higher inflation—contrary to rational expectations' predictions. That occurs because shallow agents better understand a short feedback loop that amplifies cost-push shocks, relative to a long offsetting loop. In a real business cycle model, shallow thinking amplifies fluctuations in response to productivity shocks and leads to a stock market boom and crash.

At a high level, we acknowledge the immense complexity of the economy. If it has taken decades for our best economists to understand how it works—or if they are still figuring it out—we must carefully consider how much the average person understands. We focus on the qualitative aspects and positive consequences of shallow thinking. Our theory can be applied to more complex macroeconomic models to answer quantitative and normative questions, such as business cycle accounting and the design of optimal stabilization policy. It may also be fruitfully applied to general equilibrium models in trade, spatial economics, and other fields. We hope this research agenda can make meaningful contributions to economics.

Bibliography

Afrouzi, Hassan, Spencer Y. Kwon, Augustin Landier, Yueran Ma, and David Thesmar. 2023. "Overreaction in Expectations: Evidence and Theory." *The Quarterly Journal of Economics*, 138(3): 1713–1764.

Ahn, Woo-kyoung, Nancy S. Kim, Mary E. Lassaline, and Martin J. Dennis. 2000. "Causal Status as a Determinant of Feature Centrality." *Cognitive Psychology*, 41(4): 361–416.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 2022. "Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples." *The Review of Economic Studies*, 89(6): 2958–2991.

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart. 2024. "Narratives about the Macroeconomy." *Working Paper*.

Andre, Peter, Philipp Schirmer, and Johannes Wohlfart. 2024. "Mental Models of the Stock Market." *Working Paper*.

Angeletos, George-Marios, and Chen Lian. 2018. "Forward Guidance without Common Knowledge." *American Economic Review*, 108(9): 2477–2512.

Angeletos, George-Marios, and Chen Lian. 2023*a*. "Dampening General Equilibrium: Incomplete Information and Bounded Rationality." In *Handbook of Economic Expectations*. 613–645.

Angeletos, George-Marios, and Chen Lian. 2023*b*. "Determinacy without the Taylor Principle." *Journal of Political Economy*, 131(8): 2125–2164.

Angeletos, George-Marios, and Jennifer La'O. 2009. "Incomplete Information, Higher-Order Beliefs and Price Inertia." *Journal of Monetary Economics*, 56(SUPPL.): S19–S37.

Angeletos, George-Marios, and Karthik A. Sastry. 2021. "Managing Expectations: Instruments Versus Targets." *The Quarterly Journal of Economics*, 136(4): 2467–2532.

Angeletos, George-Marios, and Zhen Huo. 2021. "Myopia and Anchoring." *American Economic Review*, 111(4): 1166–1200.

Arad, Ayala, and Ariel Rubinstein. 2012. "The 11–20 Money Request Game: A Level-*k* Reasoning Study." *American Economic Review*, 102(7): 3561–3573.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models." *Econometrica*, 89(5): 2375–2408.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model." *NBER Working Paper*.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2024. "The Intertemporal Keynesian Cross." *Journal of Political Economy*, , (Accepted).

Azeredo da Silveira, Rava, Yeji Sung, and Michael Woodford. 2024. "Optimally Imprecise Memory and Biased Forecasts." *American Economic Review*, 114(10): 3075–3118.

Ballester, Coralio, Antonio Rodriguez-Moral, and Marc Vorsatz. 2024. "Cognitive Reflection in Experimental Anchored Guessing Games." *Games and Economic Behavior*, 148: 179–195.

Barberis, Nicholas, Andrei Shleifer, and Robert W. Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics*, 49(3): 307–343.

Bastianello, Francesca, and Paul Fontanier. 2024. "Expectations and Learning from Prices." *Review of Economic Studies*, , (Forthcoming).

Bauer, Michael D., Carolin E. Pflueger, and Adi Sunderam. 2024*a*. "Changing Perceptions and Post-Pandemic Monetary Policy." *Proceedings - Economic Policy Symposium - Jackson Hole*.

Bauer, Michael D., Carolin E. Pflueger, and Adi Sunderam. 2024*b*. "Perceptions About Monetary Policy." *The Quarterly Journal of Economics*, 139(4): 2227–2278.

Ben Zeev, Nadav, Valerie A. Ramey, and Sarah Zubairy. 2023. "Do Government Spending Multipliers Depend on the Sign of the Shock?" *AEA Papers and Proceedings*, 113: 382–387.

Bianchi-Vimercati, Riccardo, Martin Eichenbaum, and Joao Guerreiro. 2024. "Fiscal Stimulus with Imperfect Expectations: Spending vs. Tax Policy." *Journal of Economic Theory*, 217: 105814.

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *The Quarterly Journal of Economics*, 117(4): 1329–1368.

Blanchard, Olivier J., and Jordi Galí. 2010. "The Macroeconomic Effects of Oil Price Shocks: Why Are the 2000s so Different from the 1970s?" In *International Dimensions of Monetary Policy*. , ed. Jordi Galí and Mark Gertler, 373–421. Chicago: University of Chicago Press / (National Bureau of Economic Research Conference Report).

Bocola, Luigi, Alessandro Dovis, Kasper Jørgensen, and Rishabh Kirpalani. 2024. "Bond Market Views of the Fed." *Working Paper*.

Boivin, Jean, Michael T. Kiley, and Frederic S. Mishkin. 2010. "How Has the Monetary Transmission Mechanism Evolved Over Time?" In *Handbook of Monetary Economics*. Vol. 3, 369–422. Elsevier Ltd.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review*, 110(9): 2748–2782.

Breza, Emily, Arun Chandrasekhar, and Alireza Tahbaz-Salehi. 2018. "Seeing the Forest for the Trees? An Investigation of Network Knowledge." *NBER Working Paper*, w24359.

Camerer, Colin, and Dan Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 89(1): 306–318.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *The Quarterly Journal of Economics*, 119(3): 861–898.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko. 2024. "The Inflation Expectations of U.S. Firms: Evidence from a New Survey." *Journal of Monetary Economics*, 145: 103569.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2024. "On Binscatter." *American Economic Review*, 114(5): 1488–1514.

Cieslak, Anna. 2018. "Short-Rate Expectations and Unexpected Returns in Treasury Bonds." *The Review of Financial Studies*, 31(9): 3265–3306.

Cieslak, Anna, and Pavol Povala. 2015. "Expected Returns in Treasury Bonds." *Review of Financial Studies*, 28(10): 2859–2901.

Clarida, Richard, Jordi Galí, and Mark Gertler. 1999. "The Science of Monetary Policy: A New Keynesian Perspective." *Journal of Economic Literature*, 37(4): 1661–1707.

Cloyne, James, Ezgi Kurt, and Paolo Surico. 2023. "Who Gains from Corporate Tax Cuts?" *NBER Working Paper*, w31278.

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico. 2023. "Do Tax Increases Tame Inflation?" *AEA Papers and Proceedings*, 113: 377–381.

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico. 2024. "Taxes, Innovation and Productivity." *Working Paper*.

Cochrane, John H., and Monika Piazzesi. 2002. "The Fed and Interest Rates—A High-Frequency Identification." *American Economic Review*, 92(2): 90–95.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8): 2644–2678.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten Van Rooij. 2023*a*. "How Does Consumption Respond to News about Inflation? Field Evidence from a Randomized Control Trial." *American Economic Journal: Macroeconomics*, 15(3): 109–152.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar. 2018. "How Do Firms Form Their Expectations? New Survey Evidence." *American Economic Review*, 108(9): 2671–2713.

Coibion, Olivier, Yuriy Gorodnichenko, Edward S. Knotek, and Raphael Schoenle. 2023*b*. "Average Inflation Targeting and Household Expectations." *Journal of Political Economy Macroeconomics*, 1(2): 403–446.

Coibion, Olivier, Yuriy Gorodnichenko, Saten Kumar, and Jane Ryngaert. 2021. "Do You Know That I Know That You Know. . . ? Higher-Order Beliefs in Survey Data." *The Quarterly Journal of Economics*, 136(3): 1387–1446.

Cooper, Ilan, and Richard Priestley. 2009. "Time-Varying Risk Premiums and the Output Gap." *Review of Financial Studies*, 22(7): 2801–2833.

Corsetti, Giancarlo, André Meier, and Gernot J. Müller. 2012. "Fiscal Stimulus with Spending Reversals." *Review of Economics and Statistics*, 94(4): 878–895.

D'Acunto, Francesco, Daniel Hoang, and Michael Weber. 2022. "Managing Households' Expectations with Unconventional Policies." *The Review of Financial Studies*, 35(4): 1597–1642.

Dal Bó, Ernesto, Pedro Dal Bó, and Erik Eyster. 2018. "The Demand for Bad Policy When Voters Underappreciate Equilibrium Effects." *The Review of Economic Studies*, 85(2): 964–998.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2022. "General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya." *Econometrica*, 90(6): 2603–2643.

Eusepi, Stefano, and Bruce Preston. 2018. "The Science of Monetary Policy: An Imperfect Knowledge Perspective." *Journal of Economic Literature*, 56(1): 3–59.

Evans, George W., and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics. Frontiers of Economic Research*, Princeton:Princeton University Press.

Farhi, Emmanuel, and Iván Werning. 2019. "Monetary Policy, Bounded Rationality, and Incomplete Markets." *American Economic Review*, 109(11): 3887–3928.

Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *The Quarterly Journal of Economics*, 129(4): 1661–1710.

Gabaix, Xavier. 2020. "A Behavioral New Keynesian Model." *American Economic Review*, 110(8): 2271–2327.

Gabaix, Xavier, and David I. Laibson. 2001. "The 6D Bias and the Equity-Premium Puzzle." *NBER Macroeconomics Annual*, 16: 257–312.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications - Second Edition.* Princeton University Press.

Galí, Jordi, J. David López-Salido, and Javier Vallés. 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association*, 5(1): 227–270.

García-Schmidt, Mariana, and Michael Woodford. 2019. "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis." *American Economic Review*, 109(1): 86–120.

Georganas, Sotiris, Paul J. Healy, and Roberto A. Weber. 2015. "On the Persistence of Strategic Sophistication." *Journal of Economic Theory*, 159: 369–400.

Greenwood, Robin, and Samuel G. Hanson. 2015. "Waves in Ship Prices and Investment." *The Quarterly Journal of Economics*, 130(1): 55–109.

Gürkaynak, Refet S., Brian Sack, and Eric Swanson. 2005. "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models." *American Economic Review*, 95(1): 425–436.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2023. "Designing Information Provision Experiments." *Journal of Economic Literature*, 61(1): 3–40.

Haavelmo, Trygve. 1943. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica*, 11(1): 1.

Hansen, Lars Peter, and Thomas J. Sargent. 2007. *Robustness.* . Course Book ed., Princeton, N.J:Princeton University Press.

Hanson, Samuel G., and Jeremy C. Stein. 2015. "Monetary Policy and Long-Term Real Rates." *Journal of Financial Economics*, 115(3): 429–448.

Hanson, Samuel G., David O. Lucca, and Jonathan H. Wright. 2021. "Rate-Amplifying Demand and the Excess Sensitivity of Long-Term Rates." *Quarterly Journal of Economics*, 136(3): 1719–1781.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica*, 50(1): 213.

Heckman, James, and Rodrigo Pinto. 2015. "Causal Analysis after Haavelmo." *Econometric Theory*, 31(1): 115–151.

Imbens, Guido W. 2020. "Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics." *Journal of Economic Literature*, 58(4): 1129–1179.

Iovino, Luigi, and Dmitriy Sergeyev. 2023. "Central Bank Balance Sheet Policies Without Rational Expectations." *Review of Economic Studies*, 90(6): 3119–3152.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton. 2014. "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks." *The Journal of Finance*, 69(3): 1197–1233.

Känzig, Diego R. 2021. "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements." *American Economic Review*, 111(4): 1092–1125.

Kendall, Chad, and Ryan Oprea. 2024. "On the Complexity of Forming Mental Models." *Quantitative Economics*, 15(1): 175–211.

Kilian, Lutz, and Robert J. Vigfusson. 2011. "Are the Responses of the U.S. Economy Asymmetric in Energy Price Increases and Decreases?: Are Responses of the U.S. Economy Asymmetric?" *Quantitative Economics*, 2(3): 419–453.

Kindleberger, Charles P. 1978. *Manias, Panics, and Crashes: A History of Financial Crises.* New York: Basic Books.

King, Robert G., and Sergio T. Rebelo. 1999. "Chapter 14 Resuscitating Real Business Cycles." In *Handbook of Macroeconomics*. Vol. 1, 927–1007.

Kneeland, Terri. 2015. "Identifying Higher-Order Rationality." *Econometrica*, 83(5): 2065–2079.

Lucas, Jr., Robert E. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory*, 4(2): 103–124.

Ludvigson, Sydney C., and Serena Ng. 2009. "Macro Factors in Bond Risk Premia." *Review of Financial Studies*, 22(12): 5027–5067.

Lusardi, Annamaria, and Olivia S. Mitchell. 2011. "Financial Literacy and Retirement Planning in the United States." *Journal of Pension Economics and Finance*, 10(4): 509–525.

Maćkowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal Sticky Prices under Rational Inattention." *American Economic Review*, 99(3): 769–803.

Mankiw, N. Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *The Quarterly Journal of Economics*, 117(4): 1295–1328.

Mendes, Arthur, Wataru Miyamoto, Thuy Lan Nguyen, Steven Pennings, and Leo Feler. 2024. "The Macroeconomic Effects of Cash Transfers: Evidence from Brazil." *Working Paper*.

Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review*, 103(4): 1212–1247.

Miao, Jianjun, Jieran Wu, and Eric R. Young. 2022. "Multivariate Rational Inattention." *Econometrica*, 90(2): 907–945.

Miranda-Agrippino, Silvia, and Giovanni Ricco. 2021. "The Transmission of Monetary Policy Shocks." *American Economic Journal: Macroeconomics*, 13(3): 74–107.

Molavi, Pooya. 2019. "Macroeconomics with Learning and Misspecification: A General Theory and Applications." *Working Paper*.

Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *The American Economic Review*, 85(5): 1313–1326.

Nimark, Kristoffer. 2008. "Dynamic Pricing and Imperfect Common Knowledge." *Journal of Monetary Economics*, 55(2): 365–382.

Olivei, Giovanni, and Silvana Tenreyro. 2007. "The Timing of Monetary Policy Shocks." *American Economic Review*, 97(3): 636–663.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference.* . 2 ed., Cambridge University Press.

Pennings, Steven. 2021. "Cross-Region Transfer Multipliers in a Monetary Union: Evidence from Social Security and Stimulus Payments." *American Economic Review*, 111(5): 1689–1719.

Ramey, Valerie A. 2011. "Identifying Government Spending Shocks: It's All in the Timing." *The Quarterly Journal of Economics*, 126(1): 1–50.

Ramey, Valerie A. 2013. "Government Spending and Private Activity." In *Fiscal Policy after the Financial Crisis*. , ed. Alberto Alesina and Francesco Giavazzi, 19–55. University of Chicago Press.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics*. Vol. 2, 71–162. Elsevier B.V.

Rottman, Benjamin Margolin, and Reid Hastie. 2014. "Reasoning about Causal Relationships: Inferences on Causal Networks." *Psychological Bulletin*, 140(1): 109–139.

Rozenblit, Leonid, and Frank Keil. 2002. "The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth." *Cognitive Science*, 26(5): 521–562.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics*, 50(3): 665–690.

Spiegler, Ran. 2016. "Bayesian Networks and Boundedly Rational Expectations." *The Quarterly Journal of Economics*, 131(3): 1243–1290.

Spiegler, Ran. 2020. "Behavioral Implications of Causal Misperceptions." *Annual Review of Economics*, 12(1): 81–106.

Stahl, II, Dale O., and Paul W. Wilson. 1994. "Experimental Evidence on Players' Models of Other Players." *Journal of Economic Behavior & Organization*, 25(3): 309–327.

Stahl, II, Dale O., and Paul W. Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10(1): 218–254.

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?" *The Quarterly Journal of Economics*, 136(4): 2309–2369.

Stantcheva, Stefanie. 2023*a*. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." *Annual Review of Economics*, 15(1): 205–234.

Stantcheva, Stefanie. 2023*b*. "Understanding of Trade." *NBER Working Paper*.

Tinbergen, Jan. 1930. "Determination and Interpretation of Supply Curves: An Example [Bestimmung Und Deutung von Angebotskurven. Ein Beispiel]." *Zeitschrift fur Nationalokonomie*, 1(5): 669–679.

Waldmann, Michael R., ed. 2017. *The Oxford Handbook of Causal Reasoning.* Vol. 1, Oxford University Press.

Woodford, Michael. 2003*a*. "Imperfect Common Knowledge and the Effects of Monetary Policy." In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*. , ed. Philippe Aghion, Roman Frydman, Joseph E. Stiglitz and Michael Woodford, 25–58. Princeton:Princeton University Press.

Woodford, Michael. 2003*b*. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

Wu, Lingxuan. 2023. "Mental Macro-Finance Models: Evidence and Theory." *SSRN Electronic Journal*.

Online Appendix for
"Thinking about the Economy, Deep or Shallow?"

Pierfrancesco Mei Lingxuan Wu

November 6, 2024

A Survey Details

A.1 Detailed Survey Design and Sample

Detailed survey design. We design our survey to elicit the general public's directional beliefs about changes in a host of macroeconomic variables (such as prices, labor hours, and interest rates) in response to a set of hypothetical macroeconomic shocks (such as oil shocks and monetary policy shocks).

Our survey builds on Andre et al. (2022), which ask respondents to forecast changes (in levels) of inflation and the unemployment rate in response to hypothetical macroeconomic shocks.[33] Grounded in our theory, two innovative features of our design are to inquire about a number of major macroeconomic variables and to only elicit beliefs about the directional responses. Inquiring about a host of macroeconomic variables traces out the path of shock propagation and tests if perception fails sequentially as suggested by our theory. For example, one of our questions concerns labor hours and asks respondents whether the average worker will work for more, fewer, or the same number of hours during a typical week in response to shocks. Comparing people's beliefs about labor hours to those about aggregate demand sheds light on their understanding of firms' input choice. Eliciting directional assessments instead of level forecasts lowers the cognitive strain and lets us focus on the qualitative aspect of people's mental models. For instance, their responses about directional price change reveal whether participants understand the Phillips curve, rather than their potentially different perceptions about its slope.[34]

[33] Haaland, Roth and Wohlfart (2023) discuss the potential and limitations of hypothetical vignettes.

[34] Andre et al. (2022) further ask participants to select the relevant ones from a list of potential channels and show that the selected channels predict their forecasts. For example, one channel in the oil shock vignette is "due to lower incomes or job loss, households cut back on their spending." We simply ask for directional assessments for a set of variables, without showing them any directional statement that is objectively true or false, which may influence their responses.

Respondents are randomized into three groups with probability one half, one quarter and one quarter, each receiving two shock scenarios in random order. Groups 1, 2, and 3 receive the following pairs of shocks respectively: oil shock and monetary policy shock, government spending shock and personal income tax shock, corporate income tax shock and transfer payment shock. In each scenario, we introduce a shock that realizes now and persists for 12 months (except for the transfer payment shock), and then elicit respondents' beliefs about changes in the economy over the next 12 months. We stress the exogeneity of the shocks and state clearly that the shocks are publicly announced or broadcasted and are common knowledge to everyone in the US. Appendix A.2 shows the phrasing of these hypothetical shocks.

Our questions cover a large set of macroeconomic variables, divided into four blocks presented in random order, each on a separate page. The four blocks correspond to choices and decision-relevant variables of firms, households, the central bank, and the federal government's fiscal policy. We ask people's opinions about the average US business and household, to avoid any potential peculiarity of their own situations. Each block contains variables that the block either responds to or decides on, summing up to 12 to 16 distinct variables for each shock scenario, listed in Table B1. Several key variables, such as price and total labor hours, are included in more than one block (as one type of agents' choices and as other agents' decision-relevant variables), resulting in a total of about 22 questions for each shock scenario. For each question, respondents select among "up," "down," "unchanged," or "I don't know" to indicate their perceived directional changes of the specific variable in response to the shocks. The directional responses of most variables we elicit to these shocks are well-established in the empirical literature, as surveyed in Table B2.

In addition, to contrast our depth of thinking against level-k thinking in the game theoretical literature, we play the popular game of "guess 2/3 of the average" à la Nagel (1995) to measure respondents' game theoretical sophistication. Each respondent selects a number between 0 and 100, and they are informed that the number closest to the 2/3 of the average wins the game.[35] Based on each respondent's answer g_i, we compute their

[35]Our baseline design runs this game without incentives for two reasons. First, offering a prize requires collecting respondents' email addresses, which some may be averse to, leading to selection bias. Second, since we cannot monitor whether respondents use search engines or other sources (though we ask them not to), we choose not to link compensation to performance and inform them of this. Nonetheless, we conducted an incentivized version with a subsample of 300 respondents. In this subsample, we confirm our additional finding that depth of thinking correlates across shock scenarios but does not correlate with level-k thinking.

level-k as $k_i \equiv \log\left(\frac{g_i}{50}\right)/\log\left(\frac{2}{3}\right)$, assuming that a level-0 player randomly selects a number.

We also assess respondents' financial literacy using questions from Lusardi and Mitchell (2011) and collect other information, including gender, age, race, ZIP code, household composition, education, main occupation and additional employment, political affiliation, household income, assets, and debts. Figure A1 illustrates the flow of our survey.

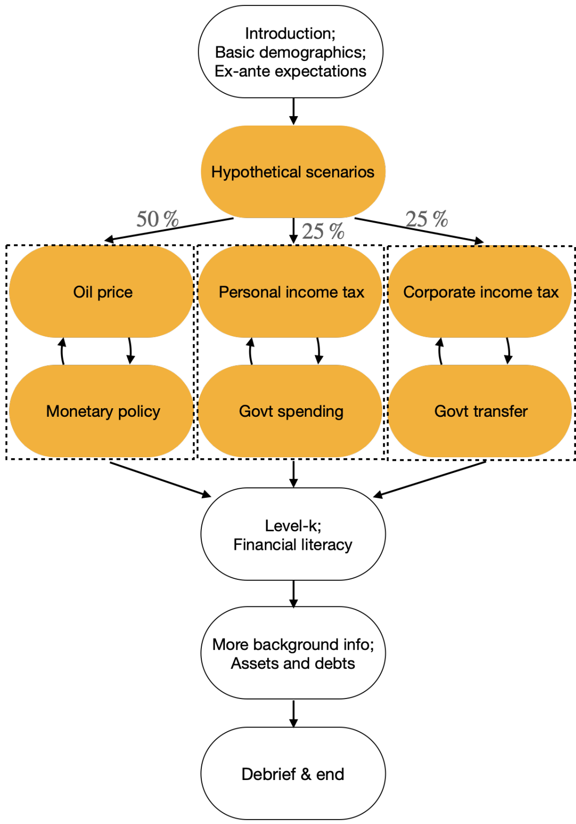


Figure A1: Survey structure

Sample. We conducted an online survey of 1,000 US households in June and October 2024. The survey was distributed through LUCID Marketplace, a platform that is widely used for research and is made up of hundreds of suppliers with a diverse set of recruitment methodologies, ensuring that the sample does not overweigh any particular segment of the population. The vast majority are double opt-in suppliers. Suppliers incentivize their respondents with loyalty reward points or gift cards or cash payments. The median completion time for our survey is 16 minutes.

We focus on US residents who are in the labor force at the time of the interview, and are aged between 25 and 65. Conditional on these characteristics, the survey is constrained through quotas to be broadly representative of the US population along the dimensions of gender, age, total gross household income, and race, based on the Annual Social and Economic Supplement data of the Current Population Survey in 2022. Table A1 notes that our sample is largely representative of the US population.[36]

We drop people who fail the attention or sanity checks placed in our survey or spend too much or too little time completing the survey to form our main sample. Our results are robust to various sample selection criteria.

Table A1: Sample statistics

	US population	Survey sample
Gender		
Male	.53	.48
Age		
25-29 years old	.13	.13
30-39 years old	.28	.29
40-49 years old	.25	.29
50-59 years old	.24	.21
60-65 years old	.10	.07
Household income		
$0-$19,999	.04	.11
$20,000-$39,999	.11	.17
$40,000-$69,999	.20	.24
$70,000-$124,999	.29	.34
$125,000+	.36	.15
Race		
White	.61	.72
Black/African-American	.12	.13
Hispanic/Latino	.18	.08
Asian/Asian-American	.07	.03
Other	.02	.03
Employment status		
Full time employed	.78	.72
Part time employed	.09	.12
Self-employed	.10	.09
Unemployed	.03	.07

Notes: Shares may not exactly add up to 1 due to rounding errors.

[36] As known in the literature (Stantcheva, 2023*a*), online samples are hard to reach the tails of income distribution and tend to skew towards white and non-Hispanic respondents. Further, we show in Appendix B.3 household income does not correlate with depth of thinking.

A.2 Hypothetical Vignettes

Group 1.

Oil shock (oil). *"Since the beginning of 2024, the average price for one barrel of WTI crude oil, which is a major benchmark for oil prices in the US, has been around $80.*

Now, imagine that the price of crude oil unexpectedly increases due to production problems in the Middle East. For the next 12 months, the price for one barrel of crude oil will be, on average, $20 higher than its current level.

This price increase is publicly broadcasted by major news outlets and is common knowledge to everyone in the US.

We will now ask you a few short questions to understand how you think the US economy would be affected by such an increase in oil price."

Monetary policy shock (MP). *"The Federal Reserve, often referred to simply as "the Fed," is the central bank of the United States that conducts the nation's monetary policy to help regulate the economy. It sets a key interest rate known as the Federal Funds Rate. This is the rate at which banks lend to each other, and it affects the economy in many ways. The Federal Funds Rate influences the interest rates for savings accounts, credit card balances, mortgages, loans, and others. As of now, the Federal Funds Rate set by the Federal Reserve is 4.75%.*

Now, imagine that the Federal Reserve unexpectedly raises the Federal Funds Rate by 0.5 percentage points, changing it to 5.25%, and announces that it will maintain this rate for the next 12 months.

This interest rate raise is publicly announced and is common knowledge to everyone in the US. The Federal Reserve clarifies that this decision is made with no changes in their assessment of economic conditions.

We will now ask you a few short questions to understand how you think the US economy would be affected by this raise in the interest rate."

Group 2.

Government spending shock (G). *"Since 2000, US federal government spending has averaged about 25% of the US Gross Domestic Product (GDP), which is the total value of all goods and services produced in the country.*

Now, imagine that the US federal government unexpectedly announces a new defense program, leading to an increase in federal government spending over the next 12 months. And the additional spending will be directed domestically. Specifically, federal government spending relative to US GDP will increase by about 2% over the next 12 months.

This increase is publicly announced and is common knowledge to everyone in the US. The government clarifies that this change is temporary and occurs without any alterations in its assessment of national security or economic conditions.

We will now ask you a few short questions to understand how you think the US economy would be affected by this increase in federal government spending."

Personal income tax shock (PIT). *"In 2023, a typical household earning the median income is subject to a 12% federal personal income tax rate. Collectively, all households paid about 2.2 trillion dollars in federal personal income taxes in 2023, which is about 8% of the US Gross Domestic Product (GDP), the total value of all goods and services produced in the country.*

Now, imagine that the US federal government unexpectedly announces a 2% increase in the federal personal income tax rate over the next 12 months.

This increase is publicly announced and is common knowledge to everyone in the US. The government clarifies that the change is temporary and occurs without any changes in its assessment of economic conditions.

We will now ask you a few short questions to understand how you think the US economy would be affected by this increase in the federal personal income tax rate."

Group 3.

Transfer payment shock (TP). *"Imagine that today the US federal government unexpectedly announces that each taxpayer will receive a one-time transfer payment worth, on average, $1,200. This one-time payment, which will not be taxed, will be available in bank accounts or as a check in mailboxes within three months.*

Taking into account that around 200 million US taxpayers will receive the payment, the total payments disbursed will be approximately 240 billion dollars, which is about 1% of the US Gross Domestic Product (GDP), the total value of all goods and services produced in the country.

The payment is publicly announced and is common knowledge to everyone in the US. The government clarifies that the payment is a one-time event and occurs without any changes in its assessment of economic conditions.

We will now ask you a few short questions to understand how you think the US economy would be affected by this transfer payment by the federal government"

Corporate income tax shock (CIT). *"Since 2017, there has been a 21% federal corporate income tax rate in place. The taxable income is a business's revenue minus expenses. In 2023, all corporations together paid about 420 billion dollars, which is about 1.5% of the US Gross Domestic Product (GDP), the total value of all goods and services produced in the country.*

Now, imagine that the US federal government unexpectedly announces a 2% increase in the federal corporate income tax rate over the next 12 months.

The increase is publicly announced and is common knowledge to everyone in the US. The government clarifies that the change is temporary and occurs without any changes in its assessment of economic conditions.

We will now ask you a few short questions to understand how you think the US economy would be affected by the increase in federal corporate income tax rate."

A.3 Other Questions

Financial literacy. Three questions are asked for each respondent.

1. *"Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?: more than today; less than today; exactly the same; don't know."*

2. *"Do you think that the following statement is 'true' or 'false'? Buying a company stock usually provides a safer return than a stock mutual fund.: true; false; don't know."*

3. *"Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?: more than $102; less than $102; exactly the same; don't know."*

Level-*k* thinking ("guess 2/3 of the average" game). *"Imagine you are playing a game with about 300 other people chosen randomly from across the United States.*

Please choose a number between 0 and 100, inclusive.

We will take your number, as well as the numbers chosen by other participants, to calculate the average number. The winning number will be the number that is closest to two-thirds (2/3) of the *average number. Specifically, we sum the chosen numbers by everyone and divide by the number of participants. Multiply the result by 2/3. The winning number is the one closest to the last result.*

The winner will get an electronic gift card at any popular merchant worth $30, which will be split when there are multiple winners. The winner will be contacted in a few days at the conclusion of this study, using the email provided below."

B Appendix to Survey Findings

B.1 Variable Directional Response and Distance

Table B1: Variables elicited in forecast part of our survey

		Group 1		Group 2		Group 3	
Variable	Abbrev.	Oil ↑	MP ↑	G ↑	PIT ↑	CIT ↑	TP ↑
Firms-related	**bus**						
Nominal marginal cost	mc	✓	✓	✓	✓	✓	✓
Demand	Y	✓	✓	✓	✓	✓	✓
Interest rate	i	✓	✓	✓	✓	✓	✓
Corporate income tax rate	CIT	✓	✓	✓	✓	✓	✓
Prices	p	✓	✓	✓	✓	✓	✓
Intermediate inputs	x	✓	✓	✓	✓	✓	✓
Investment	I	✓	✓	✓	✓	✓	✓
Total hours	N	✓	✓	✓	✓	✓	✓
Unemployment rate	u	✓	✓	✓	✓	✓	✓
Dividends/post-tax profits	div	✓	✓	✓	✓	✓	✓
Households-related	**hh**						
Interest rate	i	✓	✓	✓	✓	✓	✓
Prices	p	✓	✓	✓	✓	✓	✓
Hours	N	✓	✓	✓	✓	✓	✓
Personal income tax rate	PIT	✓	✓	✓	✓	✓	✓
Pre-tax nominal wage	W	✓	✓	✓	✓	✓	✓
Durable consumption	D	✓	✓	✓	✓	✓	✓
Non-durable consumption	ND	✓	✓	✓	✓	✓	✓
Central-bank-related	**fed**						
Unemployment rate	u	✓	✓	✓	✓	✓	✓
Inflation	p	✓	✓	✓	✓	✓	✓
Interest rate	i	✓	×	✓	✓	✓	✓
Government-related	**gov**						
Borrowing/repayment	B	✓	✓	✓	✓	✓	✓
Tax revenue	TR	✓	✓	✓	✓	✓	✓

Table B2: Literature review of directional impulse responses

		Oil price ↑	Monetary policy (MP) ↑	Transfer payment (TP)‡ ↑
Output	Y	Down (Känzig)	Down (Ramey)	Up (Pennings)
Interest rate	i	Up (Känzig)	Up (Ramey, as shock itself)	
Price	p	Up (Känzig)	Down (MAR), insignificant or up (Ramey, classic price puzzle)	
Unemployment	u	Up (Känzig)	Up (Ramey)	Down (MMNPF)
Labor hours	N	Down (BG)	Down (MAR)	
Nonresidential investment	I	Down (Känzig)	Down (BKM)	
Durable consumption	D	Down (Känzig)	Down (BKM)	Up (EHMNW)
Nondurables & services	ND	Down (Känzig)	Down (BKM)	Up (EHMNW)
Nominal wage	W	Up (BG)	Down (OT, as real wage and price both down), insignificantly down (MAR)	Up (EHMNW)
Dividend/post-tax profits	div			Up (EHMNW)
References		Känzig (2021, figs 8-10/A.7) Blanchard and Galí (2010, fig 7.6.A)	Ramey (2016, figs 2-3) Miranda-Agrippino and Ricco (2021, fig 7) Boivin, Kiley and Mishkin (2010, fig 4, post-84) Olivei and Tenreyro (2007, figs 10-14)	Pennings (2021, tab 1) Mendes et al. (2024, tab 5) Egger et al. (2022, tabs 1, 3)

		Government spending (G) ↑	Personal income tax (PIT) ↑	Corporate income tax (CIT) ↑
Output	Y	Up (Ramey16)	Down (MR)	Down (MR)
Interest rate	i	Insignificant (Ramey11)	Insignificant (MR)	Insignificant (MR)
Price	p	Up (Ramey16 unreported result)	Down (CMMS23), insignificantly down (MR)	Insignificantly up (CMMS23, MR)
Unemployment	u	Down (Ramey13*)	Up (MR)	Up (CKS), insignificantly up (MR)
Labor hours	N	Up (Ramey11*)	Down (CMMS24, MR)	Insignificantly down (CMMS24), insignificant (MR)
Nonresidential investment	I	Down (Ramey11)	Down (MR)	Down (MR)
Durable consumption	D	Down (Ramey11**)	Down (MR)	Down (CMMS24†), insignificantly up (MR)
Nondurables & services	ND	Down (Ramey11**)	Down (CMMS24†), insignificantly down (MR)	Down (CMMS24†), insignificantly up (MR)
Nominal wage	W	Up (Ramey11)	Down (CMMS24, as real wage and price both down)	Down (CMMS24, as real wage down and price insignificant)
Dividend/post-tax profits	div			Down (CKS)
Personal income tax rate	PIT	Up (Ramey11)	Up (MR, as shock itself)	Insignificantly down (MR)
Corporate income tax rate	CIT	Up (Ramey16***)	Insignificantly down (MR)	Up (MR, as shock itself)
Tax revenue	TR		Up (CMMS24‡, MR)	Up (CMMS24††), insignificant (MR)
Government spending	G	Up (Ramey16, as shock itself)	Insignificant (MR)	Insignificant (MR)
Government debt	B	Up (CMM)	Down (CMMS24, MR)	Insignificant (CMMS24, MR)
References		Ramey (2016, fig 5) Ramey (2011, fig X) Ramey (2013, figs 1.11-17) Corsetti, Meier and Müller (2012, fig 1)	Mertens and Ravn (2013, figs 2-4/9/10) Cloyne et al. (2023, fig 1) Cloyne et al. (2024, figs 1-2/B.1/H.8) Cloyne, Kurt and Surico (2023, figs 2/3B)	

Notes: This table lists directional impulse responses at about 1-year horizon across variables to shocks. Shocks considered are oil shocks, contractionary monetary policy (MP) shocks, transfer payment (TP) shocks, government spending (G) shocks, and positive shocks in personal income tax (PIT) rate and corporate income tax (CIT) rate. As we are only interested in the directions rather than the magnitudes of responses, we only impose a weak assumption that each variable responds to positive and negative shocks with opposite signs. This is weaker than assuming that the multipliers are the same for positive and negative shocks. Nonetheless, it is worth noting that while some earlier papers advocate for asymmetric multipliers, more recent papers argue that the evidence is weak (e.g., Kilian and Vigfusson (2011) for oil shocks and Ben Zeev, Ramey and Zubairy (2023) for government spending shocks). The abbreviation in parentheses indicates the main reference, usually the most recent or most cited paper. All references are listed at the bottom of each column.

‡ For the transfer payment shock, Pennings (2021) and Mendes et al. (2024) provide cross-sectional estimates instead of aggregate ones, in the US and Brazil respectively. Egger et al. (2022) study transfers in Kenya, but these transfers are funded from outside the economy. We drop this shock in a robustness version in Table B3.

* Ramey (2013) shows that the unemployment rate falls in response to government spending shocks. It is mainly driven by government employment, with the response of private employment either insignificant or negative in different specifications. We drop this variable in a robustness version in Table B3, since we elicit participants' opinions about private businesses.

** Ramey (2011, 2016) discusses extensively potential issues with previous works (e.g., Blanchard and Perotti, 2002; Galí, López-Salido and Vallés, 2007) that find a positive consumption response to government spending shocks. We drop this variable in a robustness version in Table B3.

*** Ramey (2016, fig 5) suggests that the average tax rate goes up in response to government spending shocks, calculated as federal current receipts divided by nominal GDP.

† Cloyne et al. (2024, fig 2) show positive consumption responses to decreases in PIT and CIT, but do not split into durables vs. nondurables.

†† Cloyne et al. (2024, fig H.8) show negative primary surplus response (tax revenue minus government spending) to decreases in PIT and CIT.

Table B3: Correct directional responses with robustness versions

	Oil ↑			MP ↑			G ↑			PIT ↑			CIT ↑			TP ↑		
Version	Va	Vb	Vc	Va	Vb	Vc	Va	Vb	Vc	Va	Vb	Vc	Va	Vb	Vc	Va	Vb	Vc
Y	↓	↓	↓	↓	↓	↓	↑	↑	↑	↓	↓	↓	↓	↓	↓	↑	↑	↑
i	↑	↑	↑	↑	↑	↑			↑			↓			↓			↑
p	↑	↑	↑	↓		↓	↑		↑	↓		↓			↓			↑
u	↑	↑	↑	↑	↑	↑	↓		↓	↑	↑	↑	↑		↑	↓	↓	↓
N	↓	↓	↓	↓	↓	↓	↑	↑	↑	↓	↓	↓	↓		↓	↑	↑	↑
I	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓			
D	↓	↓	↓	↓	↓	↓	↓		↓	↓	↓	↓	↓		↓	↑		↑
ND	↓	↓	↓	↓	↓	↓	↓		↓	↓		↓	↓		↓	↑		↑
W	↑	↑	↑	↓	↓	↓	↑	↑	↑	↓	↓	↓	↓	↓	↓	↑		↑
div			↓			↓			↑			↓	↓	↓	↓	↑		↑
mc			↑			↓			↑			↓			↓			↑
PIT							↑	↑	↑	↑	↑	↑						↑
CIT							↑		↑				↑	↑	↑			↑
TR									↑	↑	↑	↑	↑		↑			↑
B							↑	↑	↑	↓	↓	↓	↓		↓			↑

Notes: Three versions are constructed as follows. Version a is the baseline specification in Table 1.

1. Version a is most closely based on most up-to-date empirical estimates with clear directions.

2. Version b drops from Version a variables for which estimates are noisy or controversy exists. A few noteworthy choices are price response to monetary shocks (classic price puzzle), unemployment and consumption responses to government spending shocks (as discussed in the notes of Table B2).

3. Version c makes additional predictions based on theoretical predictions, relative to Version a.

Table B4: Model-implied variable distance with robustness versions

	Oil			MP			G			PIT			CIT			TP		
Model	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
Y	3	3	3	2	2	2	1	1	1	2	2	2	3	3	3	2	2	2
i	2	2	2	1	1	1	5	3	3	5	3	4	6	4	3	5	4	3
p	1	1	1	4	3	4	4	2	2	4	3	3	5	4	2	4	3	2
u	3	3	3	2	2	2	2	2	2	2	2	2	3	3	2	2	3	2
N	3	3	3	2	2	2	2	2	2	2	2	2	3	3	2	2	3	2
I			3			2			2			3			2			2
D	3	3	3	2	2	2	2	2	2	2	2	2	3	3	3	2	2	2
ND	3	3	3	2	2	2	2	2	2	2	2	2	3	3	3	2	2	2
W			2			3			3			3			3			3
div	4	4	3	3	3	3	2	2	2	3	3	3	2	2	2	2	2	2
mc	1	1	1	3	3	3	3	2	2	3	3	3	4	4	2	3	3	2
PIT							1	1	1	1	1	1				1	1	1
CIT							1	1	1				1	1	1	1	1	1
B							1	1	1	1	1	1	1	1	1	1	1	1

Notes: Three sets of variable distance D_{vs} are constructed as follows, with the oil price shock interpreted as a cost-push shock, as is standard in the literature. Model 2 is the baseline specification in Table 1.

1. Model 1 is the textbook New Keynesian model in the main text of this paper.

2. Model 2 extends Model 1 to feature decreasing-returns production (so that the marginal cost is increasing in quantity) and a Taylor rule of monetary policy that depends on both inflation and unemployment.

3. Model 3 extends Model 1 with capital investment by firms, price and wage rigidity (via a labor union instead of a competitive labor market).

Our directed graph representation of mental models (Figure 1) only includes labor demand and supply shifts $\hat{n}_1^d, \hat{n}_1^s$ and assumes that the real wage w_1 requires one additional step beyond demand and supply shifts. For labor hours, we assume that people perceive the change as soon as they perceive a change in either $\hat{n}_1^d$ or $\hat{n}_1^s$. Alternatively, one could assume that labor hours require an additional step beyond these shifts to be recognized as the equilibrium outcome. This difference is absorbed by variable fixed effects in our regression of correct directional belief on variable distance, and is irrelevant when applying shallow thinking in macroeconomic models, since only beliefs about the real wage are decision-relevant, while those about labor hours are not.

B.2 Additional Results of Predicting Correct Directional Belief

Table B5: Predicting correct directional belief 1_{nvs} on subsets of variables and shocks

	Main		Oil & MP		Decision-relevant		Decision-irrelevant	
Correct directional belief 1_{nvs}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable distance D_{vs}	-0.24***	-0.27***	-0.23***	-0.25***	-0.37***	-0.33***	-0.12***	-0.32***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.07)
Constant	0.99***	1.04***	0.99***	1.03***	1.24***	1.18***	0.71***	1.17***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.16)
Observations	10920	10763	7775	7775	1560	1560	4137	3902
R^2	0.24	0.63	0.26	0.66	0.56	0.85	0.24	0.63
Individual FE	Yes	Absorbed	Yes	Absorbed	Yes	Absorbed	Yes	Absorbed
Individual-variable FE		Yes		Yes		Yes		Yes
Individual-shock FE		Yes		Yes		Yes		Yes

Standard errors in parentheses
Standard errors clustered at individual level
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Columns (1, 2) present the specification for the main sample that filters attention and completion time. Columns (3, 4) focus on respondents who receive the oil and monetary policy (MP) shocks. Columns (5, 6) include only responses to the household block in the survey, which covers four decision-relevant variables: price, wage, interest rate, and personal income tax. Columns (7, 8) include only responses outside the household block and exclude the four aforementioned decision-relevant variables.

Table B6: Predicting correct directional belief 1_{nvs} with robustness versions

	Main	Full	D_{vs}^{M1}	D_{vs}^{M3}	1_{nvs}^{Vb}	1_{nvs}^{Vc}
Correct directional belief 1_{nvs}	(1)	(2)	(3)	(4)	(5)	(6)
Variable distance D_{vs}	-0.27***	-0.23***	-0.19***	-0.19***	-0.20***	-0.26***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.04***	0.91***	0.89***	0.87***	0.93***	1.06***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Observations	10763	22023	10763	12479	8335	14357
R^2	0.63	0.58	0.61	0.62	0.25	0.60
Individual FE	Absorbed	Absorbed	Absorbed	Absorbed	Yes	Absorbed
Individual-variable FE	Yes	Yes	Yes	Yes		Yes
Individual-shock FE	Yes	Yes	Yes	Yes		Yes

Standard errors in parentheses
Standard errors clustered at individual level
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Column (1) presents the main specification with individual-variable and individual-shock fixed effects on the main sample that filters attention and completion time. Column (2) examines the full sample, i.e., all respondents who completed the survey. Columns (3, 4) use the distance implied by Model 1 and Model 3 in Table B4, instead of Model 2. Columns (5, 6) use the shock-variable combinations from Version b and Version c in Table B3, instead of Version a. In the case of Version b, which includes fewer shock-variable combinations, there is not enough variation to apply individual-variable and individual-shock fixed effects.

B.3 Depth of Thinking is Individual Characteristic and Domain-Specific

We further show that the ability to understand shock propagation is indeed an individual characteristic and does not correlate with a classic measure of level-k thinking.

The previous finding that individual fixed effects matter for correct directional beliefs 1_{nvs} already suggests that some people get more variables correct than others. To further investigate this, we measure individual n's overall understanding of shock s by a *total depth score (TDS)* as

$$TDS_{ns} \equiv \sum_{v} D_{vs} \cdot 1_{nvs} \tag{B1}$$

To receive a higher TDS, a respondent needs to correctly forecast directional changes in more variables and especially more distant variables.[37]

To the extent that depth of thinking is an individual characteristic as we postulate, we expect each respondent's TDSs to correlate strongly across shocks. To test this, we rank the TDSs from the lowest to the highest for each shock, and correlate the two TDS rank measures across individuals in Table B7. Column (1) confirms this prediction.

In contrast, column (2) suggests that TDS does not correlate with a classic measure of strategic sophistication (level k), via a "guess 2/3 of the average" game we play with survey respondents. This connects to findings in the macroeconomic literature that the measured level k does not predict differential consumption response to inflation news by Dutch households (Coibion et al., 2023*a*) or first- and higher-order inflation expectations of New Zealand firm managers (Coibion et al., 2021).[38] We remark that shallow thinking likely reflects people's limited knowledge about the macroeconomy, a distinct aspect of bounded rationality from limited strategic sophistication. After all, a chess master who could anticipate opponents well may not know macroeconomics, and vice versa.

Column (3) suggests that a measure of financial literacy, based on Lusardi and Mitchell (2011), but not general education, correlates with TDS. This supports the idea that shallow thinking reflects individuals' economic knowledge, for which general education may be

[37] This TDS is a more robust measure to noise than the maximum distance of variables whose changes are understood correctly, as we have several variables for each step and respondents may coincidentally get some correct.

[38] In the experimental literature, Dal Bó, Dal Bó and Eyster (2018) show that voters prefer policy changes that bring in direct benefits but induce larger indirect costs, but their voting behavior is not correlated with level k. Georganas, Healy and Weber (2015) study stability of level k using two families of games: beauty contest games à la Nagel (1995) and undercutting games similar to Arad and Rubinstein (2012). They find that the participants' levels are consistent within the beauty contest family, but do not correlate within the undercutting game family or across two families.

Table B7: Total depth score as individual characteristic

			Main			Incentivized	Full
TDS of 2nd shock (rank)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TDS of 1st shock (rank)	0.31***				0.24***	0.30**	0.20***
	(0.05)				(0.05)	(0.11)	(0.04)
Level *k* (rank)		-0.08			-0.04	0.05	-0.02
		(0.05)			(0.05)	(0.11)	(0.03)
Financial literacy (rank)			0.28***		0.23***	0.11	0.20***
			(0.05)		(0.05)	(0.11)	(0.03)
Education (rank)			0.02		-0.02	-0.04	-0.04
			(0.05)		(0.06)	(0.12)	(0.04)
Male			-0.03		-0.03	-0.10	-0.01
			(0.03)		(0.03)	(0.06)	(0.02)
Net asset (rank)				0.05	0.05	0.04	0.08*
				(0.06)	(0.06)	(0.11)	(0.04)
Income (rank)				0.06	0.06	0.11	0.03
				(0.06)	(0.07)	(0.13)	(0.05)
Constant	0.38***	0.59***	0.37***	0.49***	0.25***	0.29*	0.26***
	(0.03)	(0.03)	(0.05)	(0.04)	(0.06)	(0.14)	(0.04)
Observations	383	383	383	383	383	118	828
R^2	0.09	0.01	0.09	0.02	0.15	0.17	0.12
Shock group FE					Yes	Yes	Yes
Age group FE					Yes	Yes	Yes

Standard errors in parentheses
Standard errors are heteroscedasticity-consistent
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Columns (1-5) use the main sample that filters attention and completion time. Column (6) focuses on a subsample of the main sample, where the "guess 2/3 of the average" game is incentivized. Column (7) studies the full sample, i.e., all respondents who completed the survey.

too noisy a proxy.

Column (4) shows that households' net asset positions or income do not significantly correlate with TDS. The two significant predictors, TDS and financial literacy, remain significant when all variables are pooled together, as indicated in column (5).

Columns (6) and (7) demonstrate that our findings hold true over a subsample where the "guess 2/3 of the average" game is incentivized, as well as in the full sample, i.e., all respondents who completed the survey.

C Appendix to the New Keynesian Model

C.1 The Infinite-Horizon New Keynesian Model

Firms. There is a continuum of firms indexed by $j \in [0, 1]$ in this economy subject to Calvo price rigidity. Each firm chooses labor demand N^s_{jt}, pays dividend DIV_{jt}, and sets price P^*_{jt} when it can, taking as given the aggregate inflation rate π_t, the real wage W_t, and the aggregate demand C_t. They agree on the steady state of the economy but may have heterogeneous beliefs about the economy's response to shocks.

Each firm produces a differentiated good, using the same constant-returns production technology using labor hours $Y_{jt} = N^d_{jt}$, which together forms a bundle with constant elasticity ε of substitution (CES) that the households consume. At the steady state, they each charge a markup $\mu \equiv \frac{\varepsilon}{\varepsilon-1}$. The log-linearized real dividend and aggregate labor demand are

$$div_t = c_t - \frac{1}{\mu - 1} w_t \tag{C1}$$

$$n^d_t = c_t \tag{C2}$$

since the price dispersion only introduces second-order changes as in Galí (2015).

Each firm resets its price with independent probability $1 - \theta$ in any period and fulfills its demand period by period. When considering its reset price P^*_{jt}, each firm maximizes its discounted sum of profits

$$\max_{P^*_{jt}} \sum_{k=0}^{\infty} \theta^k \mathbb{E}_{jt} \left[\beta^k \frac{C^{-\sigma}_{t+k}}{C^{-\sigma}_t} \frac{P^*_{jt} - W_{t+k} P_{t+k}}{P_{t+k}} Y_{j,t+k|t} \right]$$

where $\beta^k \frac{C^{-\sigma}_{t+k}}{C^{-\sigma}_t}$ is the discount factor and $W_{t+k}P_{t+k}$ is the nominal marginal cost, subject to the sequence of demand constraints

$$Y_{j,t+k|t} = \left(\frac{P^*_{jt}}{P_{t+k}} \right)^{-\varepsilon} Y_{t+k}$$

The first-order condition is

$$0 = \sum_{k=0}^{\infty} \theta^k \mathbb{E}_{jt} \left[\beta^k \frac{C^{-\sigma}_{t+k}}{C^{-\sigma}_t} \frac{(1 - \varepsilon) P^*_{jt} + \varepsilon W_{t+k} P_{t+k}}{P_{t+k}} Y_{j,t+k|t} \right]$$

and thus

$$p_{jt}^{*} = p_{t-1} + (1-\beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^{k} \mathbb{E}_{jt} \left[\sum_{l=0}^{k} \pi_{t+l} + w_{t+k} \right]$$

Aggregate inflation emerges from the pricing behavior of the $(1-\theta)$ share of resetting firms as $\pi_t = (1-\theta)\left(p_{jt}^{*} - p_{t-1}\right)$. Following the tradition, we consider a cost-push shock ϵ_t^{π} for inflation

$$\pi_t = \theta\kappa w_t + (1-\theta)\pi_t + \theta\kappa \sum_{k=1}^{\infty} (\beta\theta)^{k} \overline{\mathbb{E}}_t [w_{t+k}] + (1-\theta) \sum_{k=1}^{\infty} (\beta\theta)^{k} \overline{\mathbb{E}}_t [\pi_{t+k}] + \epsilon_t^{\pi} \tag{C3}$$

with $\kappa \equiv \frac{(1-\theta)(1-\beta\theta)}{\theta}$ capturing the slope of the Phillips curve and $\overline{\mathbb{E}}_t[\cdot]$ being the average expectations. Importantly, we do not move π_t on the right-hand side to the left. We intentionally preserve the dependence of π_t on itself, which encapsulates the within-period complementarity in individual price setting as each firm takes aggregate inflation as given.

Households. There is a continuum of households who live infinitely indexed by $h \in [0,1]$. Each household chooses consumption C_{ht} and labor supply N_{ht}^{s}, taking as given the gross nominal interest rates R_{t-1}, the inflation rate $\pi_t \equiv P_t/P_{t-1} - 1$, the real wage W_t, and the real dividend DIV_t. They agree on the steady state of the economy but may have heterogeneous beliefs about the economy's response to shocks.

They each maximize

$$\max_{\left\{C_{ht}, N_{ht}^{s}\right\}_{t\geq 0}} \mathbb{E}_{h,t=0} \sum_{t=0}^{\infty} \beta^{t} \left(\frac{C_{ht}^{1-\sigma} - 1}{1-\sigma} - \frac{\left(N_{ht}^{s}\right)^{1+\nu}}{1+\nu} \right)$$

subject to the budget constraint

$$C_{ht} + A_{ht} = \frac{R_{t-1}}{1+\pi_t} A_{h,t-1} + W_t N_{ht}^{s} + DIV_t$$

where C_{ht} is a CES bundle of goods in the economy, A_{ht} is the period-t saving.

We derive the log-linearized aggregate consumption and labor supply functions as follows. The log-linearized life-time budget constraint is

$$\mathbb{E}_{ht}\sum_{k=0}^{\infty}\beta^{k}\left(c_{h,t+k}-\underbrace{\frac{\overline{WN}}{\overline{C}}}_{\mu^{-1}}\left(w_{t+k}+n^{s}_{h,t+k}\right)-\underbrace{\frac{\overline{DIV}}{\overline{C}}}_{1-\mu^{-1}}div_{t+k}\right)-\beta^{-1}a_{h,t-1}=0$$

where the lower-case variables denote the log deviation from the corresponding steady-state values, and $\mu=\frac{\overline{C}}{\overline{WN}}$ denotes the ratio of consumption to labor income at the steady state (which equals firms' steady state markup μ).

Log-linearizing the consumption-labor FOC $W_tC_{ht}^{-\sigma}=\left(N^{s}_{ht}\right)^{\nu}$ yields

$$n^{s}_{ht}=\frac{1}{\nu}\left(w_t-\sigma c_{ht}\right)$$

Plugging this into the budget constraint gives

$$\mathbb{E}_{ht}\sum_{k=0}^{\infty}\beta^{k}\left(\left(1+\mu^{-1}\sigma\nu^{-1}\right)c_{h,t+k}-\mu^{-1}\left(1+\nu^{-1}\right)w_{t+k}-\left(1-\mu^{-1}\right)div_{t+k}\right)-\beta^{-1}a_{h,t-1}=0$$

$$\mathbb{E}_{ht}\sum_{k=0}^{\infty}\beta^{k}\left(c_{h,t+k}-\frac{\mu^{-1}\left(1+\nu\right)}{\mu^{-1}\sigma+\nu}w_{t+k}-\frac{\left(1-\mu^{-1}\right)\nu}{\mu^{-1}\sigma+\nu}div_{t+k}\right)-\frac{\beta^{-1}\nu}{\mu^{-1}\sigma+\nu}a_{h,t-1}=0$$

Log-linearizing the Euler condition $C_{ht}^{-\sigma}=\mathbb{E}_{ht}\left[\frac{\beta R_t}{1+\pi_{t+1}}C_{h,t+1}^{-\sigma}\right]$ gives

$$c_{ht}=\mathbb{E}_{ht}\left(c_{h,t+1}-\sigma^{-1}\left(i_t-\pi_{t+1}\right)\right)$$

Combining this with the budget constraint to substitute $c_{h,t+k}$ gives rise to

$$\mathbb{E}_{ht}\sum_{k=0}^{\infty}\beta^{k}\left(c_{ht}+\sum_{l=0}^{k-1}\sigma^{-1}\left(i_{t+l}-\pi_{t+l+1}\right)-\frac{\mu^{-1}\left(1+\nu\right)}{\mu^{-1}\sigma+\nu}w_{t+k}-\frac{\left(1-\mu^{-1}\right)\nu}{\mu^{-1}\sigma+\nu}div_{t+k}\right)-\frac{\beta^{-1}\nu}{\mu^{-1}\sigma+\nu}a_{h,t-1}=0$$

$$\frac{1}{1-\beta}c_{ht}-\sum_{k=0}^{\infty}\beta^{k}\mathbb{E}_{ht}\left(\frac{(1+\nu)}{\sigma+\mu\nu}w_{t+k}+\frac{(\mu-1)\,\nu}{\sigma+\mu\nu}div_{t+k}\right)+\frac{\sigma^{-1}\beta}{1-\beta}\sum_{k=0}^{\infty}\beta^{k}\mathbb{E}_{ht}\left(i_{t+k}-\pi_{t+k+1}\right)-\frac{\beta^{-1}\mu\nu}{\sigma+\mu\nu}a_{h,t-1}=0$$

which leads to the aggregate consumption function, once aggregated across all households

$$c_t = -\sigma^{-1}\beta i_t - \sigma^{-1}\beta \sum_{k=1}^{\infty} \beta^k \overline{\mathbb{E}}_t \left[i_{t+k}\right] + \sigma^{-1} \sum_{k=1}^{\infty} \beta^k \overline{\mathbb{E}}_t \left[\pi_{t+k}\right] + (1-\beta)\left[\frac{(\mu-1)\nu}{\sigma+\mu\nu} div_t + \frac{(1+\nu)}{\sigma+\mu\nu} w_t\right]$$
$$+ (1-\beta)\sum_{k=1}^{\infty} \beta^k \overline{\mathbb{E}}_t \Big[\frac{(\mu-1)\nu}{\sigma+\mu\nu} div_{t+k} + \frac{(1+\nu)}{\sigma+\mu\nu} w_{t+k}\Big] \tag{C4}$$

Using the consumption-labor FOC $n^s_{ht} = \frac{1}{\nu}(w_t - \sigma c_{ht})$ again, we get the aggregate labor supply function

$$n^s_t = \nu^{-1}\beta i_t + \nu^{-1}\beta \sum_{k=1}^{\infty} \beta^k \overline{\mathbb{E}}_t \left[i_{t+k}\right] - \nu^{-1} \sum_{k=1}^{\infty} \beta^k \overline{\mathbb{E}}_t \left[\pi_{t+k}\right] - (1-\beta)\frac{(\mu-1)\sigma}{\sigma+\mu\nu} div_t + \nu^{-1}\left(1 - \sigma\frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu}\right) w_t$$
$$- (1-\beta)\sum_{k=1}^{\infty} \beta^k \overline{\mathbb{E}}_t \Big[\frac{(\mu-1)\sigma}{\sigma+\mu\nu} div_{t+k} + \frac{(1+\nu)\sigma\nu^{-1}}{\sigma+\mu\nu} w_{t+k}\Big] \tag{C5}$$

where $\overline{\mathbb{E}}_t[\cdot]$ denotes the average expectations.

Central bank. The central bank follows a Taylor rule with a monetary policy shock ϵ^i_t,

$$i_t = \phi\pi_t + \epsilon^i_t \tag{C6}$$

Labor market. Last, to close the model, the wage arises by equilibrating labor supply and demand

$$n^s_t = n^d_t \tag{C7}$$

Equilibrium. We study a *temporary equilibrium* in which agents maximize their utilities, taking as given the average expectations $\overline{\mathbb{E}}_t[\cdot]$, and markets clear. In this New Keynesian model, equations (C1-C7) characterizes the equilibrium given the average expectations.

C.2 Proofs of Propositions

Proof of Proposition 1. It follows directly from Section 2.1. We explicitly write down the system as

$$\underbrace{\begin{pmatrix} i_1 \\ \pi_1 \\ div_1 \\ \hat{n}_1^d \\ c_1 \\ \hat{n}_1^s \\ w_1 \end{pmatrix}}_{V_1} = \underbrace{\begin{pmatrix} 0 & \phi & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\theta & 0 & 0 & 0 & 0 & \theta\kappa \\ 0 & 0 & 0 & 0 & 1 & 0 & -\frac{1}{\mu-1} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -\sigma^{-1}\beta & 0 & \frac{(1-\beta)(\mu-1)\nu}{\sigma+\mu\nu} & 0 & 0 & 0 & \frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu} \\ \nu^{-1}\beta & 0 & -\frac{(1-\beta)(\mu-1)\sigma}{\sigma+\mu\nu} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e_{n^s w}^{-1} & 0 & -e_{n^s w}^{-1} & 0 \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} i_1 \\ \pi_1 \\ div_1 \\ \hat{n}_1^d \\ c_1 \\ \hat{n}_1^s \\ w_1 \end{pmatrix}}_{V_1} + \underbrace{\begin{pmatrix} \epsilon_1^i \\ \epsilon_1^\pi \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{S_1} \tag{C8}$$

with $e_{n^s w} = \nu^{-1}\left[1 - \sigma\frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu}\right]$.□

Proof of Proposition 2. Assumptions 1 and 2 imply that the average beliefs are

$$\begin{aligned} \overline{\mathbb{E}}_0\left[V_1\right] &\equiv \sum_{d=1}^{\infty} \mathbb{P}\left(d=n\right) \mathbb{E}_0^d\left[V_1\right] \\ &= \sum_{d=1}^{\infty} \mathbb{P}\left(d=n\right) \sum_{n=1}^{d} M^{n-1} S_1 = \sum_{n=1}^{\infty} \mathbb{P}\left(d \geq n\right) M^{n-1} S_1 = \sum_{n=1}^{\infty} \lambda^{n-1} M^{n-1} S_1 \end{aligned}$$

which can be recast as (15).□

Proof of Proposition 3. (16) follows from Assumptions 2 and 3 and Definition 2. Since the variable distance D_{vs} is bounded above, an ordinary least squares estimation identifies a negative slope γ. Further, the conditional expectation (16) minimizes the mean squared error of $\mathbb{E}^{population}\left(1_{nvs} - h\left(D_{vs}\right)\right)$ among all possible predictor $h\left(D_{vs}\right)$. Since it is a case of the exponential family $b_1 \cdot b_2^{D_{vs}-1} + b_0$, we can exactly identify λ with b_2.□

Proofs of Propositions 4 and 5. Directly solving (1-7) in response to the cost-push shock ϵ_1^π and the monetary policy shock ϵ_1^i yields the equilibrium. Solving the fixed point of (1-7) with cross-variable relations dampened by λ as in (15) gives the average expectations.

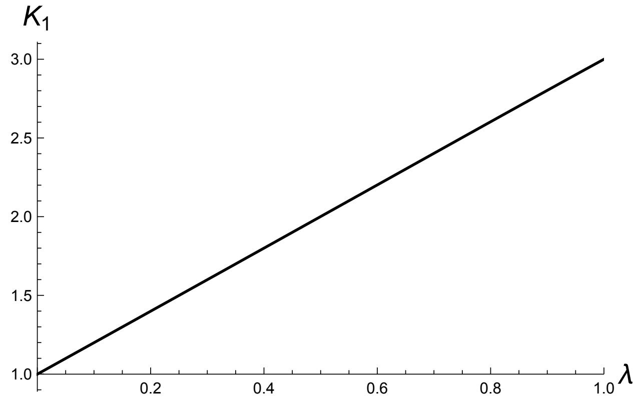


Figure C1: $K_1(\lambda)$ under our calibration

The dampening of Keynesian cross is

$$K_1(\lambda) = \frac{\dfrac{\beta(\sigma+\lambda\nu)}{\sigma\left(1-\frac{(1-\beta)(1+\nu)\sigma}{(\mu\nu+\sigma)}\right)}}{1-\lambda^2\dfrac{(1-\beta)(\mu-1)\nu}{(\mu\nu+\sigma)}-\lambda^3\dfrac{(1-\beta)\nu\left(1+\nu-\sigma-\lambda\nu+\lambda(\mu-1)\frac{(1-\beta)(1+\nu)\sigma}{(\mu\nu+\sigma)}\right)}{(\mu\nu+\sigma)-(1-\beta)(1+\nu)\sigma}} \tag{C9}$$

which depends on β, μ, σ, ν but no other parameters when $\lambda \in [0,1)$ and obtains $\sigma^{-1}(\nu+\sigma)$ at $\lambda = 1$. Under our calibration, $K_1(\lambda)$ increases in λ as illustrated in Figure C1.□

Proof of Proposition 6. It follows direction from Section 5.1 and Appendix C.1.□

Proof of Proposition 7. The proof is the same as that of Proposition 2.□

Proof of Propositions 8 and C1. Though the proposition only concerns a cost-push shock ϵ_t^π, we also consider a monetary policy shock ϵ_t^i in the proof, which does not add much complexity. This proof will nest the Proofs of Proposition 4 and 5 by setting $\rho = 0$.

We solve the average expectations using (42), which will nest the rational expectations equilibrium with $\lambda = 1$. We guess and verify that beliefs mean-revert at the same rate ρ.

Agents' actions (C1-C6) characterize $\left\{\overline{\mathbb{E}}[div_t], \overline{\mathbb{E}}\left[\hat{n}_t^d\right], \overline{\mathbb{E}}[\pi_t], \overline{\mathbb{E}}[c_t], \overline{\mathbb{E}}\left[\hat{n}_t^s\right], \overline{\mathbb{E}}[i_t]\right\}$ as

$$\overline{\mathbb{E}}[div_t] = \lambda\overline{\mathbb{E}}[c_t] - \lambda\frac{1}{\mu-1}\overline{\mathbb{E}}[w_t] \tag{C10}$$

$$\overline{\mathbb{E}}\left[\hat{n}_t^d\right] = \lambda\overline{\mathbb{E}}[c_t] \tag{C11}$$

$$\overline{\mathbb{E}}[\pi_t] = \lambda\theta\kappa\frac{\overline{\mathbb{E}}[w_t]}{1-\beta\theta\rho} + \lambda(1-\theta)\frac{\overline{\mathbb{E}}[\pi_t]}{1-\beta\theta\rho} + \epsilon_t^{\pi} \tag{C12}$$

$$\overline{\mathbb{E}}[c_t] = -\lambda\sigma^{-1}\beta\left(\frac{\overline{\mathbb{E}}[i_t]}{1-\beta\rho} - \frac{\rho\overline{\mathbb{E}}[\pi_t]}{1-\beta\rho}\right) + \lambda(1-\beta)\frac{(\mu-1)\nu}{\sigma+\mu\nu}\frac{\overline{\mathbb{E}}[div_t]}{1-\beta\rho} + \lambda(1-\beta)\frac{(1+\nu)}{\sigma+\mu\nu}\frac{\overline{\mathbb{E}}[w_t]}{1-\beta\rho} \tag{C13}$$

$$\overline{\mathbb{E}}[\hat{n}_t^s] = \lambda\nu^{-1}\beta\left(\frac{\overline{\mathbb{E}}[i_t]}{1-\beta\rho} - \frac{\rho\overline{\mathbb{E}}[\pi_t]}{1-\beta\rho}\right) - \lambda(1-\beta)\frac{(\mu-1)\sigma}{\sigma+\mu\nu}\frac{\overline{\mathbb{E}}[div_t]}{1-\beta\rho} \tag{C14}$$

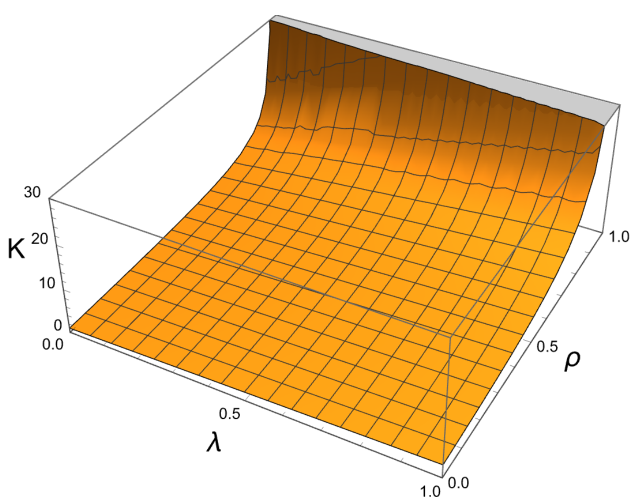
$$\overline{\mathbb{E}}[i_t] = \lambda\phi\overline{\mathbb{E}}[\pi_t] + \epsilon_t^i \tag{C15}$$

Note that in constructing $\overline{\mathbb{E}}\left[\hat{n}_t^d\right]$ and $\overline{\mathbb{E}}\left[\hat{n}_t^s\right]$, we take away their dependence on the wage w from n^d and n^s in (C2, C5).

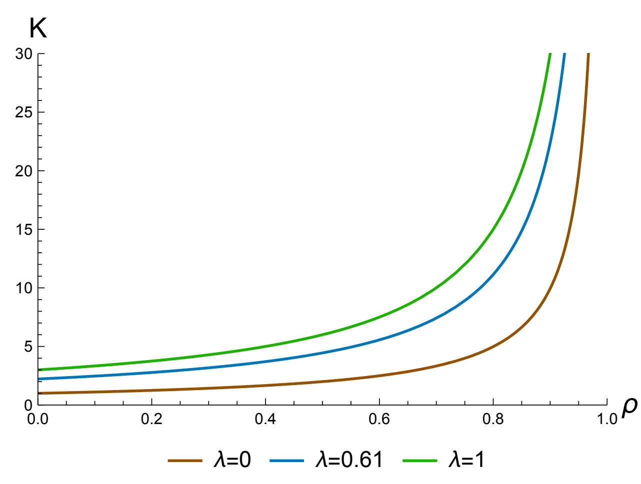
Last, the labor market clearing condition (C7) combined with the dependence of n^d and n^s on w from (C2, C5) determines $\overline{\mathbb{E}}[w_t]$

$$\nu^{-1}\left(1 - \sigma\frac{(1-\beta)(1+\nu)}{(1-\beta\rho)(\sigma+\mu\nu)}\right)\overline{\mathbb{E}}[w_t] = \lambda\overline{\mathbb{E}}\left[\hat{n}_t^d\right] - \lambda\overline{\mathbb{E}}\left[\hat{n}_t^s\right] \tag{C16}$$

Solving (C10-C16) gives the average expectations (45) and (46), which nest the rational expectations equilibrium (43) and (44) with $\lambda = 1$.



(a) $K(\lambda, \rho)$



(b) $K(\lambda, \rho)$, varying ρ under fixed λ

Figure C2: $K(\lambda, \rho)$ under our calibration

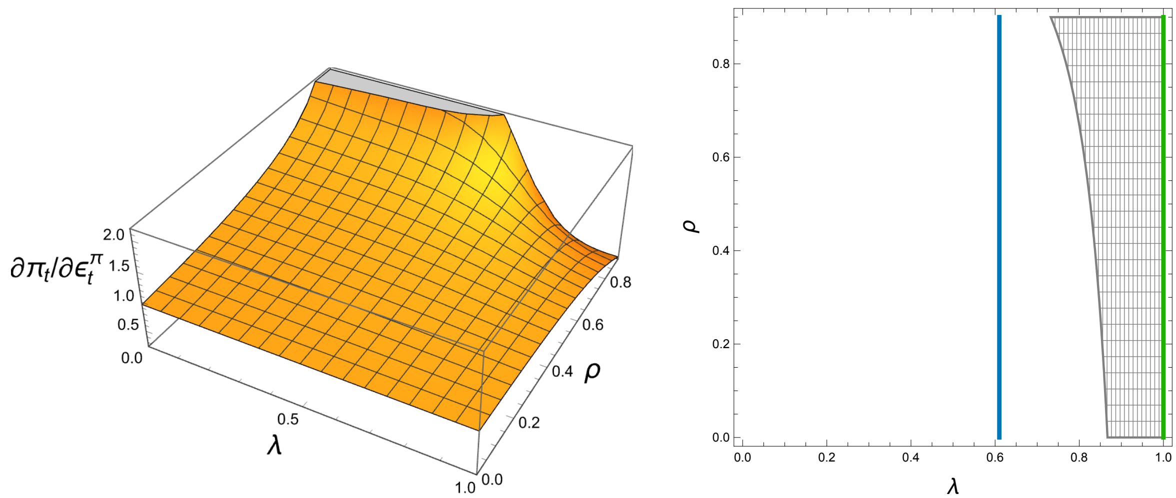
The dampening of Keynesian cross under persistent shocks is

$$K(\lambda,\rho)=\frac{\frac{\beta(\sigma+\lambda\nu)}{\sigma(1-\beta\rho)\left(1-\frac{(1-\beta)(1+\nu)\sigma}{(1-\beta\rho)(\mu\nu+\sigma)}\right)}}{1-\lambda^2\frac{(1-\beta)(\mu-1)\nu}{(1-\beta\rho)(\mu\nu+\sigma)}-\lambda^3\frac{(1-\beta)\nu\left(1+\nu-\sigma-\lambda\nu+\lambda(\mu-1)\frac{(1-\beta)(1+\nu)\sigma}{(1-\beta\rho)(\mu\nu+\sigma)}\right)}{(1-\beta\rho)(\mu\nu+\sigma)-(1-\beta)(1+\nu)\sigma}} \tag{C17}$$

which depends on $\rho,\beta,\mu,\sigma,\nu$ but no other parameters when $\lambda\in[0,1)$ and obtains $\frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}$ at $\lambda=1$. Under our calibration, $K(\lambda,\rho)$ increases both in λ and ρ as shown in Figure C2.□

C.3 Additional Results of Persistent Shocks

The effect of more persistent cost-push shocks on equilibrium inflation. Figure 43 suggests that under our calibrated λ, a more persistent cost-push shock leads to lower inflation, contrary to the rational expectations equilibrium. Figure C3 shows the equilibrium inflation response to a persistent cost-push shock, varying the shallow thinking parameter λ and the persistence of the shock ρ. Panel C3a suggests that the inflation response increases in ρ when λ is low and decreases in ρ when λ is high. Panel C3b shows, in the gray meshed area, the combination of (λ,ρ) values under which a more persistent cost-push shock leads to lower inflation than a purely transitory shock, i.e., $\frac{\partial\pi_t}{\partial\epsilon_t^\pi}<\frac{\partial\pi_t}{\partial\epsilon_t^\pi}|_{\rho=0}$. In the case of purely transitory shock ($\rho=0$), the equilibrium is independent of λ, as agents observe all variables in the same period. The vertical blue line (our calibrated λ) is completely outside the meshed region, whereas the vertical green line (rational expectations equilibrium) is entirely within it.



(a) Equilibrium inflation π_t in response to cost-push shock ϵ_t^π, varying λ and ρ

(b) Combination of (λ, ρ) values under which $\frac{\partial \pi_t}{\partial \epsilon_t^\pi} < \frac{\partial \pi_t}{\partial \epsilon_t^\pi}|_{\rho=0}$

Figure C3: Inflation response to persistent cost-push shock ϵ_t^π

Notes: Panel (a) plots the equilibrium inflation π_t (relative to the size of the shock) in response to a persistent cost-push shock $\epsilon_t^\pi = \rho^t \epsilon^\pi$, as a function of the shallow thinking parameter λ and the shock persistence ρ. Panel (b) shows, in the gray meshed area, the combination of (λ, ρ) values under which a more persistent cost-push shock leads to lower inflation than a purely transitory shock, i.e., $\frac{\partial \pi_t}{\partial \epsilon_t^\pi} < \frac{\partial \pi_t}{\partial \epsilon_t^\pi}|_{\rho=0}$. In the case of purely transitory shock ($\rho = 0$), the equilibrium is independent of λ, as agents observe all variables in the same period. The vertical line in blue indicates our calibration of λ, which is completely outside the gray area. In contrast, the vertical green line representing the rational expectations equilibrium lies entirely within the area.

Persistent monetary policy shocks.

Proposition C1. (Persistent monetary policy shock) *The rational-expectations equilibrium (REE) response to a persistent monetary policy shock $\epsilon_t^i = \rho^t \epsilon^i$ features*

$$i_t^{REE} = \left[1 + \frac{\frac{\phi\theta\kappa}{1-\beta\theta\rho} \frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}}{1 - \frac{(1-\theta)+\rho\theta\kappa\frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}}{1-\beta\theta\rho}} \right]^{-1} \epsilon_t^i \tag{C18}$$

$$\pi_t^{REE} = -\frac{\frac{\theta\kappa}{1-\beta\theta\rho} \frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}}{1 - \frac{(1-\theta)+\rho\theta\kappa\frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}}{1-\beta\theta\rho}} i_t^{REE} \tag{C19}$$

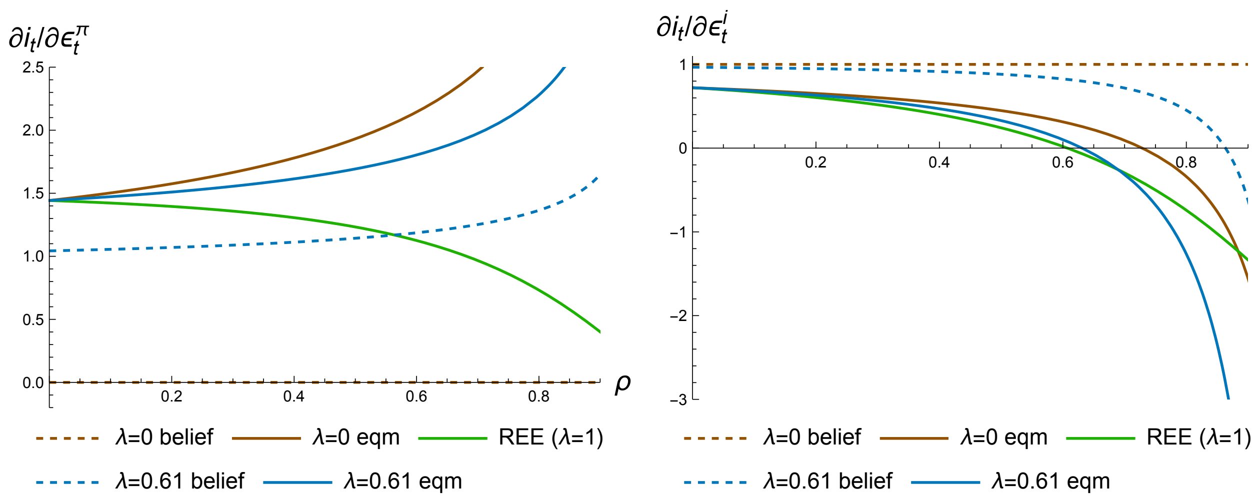
whereas the average expectations under shallow thinking are

$$\overline{\mathbb{E}}\left[i_t\right] = \left[1 + \frac{\frac{\lambda^4\phi\theta\kappa K(\lambda,\rho)}{1-\beta\theta\rho}}{1 - \frac{\lambda(1-\theta)+\lambda^3\rho\theta\kappa K(\lambda,\rho)}{1-\beta\theta\rho}}\right]^{-1} \epsilon_t^i \tag{C20}$$

$$\overline{\mathbb{E}}\left[\pi_t\right] = -\frac{\frac{\lambda^3\theta\kappa K(\lambda,\rho)}{1-\beta\theta\rho}}{1 - \frac{\lambda(1-\theta)+\lambda^3\rho\theta\kappa K(\lambda,\rho)}{1-\beta\theta\rho}}\overline{\mathbb{E}}\left[i_t\right] \tag{C21}$$

with $K(\lambda, \rho)$ increasing in λ and ρ under our calibration and $K(1, \rho) = \frac{\sigma^{-1}(\nu+\sigma)}{1-\rho}$.

This proposition nests Proposition 5 with $\rho = 0$. Figure C4 shows the responses of interest rate expectations and equilibrium interest rate to persistent cost-push shocks and monetary policy shocks.



(a) Persistent cost-push shock $\epsilon_t^\pi = \rho^t \epsilon^\pi$

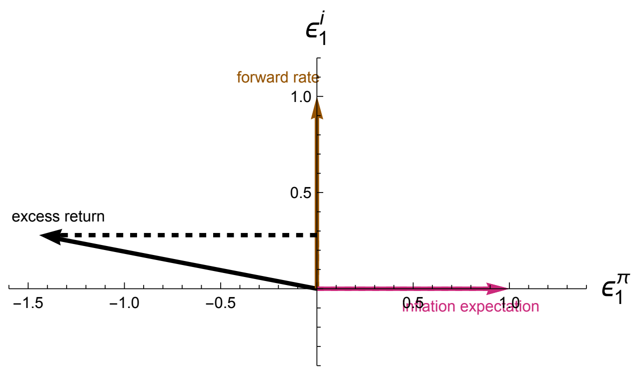
(b) Persistent monetary policy shock $\epsilon_t^i = \rho^t \epsilon^i$

Figure C4: Interest rate i_t in response to persistent shocks

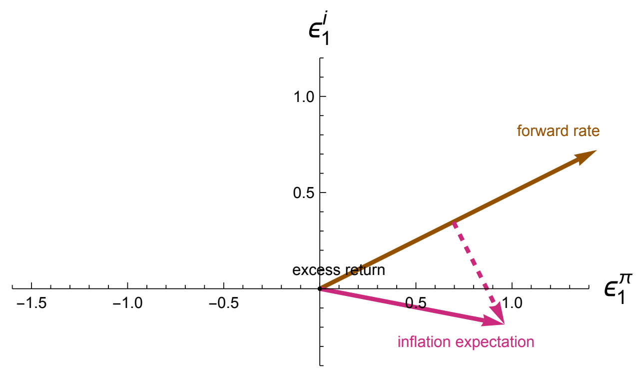
Notes: These panels plot the responses of the interest rate to persistent cost-push shocks ϵ_t^π and monetary policy shocks ϵ_t^i, parallel to Figure 10b. Under our calibration, the average interest rate expectation is higher than its equilibrium value in response to monetary policy shocks, but is lower in response to cost-push shocks, regardless of ρ.

C.4 Additional Results of Bond Returns

Bond excess returns with transitory news shocks. Figure C5 shows the loadings of bond excess returns and predictors on transitory news shocks, under $\lambda = 0, 1$, complementing Figure 8b.



(a) Loadings on news shocks ($\lambda = 0$)



(b) Loadings on news shocks ($\lambda = 1$)

Figure C5: Predictability of bond excess returns, with transitory news shocks (cont'd)

Notes: These panels plot the loadings of bond excess returns and predictors on transitory news shocks, under $\lambda = 0, 1$, complementing Figure 8b. In the case of $\lambda = 0$, the forward rate and inflation expectation point along the y- and x-axes, respectively. In the case of $\lambda = 1$, the excess return is 0.

Sensitivity of long-term rates to short-term rates with persistent shocks. We consider an economy impacted by persistent monetary policy shocks $\epsilon_t^i = \rho^t \epsilon^i$ and cost-push shocks $\epsilon_t^\pi = \rho^t \epsilon^\pi$. Suppose that in each period, the two shocks have variances σ_i^2 and σ_π^2, and are independently distributed.

In this environment, the yield of a T-period bond, $y_t^{(T)}$, is determined by the expectations hypothesis as

$$y_t^{(T)} = \frac{i_t + \sum_{k=1}^{T-1} \overline{\mathbb{E}}\left[i_{t+k}\right]}{T} \tag{C22}$$

We consider a univariate regression of the long-term yield $y_t^{(T)}$ on the short-term interest rate i_t à la Hanson, Lucca and Wright (2021)

$$y_t^{(T)} = \beta^{OLS} i_t + \alpha + \epsilon_0 \tag{C23}$$

Since the economy is affected by two different shocks, the univariate regression coefficient β^{OLS} depends on the mix of these shocks. To shed light on this, we express the responses of interest rate expectation $\overline{\mathbb{E}}[i_{t+k}]$ and equilibrium interest rate i_t to shocks as

$$\begin{aligned}\overline{\mathbb{E}}[i_{t+k}] &= \rho^k \gamma_i^{exp} \epsilon_t^i + \rho^k \gamma_\pi^{exp} \epsilon_t^\pi, \ \forall k \geq 1 \\ i_t &= \gamma_i^{ST} \epsilon_t^i + \gamma_\pi^{ST} \epsilon_t^\pi\end{aligned}$$

where the γ coefficients are implicit functions of model parameters including λ, and ρ^k stems from the fact that beliefs are mean-reverting at the same rate ρ. The γ^{exp} coefficients agree with γ^{ST} coefficients under $\lambda = 1$ (rational expectations). We denote the relative importance of monetary policy shocks in driving the short-term interest rate as $\zeta_i \equiv \frac{\left(\gamma_i^{ST}\right)^2 \sigma_i^2}{\left(\gamma_i^{ST}\right)^2 \sigma_i^2 + \left(\gamma_\pi^{ST}\right)^2 \sigma_\pi^2} \in [0,1]$, and the relative importance of cost-push shocks as $1 - \zeta_i$.

According to (C22), the long-term yield follows

$$y_t^{(T)} = T^{-1}\left(\gamma_i^{ST} + \rho \frac{1-\rho^{T-1}}{1-\rho} \gamma_i^{exp}\right)\epsilon_t^i + T^{-1}\left(\gamma_\pi^{ST} + \rho \frac{1-\rho^{T-1}}{1-\rho} \gamma_\pi^{exp}\right)\epsilon_\pi^i$$

Therefore, the univariate regression coefficient β^{OLS} from (C23) is

$$\beta^{OLS} = \frac{\operatorname{Cov}\left(y_t^{(T)}, i_t\right)}{\operatorname{Var}(i_t)} = \underbrace{T^{-1}\left(1 + \rho \frac{1-\rho^{T-1}}{1-\rho} \frac{\gamma_i^{exp}}{\gamma_i^{ST}}\right)}_{\equiv \beta_i^{IV}} \zeta_i + \underbrace{T^{-1}\left(1 + \rho \frac{1-\rho^{T-1}}{1-\rho} \frac{\gamma_\pi^{exp}}{\gamma_\pi^{ST}}\right)}_{\equiv \beta_\pi^{IV}} (1-\zeta_i) \quad \text{(C24)}$$

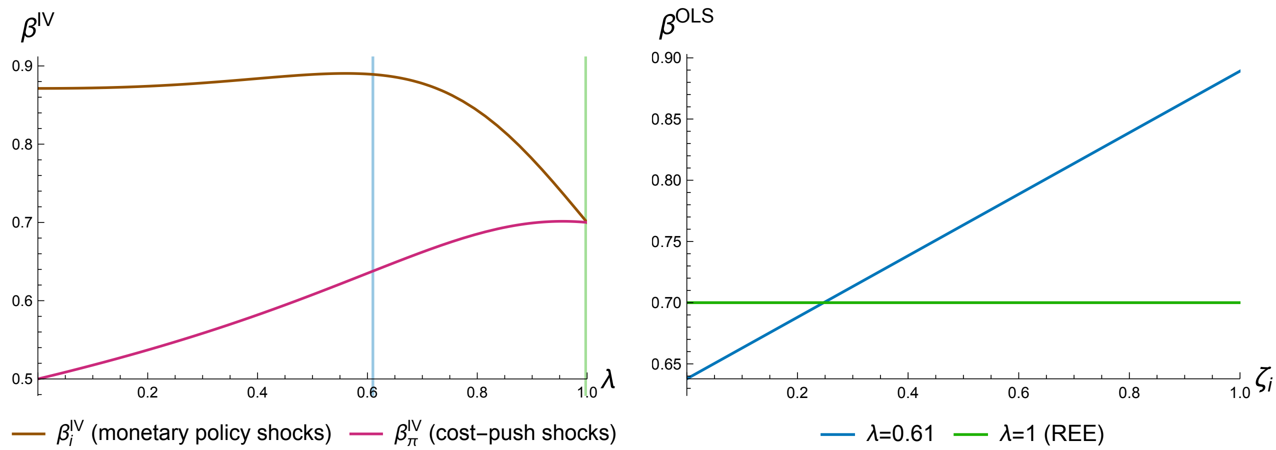
which linearly interpolates between β_i^{IV}, the conditional sensitivity under monetary shocks, and β_π^{IV}, the conditional sensitivity under cost-push shocks, based on the relative importance ζ_i.

As $\gamma_i^{exp} = \gamma_i^{ST}$ and $\gamma_\pi^{exp} = \gamma_\pi^{ST}$ under rational expectations, we have

$$\beta^{OLS,REE} \equiv T^{-1} \frac{1-\rho^T}{1-\rho} \quad \text{(C25)}$$

which is independent of the shock mix ζ_i.

Figure C6 plots these univariate regression coefficients $\beta_i^{IV}, \beta_\pi^{IV}$ and β^{OLS}, for a 2-period bond ($T = 2$) under a mild shock persistence $\rho = 0.4$. Panel C6a suggests that the conditional sensitivity to monetary policy shocks β_i^{IV} (in brown) is higher than the conditional sensitivity to cost-push shocks β_π^{IV} (in purple), as long as $\lambda < 1$. Consequently, under our calibrated λ, the unconditional sensitivity β^{OLS} increases in the relative importance of the monetary policy shocks ζ_i, as shown in panel C6b.



(a) Conditional sensitivities β_i^{IV} and β_π^{IV} as functions of shallow thinking parameter λ

(b) Unconditional sensitivity β^{OLS} as a function of relative importance of monetary policy shocks ζ_i

Figure C6: Univariate regression coefficients of long-term rates on short-term rates

Notes: Panel (a) plots the conditional sensitivities $\beta_i^{IV}, \beta_\pi^{IV}$ of the long-term rate with $T = 2$ on the short-term rate, conditional on monetary policy shocks (in brown) and cost-push shocks (in purple) respectively, under $\rho = 0.4$, as functions of the shallow thinking parameter λ.
Panel (b) plots the unconditional sensitivity β^{OLS} of the long-term rate with $T = 2$, under $\rho = 0.4$, as a function of the relative importance of monetary policy shocks ζ_i. The blue line indicates our calibrated λ, while the green line represents the rational expectations equilibrium (REE) with $\lambda = 1$.

D Appendix to the RBC Model

D.1 The RBC Model

Under our assumption that agents observe all variables in each period and have expectations formed once and for all after observing the shocks, i.e., $\mathbb{E}_t [v_s] \equiv \mathbb{E} [v_s]$ for $s > t$, we omit the time subscript on expectations for simplicity.

Firms. There is a continuum of firms with identical production technology and capital stock at the steady state but differing beliefs about impulse responses to shocks. To simplify notation, we consider a generic firm without firm-specific indexing, noting that once we aggregate across firms, the average expectations determine aggregate behavior under a first-order approximation around the steady state.

Firms maximize

$$\max_{\{N_t^d, I_t, Y_t\}} \mathbb{E} \sum_{t=0}^{\infty} \prod_{k=0}^{t-1} (1+r_k)^{-1} DIV_t$$

subject to

$$DIV_t = Y_t - W_t N_t^d - I_t - \frac{\psi}{2}\left(\frac{I_t}{K_t} - \delta\right)^2 K_t \tag{D1}$$

$$Y_t = Z_t \left(\frac{K_t}{\alpha}\right)^{\alpha} \left(\frac{N_t^d}{1-\gamma}\right)^{1-\gamma} \tag{D2}$$

$$K_{t+1} = (1-\delta) K_t + I_t \tag{D3}$$

In each period, firms choose labor demand N_t^d to satisfy

$$Z_t \left(\frac{K_t}{\alpha}\right)^{\alpha} \left(\frac{N_t^d}{1-\gamma}\right)^{-\gamma} = W_t$$

$$N_t^d = (1-\gamma)\left(\frac{Z_t}{W_t}\right)^{\frac{1}{\gamma}} \left(\frac{K_t}{\alpha}\right)^{\frac{\alpha}{\gamma}} \tag{D4}$$

We form the Lagrangian after plugging in the optimal labor demand as

$$\mathbb{E} \sum_{t=0}^{\infty} \prod_{k=0}^{t-1} R_k^{-1} \left[\gamma Z_t^{\frac{1}{\gamma}} W_t^{1-\frac{1}{\gamma}} \left(\frac{K_t}{\alpha}\right)^{\alpha/\gamma} - I_t - \frac{\psi}{2}\left(\frac{I_t}{K_t} - \delta\right)^2 K_t\right] + \prod_{k=0}^{t-1} R_k^{-1} q_t \left[(1-\delta) K_t + I_t - K_{t+1}\right]$$

Its FOC w.r.t. I_t is

$$-\left[1 + \psi\left(\frac{I_t}{K_t} - \delta\right)\right] + q_t = 0$$

which pins down the investment as

$$I_t = \left(\frac{q_t - 1}{\psi} + \delta\right) K_t \tag{D5}$$

Its FOC w.r.t. K_{t+1} is

$$\mathbb{E}\left[Z_{t+1}^{\frac{1}{\gamma}} W_{t+1}^{1-\frac{1}{\gamma}} \left(\frac{K_{t+1}}{\alpha}\right)^{\frac{\alpha}{\gamma}-1} - \frac{\psi}{2}\left(\frac{I_{t+1}}{K_{t+1}} - \delta\right)^2 + \psi\left(\frac{I_{t+1}}{K_{t+1}} - \delta\right)\frac{I_{t+1}}{K_{t+1}} + q_{t+1}(1-\delta)\right] - (1+r_t) q_t = 0$$

and thus

$$q_t = (1+r_t)^{-1}\,\mathbb{E}\left[Z_{t+1}^{\frac{1}{\gamma}} W_{t+1}^{1-\frac{1}{\gamma}} \left(\frac{K_{t+1}}{\alpha}\right)^{\frac{\alpha}{\gamma}-1} - \frac{\psi}{2}\left(\frac{I_{t+1}}{K_{t+1}} - \delta\right)^2 + \psi\left(\frac{I_{t+1}}{K_{t+1}} - \delta\right)\frac{I_{t+1}}{K_{t+1}} + (1-\delta)\,q_{t+1}\right] \tag{D6}$$

(D1-D6) implicitly characterize the dependence of output Y_t, investment (including adjustment cost) $I_t + \Psi_t$ and labor demand N_t^d on the past and current interest rates $\{r_s\}_{0\leq s\leq t}$ and wages $\{W_s\}_{0\leq s\leq t}$, as well as expectations of future interest rates $\{\mathbb{E}[r_s]\}_{s>t}$ and wages $\{\mathbb{E}[W_s]\}_{s>t}$. As we adopt a first-order approximation around the steady state, such implicit dependence is linear. As a result, once aggregated across all firms, the aggregate behavior depends on the average expectations of future interest rates $\left\{\overline{\mathbb{E}}[r_s]\right\}_{s>t}$ and wages $\left\{\overline{\mathbb{E}}[W_s]\right\}_{s>t}$.

Households. The household side of the RBC economy is the same as that of the New Keynesian economy, except that they invest in a real bond as opposed to a nominal bond. Thus the consumption and labor supply functions follow from (C4, C5) by replacing $i_{t+k} - \pi_{t+k+1}$ with $\frac{r_{t+k}-\bar{r}}{1+\bar{r}}$, where the normalization translates a level change of the real interest rate into a log change. Consumption and labor supply C_t, N_t^s depend on the interest rate, wage and dividend $\{r_t, W_t, DIV_t\}_{t\geq 0}$ as follows

$$\begin{aligned}
\frac{C_t - \overline{C}}{\overline{C}} &= -\sigma^{-1}\beta\frac{r_t-\bar{r}}{1+\bar{r}} - \sigma^{-1}\beta\sum_{k=1}^{\infty}\beta^k\overline{\mathbb{E}}\left[\frac{r_{t+k}-\bar{r}}{1+\bar{r}}\right] + (1-\beta)\left[\frac{(\mu-1)\,\nu}{\sigma+\mu\nu}\frac{DIV_t}{\overline{DIV}} + \frac{(1+\nu)}{\sigma+\mu\nu}\frac{W_t}{\overline{W}}\right] \\
&\quad + (1-\beta)\sum_{k=1}^{\infty}\beta^k\overline{\mathbb{E}}\Big[\frac{(\mu-1)\,\nu}{\sigma+\mu\nu}\frac{DIV_{t+k}}{\overline{DIV}} + \frac{(1+\nu)}{\sigma+\mu\nu}\frac{W_{t+k}}{\overline{W}}\Big]
\end{aligned} \tag{D7}$$

$$\begin{aligned}
\frac{N_t^s - \overline{N}}{\overline{N}} &= \nu^{-1}\beta\frac{r_t-\bar{r}}{1+\bar{r}} + \nu^{-1}\beta\sum_{k=1}^{\infty}\beta^k\overline{\mathbb{E}}\left[\frac{r_{t+k}-\bar{r}}{1+\bar{r}}\right] - (1-\beta)\frac{(\mu-1)\,\sigma}{\sigma+\mu\nu}\frac{DIV_t}{\overline{DIV}} + \nu^{-1}\left(1 - \sigma\frac{(1-\beta)(1+\nu)}{\sigma+\mu\nu}\right)\frac{W_t}{\overline{W}} \\
&\quad - (1-\beta)\sum_{k=1}^{\infty}\beta^k\overline{\mathbb{E}}\Big[\frac{(\mu-1)\,\sigma}{\sigma+\mu\nu}\frac{DIV_{t+k}}{\overline{DIV}} + \frac{(1+\nu)\,\sigma\nu^{-1}}{\sigma+\mu\nu}\frac{W_{t+k}}{\overline{W}}\Big]
\end{aligned} \tag{D8}$$

Goods and labor markets. The interest rates and wages arise from clearing the goods and labor markets,

$$N_t^s = N_t^d \tag{D9}$$

and

$$Y_t = I_t + \Psi_t + C_t \tag{D10}$$

Equilibrium. We study a *temporary equilibrium* in which agents maximize their utilities, taking as given the average expectations $\overline{\mathbb{E}}_t[\cdot]$, and markets clear. In this RBC model, equations (D1-D10) characterizes the equilibrium given the average expectations.

We adopt a quarterly calibration of the RBC economy, with all parameters listed in Table D1. We assume slight decreasing returns to scale in production ($\alpha < \gamma$) and the existence of a small fringe of rational agents, discussed in detail after introducing shallow thinking.

Table D1: Quarterly calibration of the RBC economy

Parameter	Description	Value	Estimate/Target
Beliefs			
λ	Continuation rate of depth of thinking	0.61	Our survey evidence
ϑ	Share of rational agents	0.1	
Firms			
Z	Productivity	0.33	Steady state $\overline{Y} = 1$
ρ	Persistence of productivity shocks	0.979	King and Rebelo (1999)
α	Capital share	0.25	
$1-\gamma$	Labor share	0.67	
δ	Capital depreciation rate	0.025	
ψ	Capital adjustment cost	1	
Households			
β	Discount factor	0.99	Steady state annual $\overline{r} = 4\%$
σ^{-1}	Elasticity of intertemporal substitution (EIS)	1	
ν^{-1}	Frisch elasticity	0.5	
φ	Labor disutility scale	0.81	Steady state $\overline{N} = 1$

D.2 Shallow Thinking and Modeling Choices

Similar to Section 5.1, we represent the RBC economy by the sequences of eight variables $\left\{r_t, Y_t, I_t, DIV_t, N_t^d, C_t, N_t^s, W_t\right\}_{t\geq 0}$, which we refer to as $\mathcal{V}$.[39] Six of them are agents' actions, which we collect as $\mathcal{V}^{action}$, whereas the other two, $\{r_t, W_t\}_{t\geq 0}$, are prices formed in the competitive goods and labor market.

We characterize the rational expectations equilibrium (REE) and introduce shallow thinking beliefs accordingly. For REE, each action $\mathbf{v} = (\{v_t\}_{t\geq 0})'$ in $\mathcal{V}^{action}$ can be represented

[39]For brevity, we ignore the adjustment cost Ψ_t, which is 0 to the first order around the steady state since the adjustment cost is quadratic.

in the sequence space as

$$\mathbf{v}^{REE} = \sum_{\mathbf{u}\in\mathcal{V}} \mathbf{J}_{\mathbf{vu}}\mathbf{u}^{REE} + \boldsymbol{\epsilon}^{v},\ \forall \mathbf{v} \in \mathcal{V}^{action} \tag{D11}$$

where $\mathbf{J}_{\mathbf{vu}}$ encodes the dependence among sequences of variables. Households' consumption $\{C_t\}_{t\geq 0}$ and labor supply $\left\{N_t^s\right\}_{t\geq 0}$ depend on the interest rate $\{r_t\}_{t\geq 0}$, dividend $\{DIV_t\}_{t\geq 0}$, and wage $\{W_t\}_{t\geq 0}$. Firms' output $\{Y_t\}_{t\geq 0}$, investment $\{I_t\}_{t\geq 0}$, dividend $\{DIV_t\}_{t\geq 0}$, and labor demand $\left\{N_t^d\right\}_{t\geq 0}$ are functions of the interest rate $\{r_t\}_{t\geq 0}$ and wage $\{W_t\}_{t\geq 0}$, subject to the productivity shock $\{Z_t\}_{t\geq 0}$. Differing from the textbook New Keynesian model with no state variable, in the RBC economy firms have a stock of capital. Thus firms' actions depend on not only current and future interest rates and wages, but also past ones. Yet (D11) is general enough to allow for it.

Instead of using (D11) together with market clearing conditions to solve for the REE, as done in Auclert et al. (2021), we use it to represent the causal relations and reinterpret the market clearing conditions as such. We pedagogically write down these expressions by associating each price with a competitive market, i.e., wages with the labor market and interest rates with the goods market.

For labor supply and demand $\mathbf{N}^s, \mathbf{N}^d$, by separating their dependence on the wage $\mathbf{W}$ from the rest, we interpret them as supply and demand curves in the sequence space

$$\mathbf{v}^{REE} = \mathbf{J}_{\mathbf{vW}}\mathbf{W}^{REE} + \hat{\mathbf{v}}^{REE},\ \mathbf{v} \in \left\{\mathbf{N}^s, \mathbf{N}^d\right\}$$

with elasticities $\mathbf{J}_{\mathbf{N}^s\mathbf{W}}, \mathbf{J}_{\mathbf{N}^d\mathbf{W}}$ and shifts $\hat{\mathbf{N}}^{s,REE}, \hat{\mathbf{N}}^{d,REE}$ defined as

$$\hat{\mathbf{v}}^{REE} = \sum_{\mathbf{u}\in\mathcal{V}\setminus\{\mathbf{W}\}} \mathbf{J}_{\mathbf{vu}}\mathbf{u}^{REE} + \boldsymbol{\epsilon}^{v},\ \mathbf{v} \in \left\{\mathbf{N}^s, \mathbf{N}^d\right\} \tag{D12}$$

In the RBC economy, both supply and demand elasticities $\mathbf{J}_{\mathbf{N}^s\mathbf{W}}, \mathbf{J}_{\mathbf{N}^d\mathbf{W}}$ are non-zero. By equalizing $\mathbf{N}^{s,REE} = \mathbf{N}^{d,REE}$, we can interpret the wage $\mathbf{W}^{REE}$ as resulting from shifts $\hat{\mathbf{N}}^{d,REE}, \hat{\mathbf{N}}^{s,REE}$

$$\mathbf{W}^{REE} = \left(\mathbf{J}_{\mathbf{N}^s\mathbf{W}} - \mathbf{J}_{\mathbf{N}^d\mathbf{W}}\right)^{-1}\left(\hat{\mathbf{N}}^{d,REE} - \hat{\mathbf{N}}^{s,REE}\right) \tag{D13}$$

For output, investment and consumption $\mathbf{Y}, \mathbf{I}, \mathbf{C}$, by separating their dependence on the interest rate $\mathbf{r}$ from the rest, we interpret them as supply and demand curves as well

$$\mathbf{v}^{REE} = \mathbf{J}_{\mathbf{vr}}\mathbf{r}^{REE} + \hat{\mathbf{v}}^{REE},\ \forall \mathbf{v} \in \{\mathbf{Y}, \mathbf{I}, \mathbf{C}\}$$

with elasticities $\mathbf{J_{Yr}}, \mathbf{J_{Ir}}, \mathbf{J_{Cr}}$ and shifts $\hat{\mathbf{Y}}, \hat{\mathbf{I}}, \hat{\mathbf{C}}$ defined as

$$\hat{\mathbf{v}}^{REE} = \sum_{\mathbf{u} \in \mathcal{V} \setminus \{\mathbf{r}\}} \mathbf{J_{vu}} \mathbf{u}^{REE} + \epsilon^{v}, \ \mathbf{v} \in \{\mathbf{Y}, \mathbf{I}, \mathbf{C}\} \tag{D14}$$

By equalizing $\mathbf{Y} = \mathbf{I} + \mathbf{C}$, we can interpret the interest rate as dependent on shifts $\hat{\mathbf{I}}^{REE}, \hat{\mathbf{C}}^{REE}, \hat{\mathbf{Y}}^{REE}$

$$\mathbf{r}^{REE} = (\mathbf{J_{Yr}} - \mathbf{J_{Ir}} - \mathbf{J_{Cr}})^{-1} \left(\hat{\mathbf{I}}^{REE} + \hat{\mathbf{C}}^{REE} - \hat{\mathbf{Y}}^{REE} \right) \tag{D15}$$

To determine beliefs, we reason with $\hat{\mathbf{v}}$ as in (D12, D14) instead of $\mathbf{v}$ in (D11). Taking stock, (D11) and (D13, D15) characterize the REE. We can similarly express it as a linear system and apply Propositions 6 and 7, as in Section 5.1. Figure D1 illustrates the system of causal relations that represent the RBC economy.

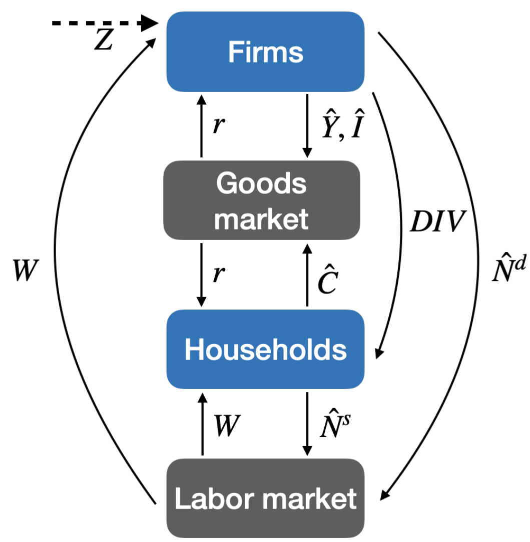


Figure D1: Causal relations of RBC economy in the sequence space

Modeling choices. We make two modeling choices regarding the returns to scale in production and the existence of a fringe of rational agents, to ensure a stationary equilibrium. We leave it for future research to explore other approaches.

Regarding firms' production, we set decreasing returns to scale with $\alpha < \gamma$. The reason is that if firms operate under constant returns, a finite-depth agent who knows the productivity shock $\{Z_t\}_{t \geq 0}$ but does not anticipate changes in wages $\{W_t\}_{t \geq 0}$ or interest rates

$\{r_t\}_{t\geq 0}$ would believe that firms expand their capital stock to a different steady level.[40] That introduces nonstationary beliefs, leading to nonstationary equilibrium responses. In such a case, agents' beliefs will be perpetually wrong, unlike in a stationary case where agents correctly understand the steady state but misjudge the temporary impulse responses. For this reason, we consider decreasing returns to production as the generic case, with constant returns approximated by a small degree of decreasing returns.

Regarding the existence of rational agents, we first note that households' forward-looking behavior is necessary for the existence of a stationary equilibrium in the RBC economy. Suppose we study a production economy with labor as a factor in fixed supply and firms like those in our RBC economy. No equilibrium with stationary $\{r_t, W_t\}_{t\geq 0}$ exists to meet the market clearing conditions $N_t^d = \overline{N}$ and $Y_t = I_t$, even if we assume that firms are rational. Hence, it is the forward-looking behavior of households that supports the RBC economy under rational expectations. In modeling shallow thinking of a dynamic general equilibrium (Proposition 6, Assumptions 1′ and 2), we nest rational expectations with $\lambda = 1$, but otherwise do not allow beliefs to be directly related to the actual equilibrium for simplicity. That is because the causal relations used in Assumption 1′ are those of the rational expectations equilibrium rather than the actual equilibrium. In the RBC economy, if households have beliefs that are entirely disconnected from the actual equilibrium, no stationary equilibrium exists. Thus, we assume that a small ϑ share of agents (both firms and households) are rational, giving rise to average expectations (for a generic variable v) $\overline{\mathbb{E}}^{mix}[v_t] = \vartheta v_t + (1-\vartheta)\overline{\mathbb{E}}[v_t]$ that determine the equilibrium, where v_t is the actual equilibrium outcome and $\overline{\mathbb{E}}[v_t]$ represents the average shallow thinking expectations from Proposition 7. Future work may improve on this ad hoc assumption by modeling shallow thinking as a fixed point, using the causal relations of the actual dynamic general equilibrium.

E Implementing Shallow Thinking with the Sequence-Space Jacobian

We pedagogically write down a 3-step procedure to implement shallow thinking in the RBC economy using the sequence-space Jacobians (SSJ) toolkit developed by Auclert et al.

[40] If there is no adjustment cost ($\psi = 0$), then the agent expects firms to infinitely expand their capital stock and production. With a positive adjustment cost ($\psi > 0$), the agent expects firms to expand their capital stock to a steady level different from before.

(2021). This procedure can also be applied to other models that can be solved in the sequence space, including models with incomplete markets and idiosyncratic shocks.

Step 1: characterizing REE. In the first step, we determine the causal relations as the Jacobians of the rational expectations equilibrium (REE) in terms of agents' best responses, i.e., (D11). We strictly adhere to expressions of agents' choices as functions of decision-relevant variables. For example, household consumption $\{C_t\}_{t\geq 0}$ and labor supply $\left\{N_t^s\right\}_{t\geq 0}$ respond to wages $\{W_t\}_{t\geq 0}$ and dividends $\{DIV_t\}_{t\geq 0}$, instead of aggregate output $\{Y_t\}_{t\geq 0}$. The REE is characterized by these Jacobians together with market clearing conditions.

Step 2: determining shallow thinking beliefs. In the second step, we characterize the average shallow thinking expectations $\overline{\mathbb{E}}[\cdot]$ of the sequences of eight variables $\mathcal{V} = \left\{\mathbf{r}, \mathbf{Y}, \mathbf{I}, \mathbf{DIV}, \mathbf{N}^d, \mathbf{C}, \mathbf{N}^s, \mathbf{W}\right\}$ with the REE Jacobians. In order to use the SSJ toolkit for Proposition 7, we construct a *modified model* as a directed *acyclic* graph (DAG) based on the REE Jacobians. We divide the system of equations in Proposition 7 into two groups, one concerning agents' actions and one about prices from competitive markets.

Based on (D11), the first group of equations concerning agents is

$$\overline{\mathbb{E}}[\mathbf{v}] = \lambda \sum_{\mathbf{u}\in\mathcal{V}} \mathbf{J}_{\mathbf{vu}} \overline{\mathbb{E}}[\mathbf{u}] + \epsilon^v, \ \forall \mathbf{v} \in \mathcal{V}^{action} \equiv \mathcal{V} \setminus \{\mathbf{r}, \mathbf{W}\} \tag{E1}$$

where $\{\mathbf{J}_{\mathbf{vu}}\}$ are the Jacobians of the REE determined in the first step. To embed this in the SSJ toolkit, we construct the *modified households and firms blocks* as follows. We replicate the households and firms blocks with REE Jacobians and modify all the Jacobians by λ.[41] Then, we create Jacobians of shifts $\overline{\mathbb{E}}\left[\hat{\mathbf{N}}^d\right], \overline{\mathbb{E}}\left[\hat{\mathbf{N}}^s\right]$ w.r.t. decision-relevant variables by replicating the Jacobians of $\overline{\mathbb{E}}\left[\mathbf{N}^d, \mathbf{N}^s\right]$ and setting their dependence on $\overline{\mathbb{E}}[\mathbf{W}]$ to zero. Similarly, we create Jacobians of shifts $\overline{\mathbb{E}}\left[\hat{\mathbf{Y}}, \hat{\mathbf{I}}, \hat{\mathbf{C}}\right]$ w.r.t. decision-relevant variables by replicating those of $\overline{\mathbb{E}}[\mathbf{Y}, \mathbf{I}, \mathbf{C}]$ and setting their dependence on $\overline{\mathbb{E}}[\mathbf{r}]$ to zero.

According to (D13, D15), the second group about markets consists of

$$\overline{\mathbb{E}}[\mathbf{W}] = \lambda \left(\mathbf{J}_{\mathbf{N}^s\mathbf{W}} - \mathbf{J}_{\mathbf{N}^d\mathbf{W}}\right)^{-1} \left(\overline{\mathbb{E}}\left[\hat{\mathbf{N}}^d\right] - \overline{\mathbb{E}}\left[\hat{\mathbf{N}}^s\right]\right)$$
$$\overline{\mathbb{E}}[\mathbf{r}] = \lambda \left(\mathbf{J}_{\mathbf{Yr}} - \mathbf{J}_{\mathbf{Ir}} - \mathbf{J}_{\mathbf{Cr}}\right)^{-1} \left(\overline{\mathbb{E}}\left[\hat{\mathbf{I}}\right] + \overline{\mathbb{E}}\left[\hat{\mathbf{C}}\right] - \overline{\mathbb{E}}\left[\hat{\mathbf{Y}}\right]\right)$$

[41] In the RBC model, the households and firms blocks have no cyclic dependence. If there were, one could turn a cycle into a target to use the SSJ toolkit.

which can be rewritten as

$$\lambda\left(\overline{\mathbb{E}}\left[\hat{\mathbf{N}}^d\right]-\overline{\mathbb{E}}\left[\hat{\mathbf{N}}^s\right]\right)-\left(\mathbf{J}_{\mathbf{N}^s\mathbf{W}}-\mathbf{J}_{\mathbf{N}^d\mathbf{W}}\right)\overline{\mathbb{E}}\left[\mathbf{W}\right]=\mathbf{0} \tag{E2}$$

$$\lambda\left(\overline{\mathbb{E}}\left[\hat{\mathbf{I}}\right]+\overline{\mathbb{E}}\left[\hat{\mathbf{C}}\right]-\overline{\mathbb{E}}\left[\hat{\mathbf{Y}}\right]\right)-\left(\mathbf{J}_{\mathbf{Yr}}-\mathbf{J}_{\mathbf{Ir}}-\mathbf{J}_{\mathbf{Cr}}\right)\overline{\mathbb{E}}\left[\mathbf{r}\right]=\mathbf{0} \tag{E3}$$

We set up *fictitious Walrasian auctioneer blocks* as simple blocks that take $\overline{\mathbb{E}}\left[\hat{\mathbf{N}}^d,\hat{\mathbf{N}}^s,\hat{\mathbf{Y}},\hat{\mathbf{I}},\hat{\mathbf{C}}\right]$ and $\overline{\mathbb{E}}\left[\mathbf{W},\mathbf{r}\right]$ as inputs and produce the residuals of (E2, E3) as outputs.

Putting together the modified households and firms blocks and the fictitious Walrasian auctioneer blocks forms a DAG, with $\overline{\mathbb{E}}\left[\mathbf{W},\mathbf{r}\right]$ as unknowns and (E2, E3) as targets. Solving this DAG yields the average shallow thinking expectations $\overline{\mathbb{E}}\left[\cdot\right]$ of the eight variables $\mathcal{V}=\left\{\mathbf{r},\mathbf{Y},\mathbf{I},\mathbf{DIV},\mathbf{N}^d,\mathbf{C},\mathbf{N}^s,\mathbf{W}\right\}$.

Step 3: calculating equilibrium given beliefs. In the last step, we determine the equilibrium given the average shallow thinking expectations. Here we allow for a slight generalization that ϑ share of agents are rational, hence the average expectations that matter for the equilibrium is

$$\overline{\mathbb{E}}^{mix}\left[v_t\right]=\vartheta v_t+(1-\vartheta)\,\overline{\mathbb{E}}\left[v_t\right] \tag{E4}$$

where v_t is the actual equilibrium outcome and $\overline{\mathbb{E}}\left[v_t\right]$ represents the average shallow thinking expectations determined above. The REE is nested by either $\vartheta=1$ or $\lambda=1$ (hence $\overline{\mathbb{E}}\left[v_t\right]=v_t$). Outside the limits, forward-looking agents with beliefs (E4) are surprised in each period t and change their behavior when a decision-relevant variable turns out different from their beliefs ($v_t\neq\overline{\mathbb{E}}^{mix}\left[v_t\right]$). They behave as

$$\mathbf{v}=\sum_{\mathbf{u}\in\mathcal{V}}\left(\mathbf{J}_{\mathbf{vu}}\overline{\mathbb{E}}^{mix}\left[\mathbf{u}\right]+\check{\mathbf{J}}_{\mathbf{vu}}\left(\mathbf{u}-\overline{\mathbb{E}}^{mix}\left[\mathbf{u}\right]\right)\right)+\epsilon^v,\ \forall\mathbf{v}\in\mathcal{V}^{action} \tag{E5}$$

where

$$\left(\check{\mathbf{J}}_{\mathbf{vu}}\right)_{ts}\equiv\begin{cases}\left(\mathbf{J}_{\mathbf{vu}}\right)_{t-s,0} & s\leq t\\ 0 & s>t\end{cases}$$

is the myopic Jacobian that captures responses to variables observed in the past.[42] Rearranging terms and plugging in (E4), we get for any $\mathbf{v} \in \mathcal{V}^{action}$

$$\begin{aligned}
\mathbf{v} &= \sum_{\mathbf{u}\in\mathcal{V}} \check{\mathbf{J}}_{\mathbf{vu}}\mathbf{u} + \sum_{\mathbf{u}\in\mathcal{V}} \left(\mathbf{J}_{\mathbf{vu}} - \check{\mathbf{J}}_{\mathbf{vu}}\right) \overline{\mathbb{E}}^{mix}\left[\mathbf{u}\right] + \boldsymbol{\epsilon}^{v} \\
&= \sum_{\mathbf{u}\in\mathcal{V}} \check{\mathbf{J}}_{\mathbf{vu}}\mathbf{u} + \sum_{\mathbf{u}\in\mathcal{V}} \left(\mathbf{J}_{\mathbf{vu}} - \check{\mathbf{J}}_{\mathbf{vu}}\right)\left(\vartheta \mathbf{u}_t + (1-\vartheta)\,\overline{\mathbb{E}}\left[\mathbf{u}_t\right]\right) + \boldsymbol{\epsilon}^{v} \\
&= \sum_{\mathbf{u}\in\mathcal{V}} \underbrace{\left(\check{\mathbf{J}}_{\mathbf{vu}} + \vartheta\left(\mathbf{J}_{\mathbf{vu}} - \check{\mathbf{J}}_{\mathbf{vu}}\right)\right)}_{\equiv \tilde{\mathbf{J}}_{\mathbf{vu}}} \mathbf{u} + \underbrace{\sum_{\mathbf{u}\in\mathcal{V}} (1-\vartheta)\left(\mathbf{J}_{\mathbf{vu}} - \check{\mathbf{J}}_{\mathbf{vu}}\right)\overline{\mathbb{E}}\left[\mathbf{u}_t\right] + \boldsymbol{\epsilon}^{v}}_{\equiv \tilde{\boldsymbol{\epsilon}}^{v}}
\end{aligned} \tag{E6}$$

where the first part $\tilde{\mathbf{J}}_{\mathbf{vu}}\mathbf{u}$ encodes the dependence among equilibrium outcomes and the second part $\tilde{\boldsymbol{\epsilon}}^{v}$ is determined independent of the equilibrium. Thus $\tilde{\boldsymbol{\epsilon}}^{v}$ is equivalent to a shock to forward-looking agents, which we call *pseudo shocks*.

To determine the equilibrium with the SSJ toolkit, we construct a model by replacing the REE Jacobians for forward-looking agents by $\tilde{\mathbf{J}}_{\mathbf{vu}}$ and adding pseudo shocks $\tilde{\boldsymbol{\epsilon}}^{v}$ in addition to the true shocks $\boldsymbol{\epsilon}^{v}$. We keep the market clearing conditions as they are. Solving for this model yields the equilibrium given shallow thinking beliefs (mixed with a ϑ share of rational beliefs).

[42] See Auclert et al. (2021) and Auclert, Rognlie and Straub (2020) for the validity of this expression to the first order.