Explaining the Macroeconomic Inertia Puzzle

Michael Cai

November 13, 2024

Click here for the most recent version

Abstract

Many macroeconomic models struggle to explain the sluggish response of aggregate variables to sudden shocks and changes in policy. While numerous theories of adjustment frictions and bounded rationality have been proposed to explain this macroeconomic inertia, no consensus has emerged among them. I show that canonical heterogeneous-agent models—the Blanchard (1985) perpetual youth and Bewley (1986) incomplete markets models—are consistent with aggregate consumption inertia if agents' average expectations align with survey expectations of income and interest rates. To determine the causes and analyze the policy implications of inertia, I adopt a model of frictional Bayesian learning that can explain patterns of forecast errors in expectations data that existing theories struggle to account for. Incorporating this form of learning into a standard heterogeneous-agent New Keynesian environment, I provide a theory for how inertia arises endogenously. Inertia results when the equilibrium amplification of an initial shock exceeds expectations, causing expectations to slowly become unanchored. This theory yields a novel drawback to inertial monetary policy rules and the delayed financing of fiscal deficits. Policy regimes that act more gradually experience longer transmission lags.

*Northwestern University. Email: michaelcai@u.northwestern.edu

I am deeply grateful to my advisors—George-Marios Angeletos, Matthias Doepke, and Matthew Rognlie—for their encouragement and guidance. I would also like to thank Patrick Adams, Bence Bardóczy, Kwok Yan Chiu, Lawrence Christiano, Keshav Dogra, Martin Eichenbaum, Wouter den Haan, Stephan Hobler, Ethan Ilzetzki, Yuriy Gorodnichenko, Joao Guerreiro, Fergal Hanks, Jonathon Hazell, Zhen Huo, Diego Känzig, Chen Lian, Xiaojie Liu, Evan Majic, Pooya Molavi, Jane Olmstead-Rumsey, Giorgio Primiceri, Ricardo Reis, Aditya Soenarjo, Amilcar Velez, Rongxuan (Dalton) Zhang and participants of the Northwestern Macro Lunch for insightful discussion and helpful comments. I gratefully acknowledge the financial support of the Alfred P. Sloan Foundation Pre-Doctoral Fellowship in Behavioral Macroeconomics awarded by the NBER.

1 Introduction

Macroeconomic variables often respond slowly and persistently to sudden shocks and policy changes[1]. Understanding the sources of this macroeconomic inertia is an important challenge for business cycle research for two reasons. First, long lags in monetary and fiscal policy transmission make them unreliable tools for stabilizing business cycle fluctuations. Second, the textbook New Keynesian model, a leading framework for business cycle analysis, cannot rationalize these slow responses and instead predicts that policy actions have immediate effects. Theories of adjustment frictions and bounded rationality have been proposed to limit agents' responsiveness and foresight, two features that prevent this model from generating inertia. However, no consensus has emerged as for which theory best accounts for the evidence.

This paper focuses on explaining the observed inertia in aggregate consumer spending because it plays a central role in facilitating monetary and fiscal policy transmission. I first adopt a semi-structural approach to analyze consumption-savings models without imposing restrictive assumptions on expectation formation or an equilibrium environment. My goal is to identify whether the basic structure of these models is consistent with aggregate consumption inertia without taking a stand on the theory that produces it. I first compute the empirical impulse responses of survey expectations and realizations of income and interest rates to an identified shock. I then test whether model-implied consumption, conditional on these measured income and interest rate paths, matches the empirical consumption impulse response.

My first main result finds that canonical heterogeneous-agent models[2] can match observed aggregate consumption inertia if agents' average expectations of income and interest rates align with average survey expectations. However, the standard representative-agent model implies an overly muted consumption response that fails to match observed consumption because of agents' low marginal propensities to consume. The consistency of heterogeneous-agent models with aggregate consumption inertia does not explain the underlying causes or counterfactual policy implications of inertia. I next adopt a fully structural approach to develop an equilibrium theory of inertia, using a model of frictional Bayesian learning disciplined to the expectations data and a New Keynesian environment.

My second main result explains why heterogeneous-agent models are not only consistent with aggregate consumption inertia but also fundamentally contribute to its emergence. Inertia arises when the "belief multiplier", a key model quantity representing the size of equilibrium amplification, interacts with learning frictions to prolong shock propagation. This interaction is summarized by a positive feedback loop following an initial shock. Consumption based on distorted beliefs induces unexpectedly large amplification in equilibrium output. This surprise reinforces the initial distortion in beliefs, which is only gradually corrected. Notably, the equilibrium amplification from the belief multiplier is only large in heterogeneous-agent economies. This feedback mechanism also makes policies that require agents to reason through the future equilibrium consequences of existing policy commitments less effective and delays policy transmission as a result.

[1] These include monetary policy shocks (Romer and Romer 2004), productivity shocks (Kurmann and Sims 2021), government spending shocks (Ramey 2011), oil price shocks (Känzig 2021), and "max-share" shocks that explain the majority of business cycle fluctuations in a large panel of macroeconomic aggregates (Angeletos et al. 2020).

[2] I focus on the perpetual youth, overlapping generations model (Blanchard 1985, Yaari 1965) and the standard incomplete markets model with idiosyncratic risk and borrowing constraints (Bewley 1986, Imrohoroğlu 1989, Huggett 1993, Aiyagari 1994).

A semi-structural approach to matching observed inertia Can we test the consistency of consumption-savings models with observed consumption inertia without relying on any particular theory of adjustment frictions or expectation formation? I address this question by building on the approaches in Auclert et al. (2020) and Bardóczy and Guerreiro (2023) to solve and estimate heterogeneous-agent models without the rational expectations assumption. I utilize an "aggregate consumption function" representation implied by these models that remains agnostic to how expectations are formed. This function takes in the full history of (cross-sectional) average subjective expectations formed for all future horizons and the realizations of aggregate income and interest rates. I then replace the average subjective expectation of agents within the model with the average expectation reported in the survey data. Given that these data are available[3], one can simply plug them into this model representation and evaluate model-implied aggregate consumption.

To assess model-implied consumption impulse responses against empirical ones, I build on the estimation approach from Barnichon and Mesters (2020) and Lewis and Mertens (2022). In contrast to traditional impulse response matching (Christiano et al. 2005), which requires a fully specified equilibrium model, I directly use the empirical impulse responses of realized and expected income and interest rates to compute the model-implied consumption response to the same identified shock. "Regressing" the model-implied impulse response of consumption on its observed response to estimate model parameters can be interpreted as an impulse response matching method that relies on fewer structural assumptions. This approach narrowly focuses on the consistency of consumption models with consumption inertia without introducing a potentially misspecified model of expectations or confounding components of the equilibrium environment that may independently induce inertia.

Given the dynamics of survey expectations are sufficient to generate model-implied consumption inertia that matches the data, what features or biases explain the expectations data? The patterns of forecast errors in survey expectations are inconsistent with full-information rational expectations, which requires the ex-ante unpredictability of ex-post forecast errors. Models of expectation formation, which uniformly predict under- or overreaction relative to the full-information rational expectations benchmark, fare no better given the varying degrees of under- and overreaction of survey expectations across variables and time.

The simplest explanation for this bias turns out to be persistent over-extrapolation of the current observation to expectations of future horizons. For example, upon observing the current period income realization conditional on the shock, average forecaster expectations of future income anchor on the current realization. If income is low today, they expect it to remain at the same level tomorrow and so on. This over-extrapolation can be fairly long-lived, in certain cases persisting for many quarters after the initial shock impact period.

Inertia as an equilibrium phenomenon Why do expectations that over-extrapolate result in aggregate consumption inertia? To jointly explain these features of the data and to consider policy counterfactuals, I now impose additional structure in the form of a model of expectation formation, disciplined to the data, and a standard New Keynesian equilibrium

[3] Far-horizon expectations that are not reported in the survey data need to be extrapolated. I discuss the details of this procedure briefly in Section 3.4 and also in Appendix A.3.

environment. I assume expectations are formed with Bayesian learning[4] in an information environment where fundamental shocks are comprised of transitory and persistent components that are imperfectly observed. When agents mistakenly attribute a transitory change in the observed variable to a persistent shock, their expectations over-extrapolate the current observation forward.

Even in the rational learning baseline, agents cannot immediately disentangle the transitory from the persistent shock components. However, the impact on equilibrium output of their consumption decisions based on these imperfect beliefs does not distort their ability to learn about these shocks. They understand how equilibrium output responds endogenously to their actions and therefore can learn about the shocks by observing output over time just as effectively as if it were a pure tracking problem.

I consider an important deviation from this baseline by imposing agents' perceived equilibrium output law of motion is "truncated" relative to the actual equilibrium output law of motion[5]. For example, if the actual equilibrium output law of motion is a function of the full history of shocks, the perceived law of motion may only account for a few recent periods. The consequence is that agents cannot fully internalize the general equilibrium consequences of their misinformed actions on the evolution of their future beliefs. Hence, overly responsive consumption to a transitory demand shock increases realized output (income) in equilibrium by more than expected, thereby reinforcing beliefs that the shock itself was persistent. This positive feedback loop results in the endogenous unanchoring of expectations, which prolongs amplification, impedes learning, and results in inertia.

The two key factors that determine consumption inertia, the size of equilibrium amplification and the degree of unanchoring, are connected by a "belief multiplier" χ. The multiplier χ increases in the income sensitivity of consumption demand but decreases in the interest rate sensitivity. Consequently, it tends to be large in heterogeneous-agent economies but small or even negative in representative-agent ones. I demonstrate formally that inertia emerges when the belief multiplier χ is positive and sufficiently large but is absent otherwise.

How inertia influences policy transmission New trade-offs in stabilization policy arise because shocks are imperfectly observed and their equilibrium consequences are imperfectly understood. A more responsive Taylor rule lowers the belief multiplier χ and tends to reduce inertia. However, an overly-responsive Taylor rule can destabilize the economy. As the Taylor rule coefficient on output[6] crosses an upper threshold, the contribution of positive future output beliefs to current output are outweighed by expected future interest rate contractions. This produces a negative feedback loop between output and beliefs that increases output volatility. I show that a Taylor rule that is not overly-responsive can balance the reduction of inertial propagation against the risk of destabilizing output and beliefs. However, to achieve this balance

[4]This form of learning has also been shown to explain systematic patterns in expectations bias in cross-sectional, experimental, and unconditional time-series evidence (Afrouzi et al. 2023, Crump et al. 2023, Farmer et al. 2024, Nagel 2024).

[5]One interpretation of this friction is that agent cognition faces a complexity limit as in Molavi (2022) or a memory constraint as in Azeredo da Silveira et al. (2024).

[6]For the monetary policy examples, I focus on real output stabilization with respect to demand shocks, given that the divine coincidence of output and inflation stabilization holds in my setting (Blanchard and Galí 2007). Therefore, to simplify analysis I adopt a "real" Taylor rule, which sets the ex-ante real interest rate, conditional on current inflation expectations, as a function of real output (Auclert et al. 2024).

the monetary authority must allow partial pass-through of the initial shock[7].

I next consider the choice of a lagged or "inertial" term in the Taylor rule. With this specification, interest rate policy responds to current output fluctuations but also passes through changes in past policy rates. In standard rational expectations settings, forward-looking agents can understand the dynamic equilibrium implications of interest rate changes stemming from both of these causes equally well. Because agents are able to accurately reason about the far-horizon equilibrium effects of interest rates, highly inertial policy rules can have powerful stabilizing effects on current output even if the current response of interest rates is muted.

I compare the counterfactual implications of demand shock transmission under two monetary policy regimes that vary by the degree of policy inertia. These regimes are chosen to obtain the same discounted sum of squared output deviations, a proxy measure for welfare loss, under rational expectations. However, under frictional learning, the "gradual regime" that has higher policy rule inertia results in larger welfare losses than the low policy inertia regime. The reason is that learning frictions impair agents' ability to comprehend the dynamic equilibrium effects that make inertial Taylor rules effective under rational expectations. When the policy response to a demand shock is delayed, agents instead perceive policy to be less responsive overall. This increases the degree to which expectations are unanchored and reduces the effective amount of stabilization.

Learning frictions also alter the transmission of deficit-financed fiscal policy through the same mechanism. Angeletos et al. (2023) show that the delayed financing of a one-time, unanticipated transfer can substantially amplify the output response to this policy shock under rational expectations. I consider this exercise in a model with learning frictions and show that the peak and cumulative responses of output to a transfer shock shift further out in time as financing is delayed. This shift may be undesirable if policymakers aim to enact a timely and short-lived fiscal stimulus.

Related literature This paper relates to a large literature that seeks to understand and quantify the sources of macroeconomic inertia. A major strand of this literature focuses on preference- and technology-based explanations for the slow adjustment of aggregate variables, such as consumption, inflation, and investment. The main preference-based explanation for consumption takes the form of habit formation in consumption spending (Fuhrer 2000, Dynan 2000, Chetty and Szeidl 2016, Havranek et al. 2017). A separate strand relaxes the full-information, rational expectations (FIRE) assumption, dampening the responsiveness of forward-looking decisions to generate inertia. Theories that depart from FIRE and generate inertia include adaptive learning (Evans and Honkapohja 1999, Williams 2003, Eusepi and Preston 2011, Milani 2011), incomplete information (Woodford 2001), sticky information and expectations (Mankiw and Reis 2002, Carroll et al. 2020), and rational and behavioral inattention (Sims 2003, Luo 2008, Maćkowiak and Wiederholt 2015, Gabaix 2019).

Auclert et al. (2020) was the first paper to point out that models with consumption habits and FIRE cannot simultaneously produce sluggish aggregate consumption adjustment and high marginal propensities to consume (MPCs) in line with microeconomic evidence. They demonstrate that heterogeneous-agent New Keynesian (HANK) models capable of matching

[7] Eusepi et al. (2024) and Christiano and Takahashi (2020) derive similar results showing overly-restrictive policy can be undesirable.

high MPCs must therefore relax the FIRE assumption to be able to match aggregate consumption inertia. Bardóczy and Guerreiro (2023) extend the methodological approach introduced by Auclert et al. (2020) and demonstrate that HANK models can be estimated by replacing a model of expectation formation directly with expectations data using the impulse response matching estimation framework of Christiano et al. (2005).

My paper builds on the approaches of Auclert et al. (2020) and Bardóczy and Guerreiro (2023) by showing that the same methodology can be applied to estimate parameters of the heterogeneous-agent consumption-savings decision without imposing a model of expectation formation or the New Keynesian equilibrium assumptions. I do this by adopting the instrumental variables approach of Barnichon and Mesters (2020) and Lewis and Mertens (2022). By using current and lagged structural shocks as the set of instruments, the instrumental variables estimation of the heterogeneous-agent model can instead be interpreted as the regression of impulse responses of observed consumption on model-implied consumption conditional on expectations data. This approach allows me to retain the impulse response matching interpretation of my results without needing to compute model-implied impulse responses within a fully-specified equilibrium model or impose a model of expectation formation.

My paper extends the literature that use survey expectations data for structural estimation following Manski (2004). Del Negro and Eusepi (2011) incorporate survey inflation expectations into a Bayesian estimation framework for representative-agent equilibrium models with rational expectations. Kosar and O'Dea (2023) discuss wide-ranging applications of expectations data in estimating models of individual and household behavior. Bardóczy and Guerreiro (2023) introduce the use of survey expectations data in HANK estimation.

The theoretical results in my paper complement the work of Angeletos and Huo (2021) and Christiano et al. (2024), which contain similar mechanisms. In these papers, when measures of equilibrium amplification or complementarity are large, learning is delayed and inertia can be prolonged. In contrast to Angeletos and Huo (2021), the model of expectation formation I adopt admits a tractable form for the belief law of motion, where the amplification parameter is a simple function of structural primitives. This allows me to analytically characterize the joint evolution of beliefs with equilibrium outcomes. Christiano et al. (2024) focuses on characterizing the speed of convergence of the perceived equilibrium law of motion to the rational expectations equilibrium. The focus of my analysis is instead on the learning behavior of Bayesian agents with fixed, potentially-flawed perceived laws of motion trying to infer imperfectly observable shocks.

Molavi (2022) is a closely related paper that demonstrates that inertia can arise when agents are constrained to entertain low-dimensional state-space representations of the equilibrium law of motion. I show that this inertia can be exaggerated when distorted beliefs formed by a similar, low-dimensional state-space model are reinforced due to an equilibrium feedback loop that is particularly strong in heterogeneous-agent economies.

Eusepi et al. (2024) demonstrates the equilibrium implications of the same model of expectation formation in a representative-agent New Keynesian model but with a different focus. While these authors prioritize illustrating the limits of short-run stabilization policy when expectations over-extrapolate, a theme I revisit briefly in the policy section of this paper, I focus on the contribution of over-extrapolating expectations in generating inertia and the consequences of this inertia for policy conduct.

2 Model-Implied Impulse Responses Using Expectations Data

This section demonstrates how to compute model-implied impulse responses from structural models with only a limited set of assumptions. The goal is to show how impulse responses using the minimal structure implied by heterogeneous-agent models can rationalize observed aggregate consumption inertia. I first demonstrate how to substitute expectations data into these models in place of a particular model of expectation formation, such as rational expectations. As opposed to specifying an equilibrium environment to determine income and interest rates, I then use empirical impulse responses of realizations and survey expectations of these variables instead. I first illustrate this approach in a familiar representative-agent example and then proceed to the heterogeneous-agent case.

2.1 Representative-agent example

To justify later substituting expectations data, I start by assuming subjective expectations E_t are arbitrary. Given this, a representative household solves the following standard consumption-savings problem

$$\max_{C_t, A_t} \sum_{t=0}^{\infty} \beta^t \zeta_t E_0 \left[\frac{C_t^{1-\sigma^{-1}}}{1-\sigma^{-1}} \right]$$
$$\text{s. to } C_t + A_t = Y_t + (1 + r_{t-1}) A_{t-1}$$

The household consumes C_t and saves in a one-period, risk-free asset A_t, taking as given real income Y_t, ex-ante real interest rates r_t, and a discount factor shock ζ_t.

Taking first-order conditions and linearizing around the steady-state $\beta(1+r) = 1$, we obtain the consumption function, given subjective expectations $E_t[\cdot]$. Let $W_t := r_{t-1}A + (1 + r)A_{t-1}$ denote financial wealth and $\gamma := \sigma\beta - (1-\beta)A$ denote the net interest rate elasticity.

$$C_t = (1-\beta) \left(\sum_{h=0}^{\infty} \beta^h E_t[Y_{t+h}] + W_t \right) - \gamma \sum_{h=0}^{\infty} \beta^{h+1} E_t[r_{t+h}] + \varepsilon_t \tag{1}$$

I will treat ε_t as the primitive demand shock, as opposed to the discount factor shock ζ_t. This shock plays the role of an unobserved source of endogeneity faced by the econometrician in trying to estimate parameters of the household problem. I assume current-period variables Y_t, r_t are directly observed by households in that period and equivalently denoted as horizon $h = 0$ expectations. That is, $E_t[Y_t] \equiv Y_t, E_t[r_t] \equiv r_t$. This serves to simplify the information structure and ease interpretation of the problem. I relax this assumption in the heterogeneous-agent case.

Suppose we want to estimate the model given in Equation (1) to match the empirical impulse response of consumption C_t to an externally-identified, exogenous shock z_t. The classic approach in the literature introduced by Christiano et al. (2005) is to impose full-information rational expectations and construct a fully-specified general equilibrium model, which includes Equation (1). Then, choosing a shock within the equilibrium model that is analogous to the externally-identified shock z_t, one can estimate the model's free parameters by fitting the model impulse response to the empirical one.

Denote the net present value of expected income and rates as

$$\mathcal{Y}_t := \sum_{h=0}^{\infty} \beta^h E_t[Y_{t+h}]$$
$$\mathcal{R}_t := \sum_{h=0}^{\infty} \beta^{h+1} E_t[r_{t+h}]$$

Instead of then imposing assumptions on expectation formation to map $\mathcal{Y}_t, \mathcal{R}_t$ to current period state variables, one could instead substitute expectations data in place of model subjective expectations. Suppose we have imperfectly measured expectations data for income and interest rates for all future horizons, denoted by the expectations operator E_t^{data}. Substituting them into the net present value expressions denoting the new quantities as $\mathcal{Y}_t^{\text{data}}, \mathcal{R}_t^{\text{data}}$ and collecting the differences into the measurement error term $\mu_t := \mathcal{Y}_t - \mathcal{Y}_t^{\text{data}} + \mathcal{R}_t - \mathcal{R}_t^{\text{data}}$, we obtain

$$C_t = (1-\beta)\left(\mathcal{Y}_t^{\text{data}} + W_t\right) - \gamma \mathcal{R}_t^{\text{data}} + \varepsilon_t + \mu_t \tag{2}$$

How do we compute model-implied impulse responses from Equation (2)? Following Barnichon and Mesters (2020), I adopt a semi-structural approach for computing model-implied impulse responses. A key intuition from this paper is the equivalence of instrumental variables estimation of Equation (2) using current and lagged values of the shock $\{z_{t-\ell}\}_{\ell \geq 0}$ and ordinary least-squares regression on impulse responses of the variables to the shock z_t. I make the following standard instrumental variables assumptions

$$\begin{aligned} \text{Exogeneity} \quad & \mathbb{E}[\varepsilon_t z_{t-\ell}] = \mathbb{E}[\mu_t z_{t-\ell}] = 0, \ \forall \ell \geq 0 \\ \text{Relevance} \quad & \mathbb{E}\left[\mathcal{R}_t^{\text{data}} z_{t-\ell}\right] \neq 0 \end{aligned}$$

Note $\mathcal{Y}_t^{\text{data}}$ does not show up in the relevance condition because the discount factor $\beta = (1+r)^{-1}$ is a known quantity. The net interest rate elasticity γ contains the elasticity of intertemporal substitution (EIS) σ, which is the only free parameter to estimate. Post-multiplying and taking expectations we see immediately that the new explanatory and response variables are unnormalized estimands of a local projection onto the shock z_t or equivalently their impulse responses

$$\mathbb{E}[C_t z_{t-\ell}] = (1-\beta)\left(\mathbb{E}[(\mathcal{Y}_t^{\text{data}} + W_t) z_{t-\ell}]\right) - \gamma \mathbb{E}[\mathcal{R}_t^{\text{data}} z_{t-\ell}], \ \ell \geq 0$$

The left-hand side of the above expression can be interpreted as the impulse response of consumption to the shock $\{z_{t-\ell}\}_{\ell \geq 0}$ implied by the representative-agent model.

To write a more general representation of the consumption function that will lead us to the heterogeneous-agent case, let's re-consider Equation (2). Model-implied consumption C_t is the sum of endogenous and exogenous components, $\mathcal{C}_t, \mathcal{E}_t$ respectively

$$C_t = \mathcal{C}_t(\mathcal{Y}_t^{\text{data}}, \mathcal{R}_t^{\text{data}}, W_t; \sigma) + \mathcal{E}_t(\varepsilon_t, \mu_t) \tag{3}$$

Using the household budget constraint we can recursively substitute the asset state A_{t-1} out of the above expression we obtain an alternate representation

$$C_t = \mathcal{C}_t(\{\mathcal{Y}_\tau^{\text{data}}, \mathcal{R}_\tau^{\text{data}}, r_{\tau-1}A\}_{\tau\leq t};\sigma) + \mathcal{E}_t(\{\varepsilon_\tau, \mu_\tau\}_{\tau\leq t}) \tag{4}$$

These representations are connected by a simple but useful intuition. Current financial wealth accumulates the *total* effect of past consumption-savings decisions based on past expectations and realizations on current consumption. By integrating out this state variable, we can instead compute the *individual* contribution of a particular past belief, say $E_\tau^{\text{data}}[Y_{\tau+h}]$, on past expected income $\mathcal{Y}_t^{\text{data}}$ and in turn on current consumption C_t through Equation (4).

While state variables like W_t are low-dimensional in representative-agent models, they can be function-valued in heterogeneous-agent models. This complicates both measuring these variables in the data and obtaining solution representations in the form of Equation (3). However, as I will show in the following section, with a few limited assumptions one can obtain an analogous aggregate consumption function representation to Equation (4) in heterogeneous-agent models, like the standard incomplete markets model.

To conclude this example, applying our instrumental variables conditions to Equation (4) we have the moment condition that encapsulates model-implied impulse response estimation

$$\mathbb{E}[(C_t - \mathcal{C}_t(\{\mathcal{Y}_\tau^{\text{data}}, \mathcal{R}_\tau^{\text{data}}, r_{\tau-1}A\}_{\tau\leq t};\sigma_0))z_{t-\ell}] = 0,\ \ell \geq 0 \tag{5}$$

when $\sigma = \sigma_0$, the true parameter value. I proceed to set up the general moment conditions that I will use in model estimation in Section 3, which include Equation (5) as a special case.

2.2 Heterogeneous-agent consumption functions and moment conditions

I now derive a moment condition akin to Equation (5) for a class of heterogeneous-agent models that include idiosyncratic risk, incomplete markets, and borrowing constraints (Bewley 1986, Imrohoroğlu 1989, Huggett 1993, Aiyagari 1994). Doing so allows us to assess whether the heterogeneous-agent, model-implied impulse response of aggregate consumption matches the empirical impulse response well. I first follow Auclert et al. (2020), Auclert et al. (2021), and Bardóczy and Guerreiro (2023)[8] to demonstrate how aggregated consumption decisions from a first-order solution of heterogeneous-agent models can be written as functions of histories of average subjective expectations and realizations. I then adopt a separate estimation approach and outline the assumptions required to construct the moment conditions.

Dynamic programming problem setup Suppose individuals-i have time-t subjective expectations $E_{i,t}[\cdot]$. Let an individual-i's optimization problem be defined by the following value function with common structural parameters $\boldsymbol{\theta}$ that we seek to estimate

$$\mathbf{v}_{i,t} = v(E_{i,t}[\mathbf{v}_{i,t+1}], \mathbf{S}_{i,t}; \boldsymbol{\theta}) \tag{6}$$

An individual-i's state $\mathbf{S}_{i,t} = (\mathbf{s}_{i,t}, \mathbf{X}_t)$ has an idiosyncratic component $\mathbf{s}_{i,t}$ and an aggregate

[8] The problem setup that leads to the consumption function derivation is also outlined in Bardóczy and Guerreiro (2023), albeit focusing on a different representation of this consumption function in terms of forecast errors and revisions. These representations are equivalent.

one $\mathbf{X}_t$. Individuals-i take aggregates $\mathbf{X}_t$ as given[9], which represent either aggregate exogenous shocks or aggregate endogenous variables determined in equilibrium. I purposefully deviate from the usual definition of the state $\mathbf{S}_{i,t}$, which includes the full information set of expectation $E_{i,t}$, to help conceptually separate variables which directly influence the decision problem in Equation (6) via current constraints and flow payoffs, denoted $\mathbf{S}_{i,t}$, and those that may also influence expectation formation, implicit in $E_{i,t}$. This distinction allows me to state assumptions more clearly.

Assumption 1. (**Idiosyncratic rationality**) Exogenous idiosyncratic states $\mathbf{x}_{i,t} \subset \mathbf{s}_{i,t}$ are stochastic processes described by a finite-state, positive recurrent Markov chain that is common and known across individuals-i.

Because I focus on the effects of the bias in average expectations on aggregate realizations, I reduce the generality of this problem setup, only considering deviations from full-information, rational expectations at the aggregate level. By Assumption 1, subjective expectations of functions of idiosyncratic states are therefore taken with respect to their true density. I place the following additional structure on subjective expectations.

Assumption 2. *Individual-i, time-t subjective conditional expectations $E_{i,t}[\cdot]$ satisfy*

a) **Consistency** $E_{i,t}[\mathbf{v}_{i,t+h}] = E_{i,t}[v(E_{i,t+h}[\mathbf{v}_{i,t+h+1}], \mathbf{S}_{i,t+h}; \boldsymbol{\theta})]$ for $h > 1$ via Equation (6)

b) **Independence** variables inducing heterogeneity in $E_{i,t}$ are independent of $\mathbf{s}_{i,t}$

c) **Law of iterated expectations**

The main substantive assumption is independence, which restricts correlated heterogeneity in expectations of aggregate variables with individual-level characteristics. An example of a violation of this assumption would be if individual attention correlated with the incidence of aggregate income on individual income, as in Guerreiro (2023). While this restriction is not innocuous, I proceed with it nonetheless given many theories of bounded rationality with respect to aggregate variables, such as sticky or exogenous noisy information, satisfy this condition.

Let $\mathcal{I}$ denote the index set of all individuals-i. The law of motion of the full distribution of individual-i state variables $\mathbf{D}_t$ is then defined by the transition equation

$$\mathbf{D}_{t+1} = \Lambda(\{\mathbf{v}_{i,t}\}_{i\in\mathcal{I}}, \mathbf{X}_t, \mathbf{D}_t) \tag{7}$$

The aggregation of individual decision rules $y_{i,t}(E_{i,t}[\mathbf{v}_{i,t+1}], \mathbf{X}_t; \boldsymbol{\theta}_i)$ is given by

$$Y_t = \mathcal{Y}(\{y_{i,t}(\mathbf{v}_{i,t})\}_{i\in\mathcal{I}}, \mathbf{D}_t) \tag{8}$$

for a scalar, aggregate output variable Y_t.

Definition 1. A **steady state** $(\{\mathbf{v}_i, \mathbf{s}_i\}_{i\in\mathcal{I}}, \mathbf{D}, \mathbf{X}, Y)$ is the constant-valued fixed point consistent with (6, 7, 8).

Given **Assumptions** 1, 2 and the system of equations (6, 7, 8), the first-order response (locally

[9]That is, they act as if they are atomistic with respect to aggregate outcomes $\mathbf{X}_t$, even though their actions may collectively determine $\mathbf{X}_t$ variables in equilibrium.

around a steady state) of aggregated decisions Y_t to an aggregate shock $\mathbf{X}_t$ is given by

$$Y_t = \sum_{\tau \leq t} \sum_{h \geq 0} \sum_{X \in \mathbf{X}} \mathcal{F}^X_{t-\tau,h}(\boldsymbol{\theta}) E_\tau [X_{\tau+h}] \tag{9}$$

where $\mathcal{F}^X_{t-\tau,h}(\boldsymbol{\theta})$ is the $(t-\tau, h)$ entry of the "fake news" matrix $\mathcal{F}^X$, defined in Auclert et al. (2021) for an aggregate output variable Y with respect to an aggregate input variable X.

Our earlier representative-agent consumption function fits into the general representation given by Equation (9). Given the standard incomplete markets model of household consumption-savings (Bewley 1986, Imrohoroğlu 1989, Huggett 1993, Aiyagari 1994) can be defined by the above system, we can derive a consumption function for this model under the same limited set of assumptions on expectation formation.

Let us consider the heterogeneous-agent consumption function implied by Equation (9) and interpret the coefficients on its typical inputs: income, interest rates, and demand shocks[10].

$$C_t = \sum_{\tau \leq t} \sum_{h \geq 0} \mathcal{F}^Y_{t-\tau,h} E_\tau [Y_{\tau+h}] + \mathcal{F}^r_{t-\tau,h} E_\tau [r_{\tau+h}] + \mathcal{F}^\varepsilon_{t-\tau,h} E_\tau [\varepsilon_{\tau+h}]$$

Similarly to the representative-agent case in Equation (4), aggregate consumption is a function of both current and past (average) expectations of income, interest rates, and demand shocks. The coefficient $\mathcal{F}^Y_{t-\tau,h}$ can be interpreted as the effect on consumption at time-t of a change in expectations formed at an earlier time-τ of time-$\tau+h$ income holding all other beliefs and realizations across time and variables fixed. While Auclert et al. (2021) offers a detailed explanation of these terms, I offer an intuitive explanation to help understand its construction.

Suppose at time-τ households suddenly thought income at a future time-$\tau+h$ would be higher. Given these beliefs, they would borrow in advance of the future income receipt to smooth consumption. However, given other beliefs and realizations are fixed, in the next period time-$\tau+1$ households realize their expectation in the prior period was mistaken, hence the name "fake news". Nonetheless, the consequences of dissaving or borrowing in the prior period affect the wealth they inherit in the current period. This in turn affects their subsequent savings decisions many periods afterward, resulting in consequences that last through time-t.

The logic of the coefficients in $\mathcal{F}$ applies identically whether expectations are formed with full-information rational expectations, or any other deviation satisfying the above assumptions. If households collectively expect future income will be higher, the average effect on consumption from the mean belief[11] is the same irrespective of whether the mean belief is accurate or distorted. Likewise, when the future period arrives and realized income is lower or higher than expected, the effect on household spending is equivalent to the effect of an unanticipated income shock. Spending out of this unanticipated income shock is the sum of spending out of realized current income, which is not a function of expectations, and spending out of anticipated future income due to potentially revised mean beliefs following the shock. As before,

[10] The fake news matrix for ε_t depends on the particular micro-foundation one uses for the primitive demand shock that comprises this exogenous intercept term. For example, if ε_t is a linear combination of discount factor shocks, then $\mathcal{F}^\varepsilon$ will be functions of the interest rate matrix $\mathcal{F}^r$, since discount factor shocks alter consumption similarly to perturbations to the ex-ante interest rate.

[11] The average subjective belief is the only moment from the subjective probability distribution over aggregate variables that determines household consumption decisions because of certainty equivalence from the linearization.

the latter response is the same whether mean beliefs are accurate or distorted.

Moment conditions I proceed to construct moment conditions to estimate heterogeneous-agent models using expectations data, following Barnichon and Mesters (2020). Suppose we have a vector of current and lagged structural shocks $\mathbf{z}_t = \{z_{t-\ell}\}_{\ell=0,\ldots,N_\ell}$ to use as instruments.

Assumption 3. **(Serially uncorrelated)** *$z_{t-\ell}$ are serially uncorrelated across ℓ*

Partition $\mathbf{X}_t$ into variables unobserved by the econometrician, $\boldsymbol{\varepsilon}_t$, and those observed, $\mathbf{W}_t$.

Assumption 4. **(Exogeneity)** $\mathbb{E}[E_\tau[\varepsilon_{\tau+h}]\,z_{t-\ell}] = 0, \quad \forall\, h \geq 0,\, \varepsilon_{\tau+h} \in \boldsymbol{\varepsilon}_{\tau+h},\, z_{t-\ell} \in \mathbf{z}_t$

This exogeneity condition is slightly more general than the one in the representative-agent example, encompassing the case where shocks may be imperfectly observed by economic agents. For $\tau < t-\ell$, if $z_{t-\ell}$ instruments are not systematically predictable by information available prior to time-$t-\ell$, it is natural to assume orthogonality to measurable functions of earlier information sets, i.e. $E_\tau[\varepsilon_{\tau+h}]$ for $\tau < t-\ell$. For $\tau \geq t-\ell$, this assumption requires agents to be aware that the instrument $z_{t-\ell}$ is orthogonal to time-t information relevant for determining $\varepsilon_{\tau+h}$. Alternatively, one could directly assume the shocks $\varepsilon_{\tau+h}$ are observable by agents but not the econometrician, and that the shocks $\varepsilon_{\tau+h}$ are known to be orthogonal to $z_{t-\ell}$. An example of this would be if ε_{t+h} were household preference shocks known to be orthogonal to a supply or policy shock $z_{t-\ell}$.

Assume **Assumptions 1, 2, 3, 4**. Given cross-sectional average, subjective expectations $E_t[\cdot]$, a vector of instruments $\mathbf{z}_t$, and applying Equation (9), where $\mathbf{W} \subseteq \mathbf{X}$ denotes the subset of aggregate inputs $\mathbf{X}$ that are observable to the econometrician, we obtain the following N_ℓ moment conditions

$$\mathbb{E}\left[\left(Y_t - \sum_{\tau=t-\ell}^{t}\sum_{h\geq 0}\sum_{W\in\mathbf{W}} \mathcal{F}^{W}_{t-\tau,h}(\boldsymbol{\theta}_0)E_\tau[W_{\tau+h}]\right)z_{t-\ell}\right] = 0, \quad \text{for } z_{t-\ell} \in \mathbf{z}_t$$

Where the moment condition equals zero uniquely at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Suppose we have data on realizations and cross-sectional average expectations of income and interest rates up to a finite horizon H, denoted by $\{E_t^{\text{data}}[\mathbf{W}_{t+h}]\}_{h\leq H}$. To be able to evaluate the above moment condition, we need to extrapolate missing expectations data horizons.

Let $F_t[\mathbf{W}_{t+h};\boldsymbol{\vartheta}]$ denote an auxiliary model of subjective expectation $E_t[\mathbf{W}_{t+h}]$, with parameter vector $\boldsymbol{\vartheta}$.

Assumption 5. **(Shape restriction)**

$$\mathbb{E}[(E_t^{\text{data}}[W_{t+h}] - F_t[W_{t+h};\boldsymbol{\vartheta}_0])z_{t-\ell}] = 0, \text{ for } h \leq H, W_{t+h} \in \mathbf{W}_{t+h},\, z_{t-\ell} \in \mathbf{z}_t \tag{10}$$

We can then estimate $\boldsymbol{\vartheta}$ from the auxiliary model using the HN_ℓ moment conditions implied by **Assumption 5**. For example, $\boldsymbol{\vartheta}$ could be coefficients of an AR(p) process fit to the impulse response of each subjective expectation $E_t^{\text{data}}[W_{t+h}]$ across horizons-h and impulse response periods-ℓ to shocks $z_{t-\ell}$.

Let expectations data augmented with missing horizons extrapolated from the auxiliary model be

$$E_t^{\text{data}}[\mathbf{W}_{t+h};\boldsymbol{\vartheta}] := \begin{cases} E_t^{\text{data}}[\mathbf{W}_{t+h}] & \text{if } h \leq H \\ F_t[\mathbf{W}_{t+h};\boldsymbol{\vartheta}] & \text{if } h > H \end{cases}$$

Assumption 6. (Measurement error exogeneity)

$$\mathbb{E}[(E_t[W_{t+h}] - E_t^{\text{data}}[W_{t+h};\boldsymbol{\vartheta}_0])z_{t-\ell}] = 0, \text{ for } W_{t+h} \in \mathbf{W}_{t+h}, z_{t-\ell} \in \mathbf{z}_t$$

With **Assumptions 1, 2, 3, 4, 5, 6** and given data $\{\mathbf{z}_t, Y_t, \{E_t^{\text{data}}[\mathbf{W}_{t+h}]\}_{h\leq H}\}$, we obtain

$$\mathbb{E}\left[\left(Y_t - \sum_{\tau=t-\ell}^{t}\sum_{h\geq 0}\sum_{W\in\mathbf{W}}\mathcal{F}_{t-\tau,h}^{W}(\boldsymbol{\theta}_0)E_\tau^{\text{data}}[W_{\tau+h};\boldsymbol{\vartheta}_0]\right)z_{t-\ell}\right] = 0, \quad \text{for } z_{t-\ell} \in \mathbf{z}_t \tag{11}$$

Equation (11) is a collection of unconditional moment conditions, which can be estimated with two-step generalized method of moments (Newey and McFadden 1994). The first step is the estimation of the nuisance parameter $\boldsymbol{\vartheta}$ used to extrapolate missing data using condition (10). Following that one can evaluate the moment condition to estimate the structural parameters of interest $\boldsymbol{\theta}$. Using Equation (11), I construct the following set of moment conditions for the heterogeneous-agent models that I estimate in the following section.

$$\mathbb{E}\left[\left(C_t - \sum_{\tau=t-\ell}^{t}\sum_{h\geq 0}\mathcal{F}_{t-\tau,h}^{Y}(\boldsymbol{\theta}_0)E_\tau[Y_{\tau+h};\boldsymbol{\vartheta}_0] + \mathcal{F}_{t-\tau,h}^{r}(\boldsymbol{\theta}_0)E_\tau[r_{\tau+h};\boldsymbol{\vartheta}_0]\right)z_{t-\ell}\right] = 0, \quad \text{for } z_{t-\ell} \in \mathbf{z}_t \tag{12}$$

As before, we can interpret the quantities contained in the moment conditions as impulse response coefficients with respect to $z_{t-\ell}$. The moment condition derived earlier in Equation (5) for the representative-agent model can be seen as a special case of the above expression.

Remarks The intuition of the additional Assumption 6 is that the expectations data we use, even for directly observed horizons $h \leq H$, may be an imperfect measurement of the model subjective expectations. This assumption will typically hold for measurement error due purely to noise, such as classical measurement error. However, I will later provide substantive and specific arguments for its validity because of my need to use professional forecaster expectations in place of household expectations when estimating consumption functions. This assumption states that the impulse response of expectations data must resemble the response of agents expectations in the model to the identified shock used as an instrument. This will inform my choice of instruments, which I discuss in depth in the following section.

There is typically sparse availability of expectations data with a large set of horizons H. Therefore, choosing an auxiliary model with a large number of parameters $\boldsymbol{\vartheta}$ may overfit the noise in expectations data, potentially violating Assumption 6. Because this assumption is not explicitly testable, I err on the side of caution by testing robustness of $\boldsymbol{\theta}$ estimates against multiple auxiliary models that are sparsely parameterized. The choice of these models is informed by the impulse response interpretation of moment conditions (10), (11) when using exogenous structural shocks and their lags as instruments. Relying on the typical assumption that impulse response functions are smooth justifies the focus on low-dimensional, smooth auxiliary models for extrapolation.

3 Estimating Model-Implied Consumption Impulse Responses

This section evaluates model-implied impulse responses of aggregate consumption from a set of standard consumption-savings models against the response of realized consumption to an externally-identified, exogenous supply shock. Typical impulse response matching as in Christiano et al. (2005) is done within a fully-specified equilibrium model, requiring assumptions on expectation formation and the equilibrium environment. I instead adopt a semi-structural approach, using moment conditions derived in Section 2 as a set of structural restrictions in an otherwise empirical impulse response estimation framework. This allows me to focus solely on these models' ability to match realized consumption dynamics, taking as given data on realized and expected income and interest rates.

3.1 Consumption functions from structural consumption-savings models

While my main focus is to demonstrate the fit of heterogeneous-agent models to macroeconomic data, it is useful to contrast them with a representative-agent benchmark. Recall the earlier-derived representative-agent consumption function in Equation (1).

$$C_t = (1-\beta)\left(\sum_{h=0}^{\infty}\beta^h E_t[Y_{t+h}] + W_t\right) + \gamma\sum_{h=0}^{\infty}\beta^{h+1}E_t[r_{t+h}]$$

The first heterogeneous agent model I consider is the tractable perpetual youth, overlapping generations model from Angeletos et al. (2023), which builds on Yaari (1965), Blanchard (1985). The consumption function from the model solution linearized around steady state $\beta(1+r) = 1$ is

$$C_t = (1-\beta\omega)\left(\sum_{h=0}^{\infty}(\beta\omega)^h E_t[Y_{t+h}] + W_t\right) - \gamma\sum_{h=0}^{\infty}(\beta\omega)^h E_t[r_{t+h}] \tag{13}$$

where aggregate wealth is given by W_t and the net interest elasticity $\gamma := \sigma\beta\omega - (1-\beta\omega)\beta A$.

The attractiveness of this model is it closely mirrors its representative-agent counterpart, albeit with one additional degree of freedom, $\omega \in [0,1]$, the perpetual youth hazard rate. When $\omega < 1$ the overlapping generations model exhibits greater income sensitivity of consumption, for example as measured by the current marginal propensity to consume (MPC) out of unearned income. Given the discount factor β is pinned down in both models by the steady state real interest rate r, it is not a degree of freedom for estimation. It will prove useful to compare these two models which differ along this single dimension given their otherwise similar structure.

The second heterogeneous-agent model I consider is the standard incomplete markets model of Bewley (1986), Imrohoroğlu (1989), Huggett (1993), Aiyagari (1994). In this model, a unit mass of households face idiosyncratic income risk, borrowing constraints, and incomplete markets in the form of a one-period risk-free asset. The individual-i household problem is given by

$$
\begin{aligned}
V(e_{i,t}, a_{i,t-1}) &= \max_{c_{i,t}, a_{i,t}} \frac{c_{i,t}^{1-\sigma^{-1}}}{1-\sigma^{-1}} + \beta E_t[V(e_{i,t+1}, a_{i,t})|e_{i,t}] \\
c_{i,t} + a_{i,t} &= e_{i,t} Y_t + (1 + r_{t-1}) a_{i,t-1} \\
a_{i,t} &\geq 0
\end{aligned}
$$

where E_t denotes the time-t subjective expectation. Idiosyncratic productivity $e_{i,t}$ is a stationary, commonly-known Markov process with persistence ρ_e and variance σ_e^2. It has a fixed transition matrix $\Pi(e, e')$ with an associated stationary distribution $\pi(e)$ and a stationary mean normalized to one, i.e. $\sum_e \pi(e) e = 1$.

I restate here the aggregate consumption function derived in the previous section

$$
C_t = \sum_{\tau \leq t} \sum_{h \geq 0} \mathcal{F}^Y_{t-\tau,h} E_\tau[Y_{\tau+h}] + \mathcal{F}^r_{t-\tau,h} E_\tau[r_{\tau+h}] \tag{14}
$$

To map model expectations to their data equivalents I will assume that horizon-0 expectations and realizations coincide, i.e. $E_t^{\text{data}}[Y_t] = Y_t$ and $E_t^{\text{data}}[r_t] = r_t$ for all times-t. In general, these consumption functions do not necessarily enforce horizon-0 expectations to align with realizations, where allowing for a wedge between them may be reasonable in certain information settings. One example is a consumption-savings problem where households are rationally inattentive and cannot perfectly observe state variables even after the period they are realized, as in Luo (2008).

All linearized consumption functions are local approximations about a steady state, which I calibrate following McKay et al. (2016).

Parameter	Description	Value
r	Real interest rate	0.005
$\frac{\text{Assets}}{\text{Income}}$	Assets to disposable income ratio	1.4
ρ_e	Idiosyncratic productivity persistence	0.966
σ_e^2	Idiosyncratic productivity variance	0.504

Table 1: Steady state model calibration

Note: The real interest rate and disposable income are both listed at a quarterly frequency.

The representative-agent model discount factor β is pinned down by the steady state real interest rate $\beta(1 + r) = 1$. The standard incomplete markets model discount factor β must be calibrated to hit the asset-to-disposable income calibration target for a given value of the elasticity of intertemporal substitution. This leaves these already sparsely parameterized consumption-savings models with few degrees of freedom to estimate, which I report in Table 2. This is precisely the goal of this estimation: to test the minimal structure implied by these models, without assumptions on expectation formation or a surrounding equilibrium environment.

Parameter	Description
σ	Elasticity of intertemporal substitution (EIS)
ω^{*}	Perpetual youth hazard rate

Table 2: Parameters to estimate

Note: $*$ This parameter is only available for estimation in the perpetual youth overlapping generations model.

3.2 Data

The consumption function representations of all models require data on realized and expected real disposable income and interest rates.

Realizations data The income measure I use is real disposable personal income (DSPIC96), sourced from the Bureau of Economic Analysis. The interest rate measure I use is the nominal federal funds rate (DFF), deflated by one-period ahead realized consumer price index inflation (CPIAUCSL), sourced from the Bureau of Labor Statistics.

In typical linearized macroeconomic models without financial frictions, there is a single interest rate due to no-arbitrage conditions on asset choice. In reality, households face different interest rates for saving or borrowing products, which makes the choice of a single interest rate data series non-obvious. Given our consumption functions are local approximations around a steady state where all households hold either strictly positive wealth or are borrowing constrained, a savings rate is the best analog to the model interest rate. Due to this, I use the federal funds rate as the data series proxying for the model-based savings rate, given savings rates tend to move closely with the federal funds rate.

Expectations data I use consensus expectations reported by the Bluechip Economic Indicators and Financial Forecasts for real disposable personal income, the nominal federal funds rate, and CPI inflation. As mentioned earlier, because I am estimating household consumption functions, a household-level survey like the Michigan Survey of Consumers or the Survey of Consumer Expectations conducted by the Federal Reserve Bank of New York would be most ideal. However, given neither of these sources nor other commonly-used household surveys elicit point forecasts of interest rates, we would have to make auxiliary assumptions to map qualitative responses about interest rates in these surveys to point forecasts. In addition, household surveys tend to report a single shorter horizon, typically one-year ahead, and a longer five-to-ten year horizon forecast. Because of the need to extrapolate horizons it is useful to have a more complete term structure of near-term expectations from the Bluechip. For consistency across available data periods, I use Bluechip expectations for one through four-quarters ahead for each forecasted variable.

Instrumental variables There are other substantive reasons that may alleviate some concerns of using forecaster as opposed to household survey expectations. Rozsypal and Schlafmann (2023) analyze household income expectations from the Michigan survey and find evidence of over-persistence bias, where households extrapolate expectations of future income from realized current income. This is precisely the form of bias I document empirically in upcoming results on forecaster expectations of real disposable income. In recent work, De Silva and Mei (2024) document that household interest rate expectations tend to be close to forecaster expectations during periods where they make durables purchases.

Importantly, household and forecaster expectations have been documented to exhibit some systematic differences. Candia et al. (2020) and Kamdar and Ray (2023) find that household expectations overweight "supply-side" narratives, which emphasize the negative co-movement of real variables like real output and inflation, and underweight "demand-side" narratives. Andre et al. (2022) document the mental models households use to understand and form expectations of the economic effects of supply shocks, such as sudden changes in oil prices, are similar to those of forecasters but differ materially for monetary and fiscal policy shocks.

Using shocks which are interpreted in a systematically different way by households and forecasters would violate measurement error exogeneity as stated in **Assumption 6**. Using a supply shock to instrument forecaster expectations is the best way to address this concern given forecaster and household expectations exhibit qualitatively similar co-movements in response to these shocks. Therefore, I estimate model-implied impulse responses with respect to an oil supply news shock from Känzig (2021). Identified using a high-frequency identification approach, this shock captures variation in oil futures prices around a narrow time window of OPEC production announcements.

3.3 Empirical impulse response estimation

To estimate impulse responses of macroeconomic variables and their forecasts, I adopt the proxy structural vector autoregression (VAR) approach (Stock and Watson 2012, Mertens and Ravn 2013) and follow the empirical setup and notation from Känzig (2021) closely. I first estimate a reduced-form VAR, with a constant and a deterministic linear trend

$$\mathbf{Y}_t = \boldsymbol{\alpha} + \delta t + \sum_{l=1}^{p} \boldsymbol{\beta}_l \mathbf{Y}_{t-l} + \mathbf{u}_t$$

where $\boldsymbol{\alpha}, \delta t, \{\boldsymbol{Y}_{t-l}\}_{l=0}^{p}, \boldsymbol{u}_t$ are vectors of length n and $\boldsymbol{\beta}_l$ is a matrix of dimension $n \times n$.

I assume invertibility, in that the reduced-form residuals $\mathbf{u}_t$ are a linear combination of i.i.d structural shocks $\boldsymbol{\varepsilon}_t$

$$\boldsymbol{u}_t = \mathbf{S}\boldsymbol{\varepsilon}_t$$

where $\mathbb{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{\Omega}$, a positive, diagonal matrix.

Assuming an instrument z_t satisfies the standard identifying assumptions

$$\mathbb{E}[\varepsilon_{1,t} z_t] = \alpha \neq 0$$
$$\mathbb{E}[\boldsymbol{\varepsilon}_{2:n,t} z_t] = \mathbf{0}$$

where the structural shock we are identifying is ordered first in the VAR, without loss of generality. I can identify the first column of $\mathbf{S}$ up to sign and scale, which I denote $\mathbf{s}_1$, given by

$$\mathbf{s}_1 = \mathbb{E}[\mathbf{u}_t z_t]$$

Finally, to pin down the sign and scale factor $s_{1,1} := \frac{\mathbb{E}[u_{1,t}z_t]}{x}$ for the econometrician's desired value x, I can normalize the impact effect of the identified shock on variable $y_{1,t} = x$, using the re-scaled structural impact vector $\tilde{\boldsymbol{s}}_1 = \boldsymbol{s}_1/s_{1,1}$ provided $\mathbb{E}[u_{1,t}z_t] \neq 0$.

Specification The variables included in the baseline specification are real gross domestic product, real disposable income, the consumer price index (CPI), the nominal federal funds rate, real oil price and world oil production measures. The real oil price is the WTI crude oil price deflated by CPI inflation and will be the proxy SVAR's first-stage variable, scaled such that the on-impact effect of a positive oil supply news shock increases real oil prices by 10 percentage points. I then augment the baseline specification with real personal consumption expenditures and Bluechip expectations at each horizon-h period ahead one variable at a time. The data are measured at a quarterly frequency and in log-levels, aside from the federal funds rate. The time period spans 1985-Q1 through 2017-Q3, due to availability of the Bluechip data.

3.4 Impulse responses of income and interest rates to an oil shock

The upper panel of Figure 1 displays impulse responses of realizations and Bluechip expectations of real income and real interest rates. In response to an inflationary oil shock, realized real income and real interest rates (black lines) exhibit a prolonged decline.

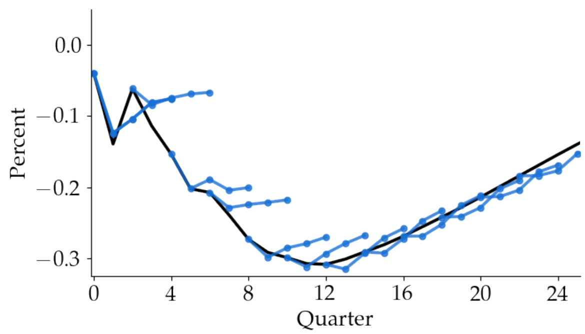
The lower panel of Figure 1 displays the impulse responses of CPI inflation and the nominal federal funds rate, the two variables used to construct our real interest rate measure. From these responses, we see that the majority of the decline in real interest rates is driven by elevated CPI inflation. The literature on the real effects of oil price shocks point to the central role of contractionary response of systematic monetary policy (Bernanke et al. 1997, Gagliardone and Gertler 2023). However, the response of the nominal federal funds rate in our limited sample period beginning in 1985 is relatively muted compared to these papers, whose sample periods typically extend back to the early 1970s.

The blue lines extending outward from the realized impulse responses of each variable represent the impulse response of expectations *across horizons* for a fixed impulse response period. This is a useful way to visualize the bias in expectations across the term structure, as shown in Bardóczy and Guerreiro (2023). To clarify interpretation, consider the definition of the impulse response across periods-ℓ of a horizon-h income expectation

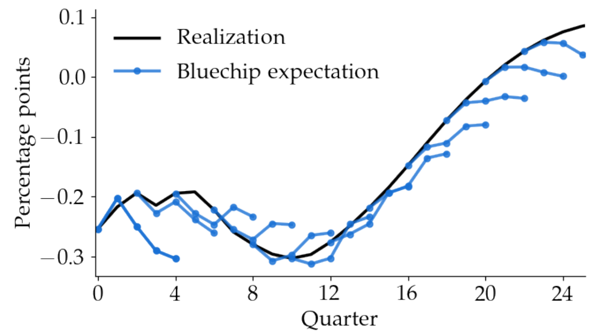
$$\Psi(E_t[Y_{t+h}], \varepsilon_{t-\ell}) := \mathbb{E}[E_t[Y_{t+h}]|\varepsilon_{t-\ell} = 1] - \mathbb{E}[E_t[Y_{t+h}]|\varepsilon_{t-\ell} = 0] \tag{15}$$

where $\mathbb{E}[\cdot]$ is the unconditional expectation across time-t and the index ℓ denotes the periods elapsed between the shock onset and the time-t that expectations are formed. The blue line is given by fixing the elapsed period-ℓ since the shock and plotting expectations across horizons-h. This construction is interpretable as the average response of expectations across all horizons-h formed ℓ-periods after an initial shock.

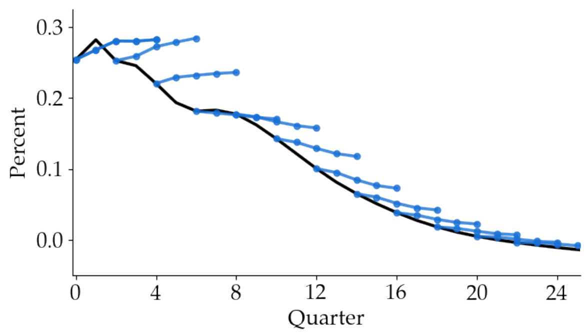
I proceed to analyze the systematic patterns of ex-post forecast errors in the expectations data shown in Figure 1 and consider whether they can be explained by existing models of expectation formation. To do so it is useful to consider the behavior of subjective expectations if they were formed under full-information rational expectations with respect to a hypothetical equilibrium economy. Let us define full-information to mean either $\varepsilon_{t-\ell}$ is directly public in-



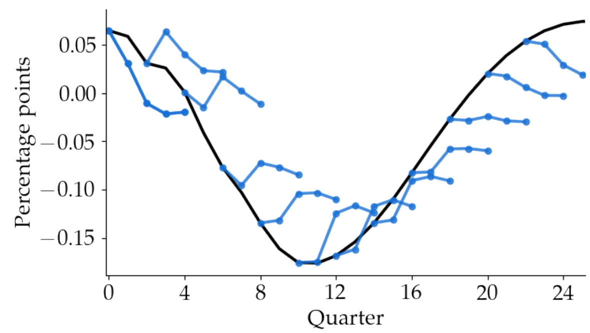
(a) Real Disposable Income



(b) Real Federal Funds Rate



(c) Consumer Price Inflation



(d) Nominal Federal Funds Rate

Figure 1: Realized and Bluechip impulse responses to a Känzig (2021) oil shock

Note: each panel contains an impulse response function of realizations (black) to a positive Känzig (2021) oil price news shock. Each expectation "hair" (blue) collects quarter-ℓ impulse response coefficients of Bluechip survey expectations for horizons $h = 1, ..., 4$ quarters ahead. The real federal funds rate is the nominal federal funds rate deflated by consumer price inflation.

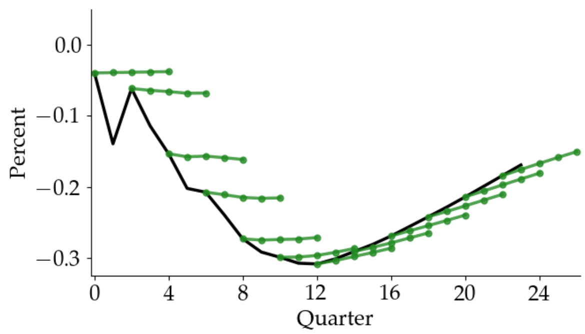
formation as of time-$t - \ell$, or that $\varepsilon_{t-\ell}$ is measurable in the set of publicly-observed time-$t - \ell$ variables in the economy. Assuming the equilibrium state variables are jointly stationary, the law of iterated expectations applies to Equation (15) and yields

$$\Psi(Y_{t+h}, \varepsilon_{t-\ell}) := \mathbb{E}[Y_{t+h}|\varepsilon_{t-\ell} = 1] - \mathbb{E}[Y_{t+h}|\varepsilon_{t-\ell} = 0], \quad \forall \ell, h \geq 0$$

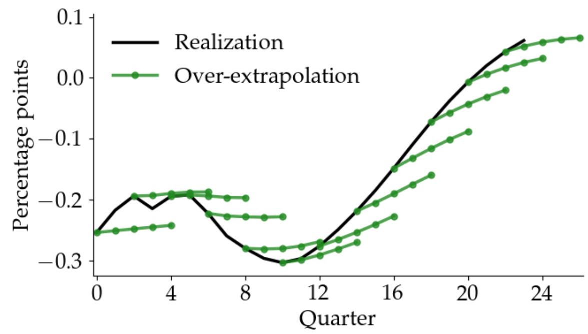
In words, the impulse response of the horizon-h expectation conditional on an ℓ-period past shock should equal the realized impulse response in the horizon period h. Given this, we can interpret the vertical gaps between the black and blue lines in Figure 1 as suggestive evidence that expectations deviate from the full-information rational expectations assumption.

Many theories of bounded rationality result in subjective expectations that systematically under- or over-react relative to the full-information rational expectations benchmark. Systematic under-reaction corresponds to expectation "hairs" (blue) that extend from the realization (black) always remaining closer to zero than the realization. Over-reaction conversely has the "hairs" remaining further from zero than the realization.

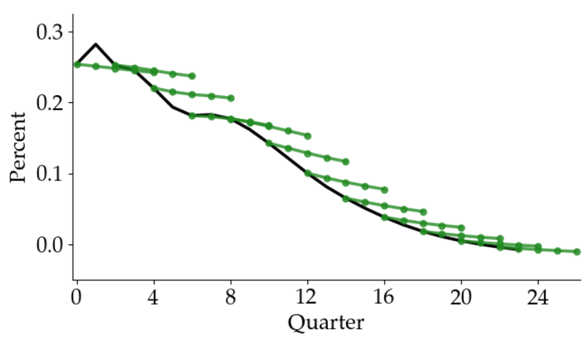
Theories of systematic under-reaction, such as sticky information (Mankiw and Reis 2002) or cognitive discounting (Gabaix 2020), are unable to explain the over-reaction of measured expectations to certain variables, such as the response of CPI inflation expectations in Figure 1d. Conversely, while theories of systematic over-reaction, such as diagnostic expectations (Bordalo et al. 2018), can explain the response of CPI inflation, they are unable to match the initial under-reaction of real disposable income expectations displayed in Figure 1a. Angeletos et al. (2021) and Bardóczy and Guerreiro (2023) convey this point as well, documenting similar patterns in impulse responses from the Survey of Professional Forecasters for other variables and shocks. I display the responses of a number of these other models and discuss them further in Appendix A.2.



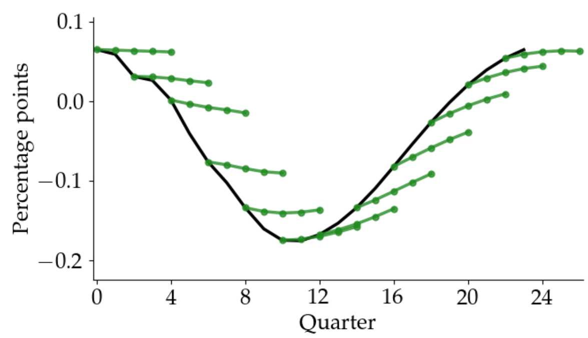
(a) Real Disposable Income



(b) Real Federal Funds Rate



(c) Consumer Price Inflation



(d) Nominal Federal Funds Rate

Figure 2: Realized and "over-extrapolation" model impulse responses to a Känzig (2021) oil shock

Note: each panel contains an impulse response function of realizations (black) to a positive Känzig (2021) oil price news shock. Each expectation "hair" (red) collects quarter-ℓ impulse response coefficients of the parametric "over-extrapolation" model given by Equation (16) for horizons $h = 1, ..., 4$ quarters ahead. The real federal funds rate is the nominal federal funds rate deflated by consumer price inflation.

While it is difficult to summarize these biases solely as systematic under- or over-reaction relative to the full-information rational expectation, they can however be rationalized as over-extrapolation of the most recent realization. I demonstrate this with an illustrative two-parameter model fit across all variables, time-ℓ and horizons-h. The model parameterizes the impulse response (IR) of subjective expectations $\Psi(E_t[Y_{t+h}], \varepsilon_{t-\ell})$, in the following way

$$\Psi(E_t[Y_{t+h}], \varepsilon_{t-\ell}) = \omega^{\ell}\theta^{h} \underbrace{\Psi(Y_t, \varepsilon_{t-\ell})}_{\text{Extrapolation IR}} + (1-\omega^{\ell}) \underbrace{\Psi(Y_{t+h}, \varepsilon_{t-\ell})}_{\text{Full-information rational expectation IR}} \tag{16}$$

The parameter θ determines the persistence of extrapolation from the current time-t observation across expectation horizons-h and ω determines how long this bias lasts across response periods-ℓ. The parameter $\omega \in (0, 1)$ ensures that extrapolation bias will eventually diminish and expectations will converge back to the rational benchmark as time-ℓ progresses.

Figure 2 displays the impulse response of expectations implied by this simple over-extrapolation model, which replicates Figure 1 quite well. This observation motivates my later adoption of a Bayesian learning model that exhibits a similar form of over-extrapolation bias.

In addition to the fitted model producing Figure 2, I consider a battery of other simple parametric models fit to the same data to extrapolate missing horizons. One could view the use of these models to extrapolate missing horizons as shape restrictions on expectational impulse responses across horizons-h, as the simple parametric model I fit demonstrates. As a baseline, I use an estimated AR(2) process, constrained to be stationary, to extrapolate missing horizons. The results in the following section on consumption function estimation are robust to alternate choices. Details for the choice of auxiliary models for extrapolation, their estimation, and resulting structural parameter estimates are in Appendix A.3.

3.5 Empirical vs. model-implied impulse responses of consumption

The expectation impulse responses plotted in Figure 1 correspond to the impulse response estimands reported in the model-implied consumption moment condition (11). Figure 3 displays the model-implied impulse responses of consumption from the estimated representative-agent, perpetual youth overlapping generations, and standard incomplete markets models. Recall these models only have one or two structural parameters to be used as degrees of freedom for estimation across numerous impulse response periods. The elasticity of intertemporal substitution σ is estimated across all models, and the hazard rate ω, which allows for heightened income sensitivity of consumption, is additionally estimated for the perpetual youth model.

My baseline estimates for the elasticity of intertemporal substitution (EIS) across models are low. For the representative-agent benchmark the estimated EIS approaches zero, while for the heterogeneous-agent models the estimated EIS is around 0.1. The main intuition stems from observing in Figure 1b that realized and expected real interest rates decline in response to the oil price shock. The estimated EIS is pushed downward to mitigate the positive intertemporal substitution response of model-implied consumption and amplify the negative income effects from lower rates.

One key reason why the estimated EIS is low across all models is that the magnitude of the observed decline in consumption to this shock exceeds that of disposable income. Hence, for model-implied consumption to match the observed decline, negative income effects from interest rates must be sufficiently large.

While an EIS estimate of 0.1 is low it is not unprecedented. In a quasi-experimental setting, Best et al. (2020) exploit borrower bunching behavior around loan-to-value thresholds used to price mortgages and also find an estimated EIS of 0.1. Likewise, Ring (2024) finds evidence for a

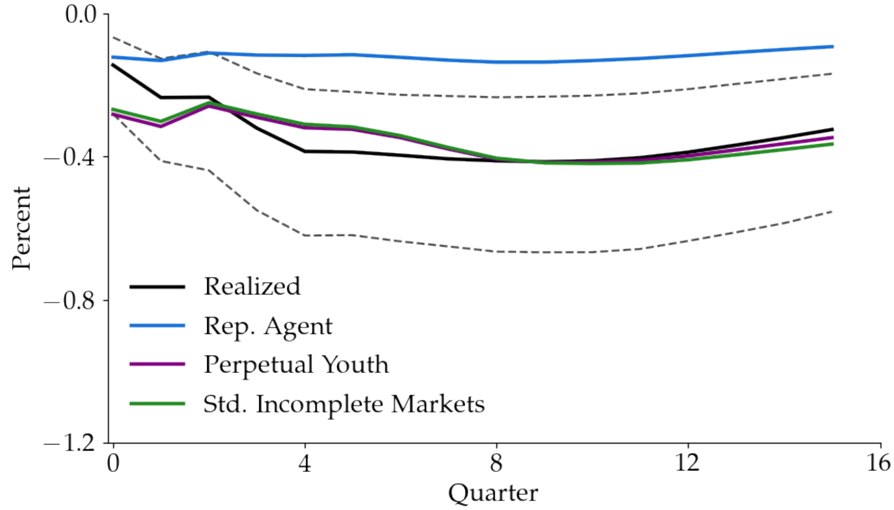


Figure 3: Estimated model-implied consumption impulse responses to a Känzig (2021) oil shock

Note: the realized (black) consumption impulse response and model-implied consumption impulse responses to a positive Känzig (2021) oil price news shock. Model-implied responses are produced by evaluating each models consumption function using empirical impulse responses of realized and expected income and interest rates. The dashed lines are 68% confidence bands produced using moving block bootstrap by Jentsch and Lunsford (2019).

similarly low EIS using Norwegian administrative data and geographic variation to investigate the relative size of substitution and income effects of wealth taxation on savings behavior.

Recall that most of the decline in the real interest rate from the oil price shock is due to the increase in realized and expected inflation. One concern might be that forecasters' inflation expectations respond differently than household expectations due to this shock. Figure 7 of Känzig (2021) provides suggestive evidence that the magnitude of households' inflation expectation responses to oil shocks may be larger than that of forecasters'. However, this would imply an even lower expected real rate taking account of household expectations reinforcing the need for a low estimated EIS.

The income sensitivity of consumption, as measured by the current marginal propensity to consume (MPC) out of unearned income, is small in the representative-agent model by construction. In contrast, heterogeneous-agent models can have substantially higher MPCs, and indeed I find this to be the case in this estimation. The estimated hazard rate for the perpetual youth model implies an MPC of approximately five percent at a quarterly frequency. While the standard incomplete markets model did not have an independent degree of freedom from the EIS to estimate, due to the discount factor being used to target the steady-state level of assets, its MPC nonetheless matches that of the perpetual youth model at five percent. As Figure 3 demonstrates, the higher MPC in the heterogeneous-agent models proves crucial to match the pronounced consumption contraction due to the oil shock. Due to the much longer effective horizons for income smoothing, the representative-agent models' response to the shock is less severe.

Consumption-Savings Models

Parameter	Perpetual Youth	Standard Incomplete Markets	Rep. agent
EIS	0.08	0.09	0.00
MPC	0.04	0.05	0.005
$\frac{\text{Assets}}{\text{Income}}$	1.4	1.4*	1.4
EIS	0.80	0.05	-
MPC	0.2*	0.2*	-
$\frac{\text{Assets}}{\text{Income}}$	1.4	0.425	-

Table 3: Estimated/targeted parameters from consumption-savings models

Note: The top panel contains estimated parameters enforcing that the steady state assets-to-income ratio is equal to the initial calibration target. The bottom panel contains estimated parameters when models instead target a higher marginal propensity to consume (MPC). The EIS σ is the elasticity of intertemporal substitution. MPC and income are reported at a quarterly frequency.

MPCs in micro-calibration versus macro-estimation It is well-known that the standard incomplete markets model is unable to simultaneously match typical microeconomic estimates of the current MPC and the steady state level of household assets (Kaplan and Violante 2022). By restricting the estimated model to match the latter, I attain an implied MPC of around 0.05 at a quarterly frequency, which is lower than typical microeconomic estimates which range from 0.15 to 0.25. However, this MPC that is consistent with our targeted macroeconomic impulse responses and has been shown to be consistent with a broader range of macroeconomic moments in full-information HANK estimation on macroeconomic time series Bayer et al. (2024).

I consider how calibrating the MPC in both the perpetual youth and standard incomplete markets models to 0.2, in line with microeconomic evidence, affects their implied impulse response fit. In the standard incomplete markets model I calibrate the discount factor β to now match the MPC target. While the fit deteriorates, as shown in Figure 4, they both still remain within a one standard deviation bound of the empirical impulse response of consumption. However, the estimated parameters for the EIS now diverge between these models. The estimated EIS is an order of magnitude larger in the perpetual youth model, while in the standard incomplete markets model it is slightly lower. In addition, the perpetual youth model now overshoots the empirical response, whereas the standard incomplete markets model undershoots.

To explain the reason for this change, consider an important difference between these two models: given the linearization in aggregates, the perpetual youth model lacks a precautionary savings motive. Recall in the perpetual youth model that the steady state level of assets is independent of the MPC. Therefore, changing the EIS only scales the relative size of the substitution versus income effects in response to the discounted value of expected interest rate changes, as shown in Equation (13). Given the large, prolonged decline in realized and expected real interest rates in response to the shock shown in Figure 1, higher MPCs in the perpetual youth model at the original, lower EIS estimate would have excessively amplified the negative income effect from lower rates.

The standard incomplete markets model requires a lower discount factor to attain a high MPC, which in turn reduces steady state asset demand because agents are less patient. Whether

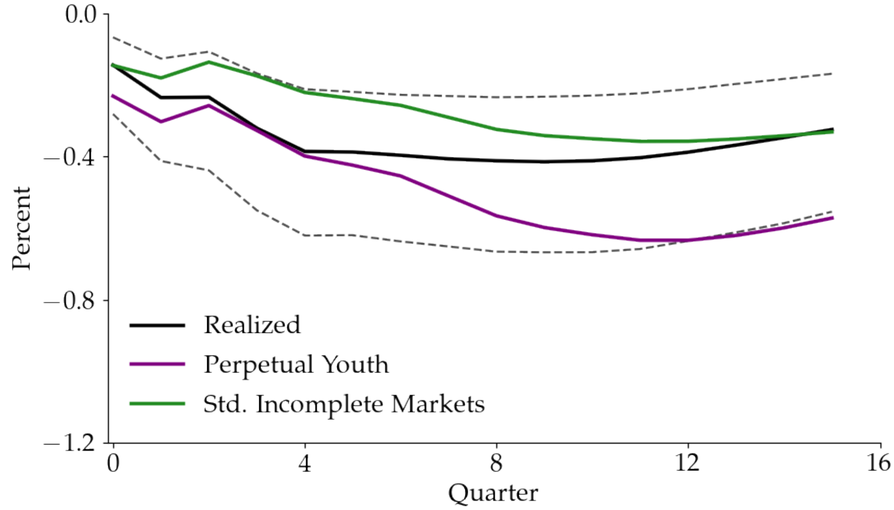


Figure 4: Calibrated model-implied consumption impulse responses to a Känzig (2021) oil shock

Note: the realized (black) consumption impulse response and model-implied consumption impulse responses to a positive Känzig (2021) oil price news shock. Model-implied responses are produced by evaluating each models consumption function using empirical impulse responses of realized and expected income and interest rates. The dashed lines are 68% confidence bands produced using moving block bootstrap by Jentsch and Lunsford (2019).

the magnitude of interest rate income effects increases due to the higher MPC or decreases due to the lower stock of steady state assets is a quantitative question. In this case, the lower discount factor reduces the magnitude of the negative interest rate income effect, requiring an even lower EIS. Because I directly use the canonical standard incomplete markets model, I cannot resolve the fundamental tension between these parameter calibrations. Nonetheless, I show Figures 3 and 4 that conditioning directly on expectations data, the models are similarly able to rationalize the observed inertia in aggregate consumption.

Full-information, rational expectations comparison It is natural to consider how model-implied consumption responses may differ in comparing those using expectations data with those formed via full-information, rational expectations (FIRE). However, without the complete specification of an equilibrium model we are not able to consider this counterfactual because of the Lucas critique. In Section 3.4, we could directly assess the validity of the FIRE assumption by making relative comparisons of average expectations and realizations. Here, however, we are unable to consider the counterfactual implications of full-information rational expectations versus data-based expectations without understanding how it changes the data-generating process of realizations themselves. Given this, I proceed by adding structure in the form of a model of expectation formation informed by the data and an equilibrium environment to understand the joint determinants of observed expectations biases and consumption inertia.

4 An Equilibrium Macroeconomic Model with Inertia

In this section, I develop a theory for how inertia arises endogenously within a heterogeneous-agent New Keynesian (HANK) equilibrium model. I begin by defining the temporary equilibrium, given subjective expectations in a standard HANK environment and proceed to adopt a form of frictional Bayesian learning as my model of expectation formation. I demonstrate how the interactions between learning frictions and the core features of HANK — high income sensitivity and low interest rate sensitivity of consumption demand — play a key role in generating macroeconomic inertia.

4.1 Temporary equilibrium definition

I begin by defining a temporary equilibrium, an intermediate step toward a fully-specified general equilibrium that does not yet place restrictions on how forward-looking agents form expectations. As in Woodford (2013), I immediately resort to using the linearized equilibrium[12], whose deviations are given by time-indexed variables, e.g. C_t, around a non-stochastic steady state, whose notation is given by non-time-indexed variables, e.g. C.

The equilibrium environment closely follows Angeletos et al. (2023), although I simplify along a few dimensions that are not central to my analysis. I briefly discuss the shared equilibrium ingredients, such as the firm problem, policy rules, and market clearing conditions and elaborate only on my points of departure.

Households and firms I use the Angeletos et al. (2023) specification of the perpetual youth overlapping generations model, resulting in the earlier-estimated consumption function in Equation (13).

Labor unions intermediate labor markets, ensuring households supply an identical quantity of labor and equalizing the real wage and the average marginal rate of substitution between aggregate consumption and labor supply. Households therefore receive the identical labor income. Firm production follows the textbook New Keynesian model (Galí 2015), where identical monopolistically-competitive firms operate a linear-in-labor production technology and face Calvo price-setting frictions. This gives rise to an aggregate price inflation New Keynesian Phillips curve linearized around a zero-inflation steady state.

$$\pi_t = \kappa Y_t + \beta E_t[\pi_{t+1}] \tag{17}$$

Firms distribute dividends identically, ensuring households receive the same profit income.

Policy and market clearing The real interest rate is determined by the real Taylor rule

$$i_t - E_t[\pi_{t+1}] \equiv r_t = \phi Y_t \tag{18}$$

The monetary authority sets nominal interest rates accounting for the equilibrium consequences on subjective inflation expectations $E_t[\pi_{t+1}]$ to achieve a real interest rate target of ϕY_t. A rule

[12] The consumption functions written earlier are in level as opposed to log deviations. To maintain this notation, I normalize steady state output $Y = 1$ such that level and log deviations for the below-defined variables can be interpreted identically.

of this form allows the monetary authority to conduct policy as if it maintained direct control of the ex-ante real interest rate. Adopting this rule therefore allows us to focus on the equilibrium determination of household consumption as a function of real interest rates without needing to separately account for the dynamics of subjective inflation expectations.

Market clearing in the goods market is given by

$$C_t = Y_t \tag{19}$$

Definition 2. A (linearized) **temporary equilibrium** consists of sequences of prices $\{i_t, \pi_t\}$ and quantities C_t, Y_t that satisfy (13), (17), (18), (19) for all periods t, given subjective expectations of $\{E_t[Y_{t+h}], E_t[\pi_{t+h}], E_t[i_{t+h}]\}_{h>0}$.

Assets are in zero net supply so households hold zero wealth in equilibrium.

4.2 General equilibrium dynamics with Bayesian learning

Closing the temporary equilibrium defined in Section 4.1, I assume households form expectations with a particular class of Bayesian learning models. These models have been shown to be consistent with multiple dimensions of evidence on expectation formation, including in the cross-section of households (Nagel 2024), experiments (Afrouzi et al. 2023), and unconditional time-series (Farmer et al. 2024, Crump et al. 2023).

In these models, Bayesian agents observe variables such as output or interest rates, understanding them to be driven by unobserved shocks that each are the sum of a transitory and a persistent component[13]. Because they cannot immediately distinguish these components, they gradually update their beliefs about them by solving a filtering problem given the time series of observed variables.

I assume that the only exogenous shock in the baseline economy is a demand shock ε_t, which is the sum of two unobserved mean-zero AR(1) components

$$\varepsilon_t = \lambda_t + \eta_t$$

with persistence parameters $\rho_\lambda > \rho_\eta$ and mean-zero Gaussian i.i.d innovations with variances $\sigma_\lambda^2, \sigma_\eta^2$. I refer to λ_t as the "persistent" component, due to its higher persistence, and η_t as the "transitory" component. I assume household have common knowledge of the functional form of the demand shock and the values of its parameters.

What remains to be specified is how agents draw inference about the underlying shock components λ_t, η_t from observing variables Y_t, π_t, i_t. Define the perceived laws of motion of Y_t, π_t, i_t as a set of stochastic variables $\tilde{Y}_t, \tilde{\pi}_t, \tilde{i}_t$, which are also linear functions of the histories of the unobserved components $\{\lambda_{t-\ell}, \eta_{t-\ell}\}_{\ell \geq 0}$. However, I permit the perceived laws to differ from the actual laws of motion, that is, their function coefficients[14] on their component histories $\{\lambda_{t-\ell}, \eta_{t-\ell}\}_{\ell \geq 0}$ may differ.

[13]Some of the cited literature adopt the convention that the persistent component is a non-stochastic, long-run mean parameter, which agents are nonetheless uncertain about. The inference problem is similar to the case I study here and results in similar forms of over-extrapolation that lie at the core of my analysis.

[14]I do not consider subjective uncertainty over and learning of the coefficients of the perceived law of motion itself, which is the subject of a separate literature on equilibrium learning. I assume the coefficients of the perceived law of motion are constant.

Subjective expectations are evaluated with respect to the perceived laws of motion. For example, if $\tilde{Y}_t = \lambda_t + \eta_t$ then expected one-period ahead output is given by

$$E_t[Y_{t+1}] = E_t[\lambda_{t+1}] + E_t[\eta_{t+1}]$$

Agents' beliefs can be represented by the following state-space model with the AR(1) coefficients collected in the diagonal matrix $\mathbf{F}$ and the perceived law of motion coefficients collected in the lag[15] polynomial matrix $\tilde{\mathbf{A}}(L)$

$$\begin{bmatrix} \lambda_{t+1} \\ \eta_{t+1} \end{bmatrix} = \mathbf{F} \begin{bmatrix} \lambda_t \\ \eta_t \end{bmatrix} + \begin{bmatrix} u_{\lambda,t+1} \\ u_{\eta,t+1} \end{bmatrix}, \qquad \begin{bmatrix} \tilde{Y}_t \\ \tilde{\pi}_t \\ \tilde{i}_t \end{bmatrix} = \tilde{\mathbf{A}}(L) \begin{bmatrix} \lambda_t \\ \eta_t \end{bmatrix}$$

Due to linearity and normality of innovations, the subjectively optimal state estimates, conditional on the perceived state-space model and past observations, are given by the Kalman update equation

$$\begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \mathbf{F} \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix} + \mathbf{G} \left(\begin{bmatrix} Y_t - E_{t-1}[Y_t] \\ \pi_t - E_{t-1}[\pi_t] \\ i_t - E_{t-1}[i_t] \end{bmatrix} \right)$$

where $\mathbf{G}$ is the steady-state Kalman gain implied by the agents' perceived state-space model.

I make the typical learning assumption that the information set for time-t decisions determining equilibrium Y_t, π_t, i_t is the history of past observables $\{Y_{t-\ell}, \pi_{t-\ell}, i_{t-\ell}\}_{\ell \geq 1}$. To reflect this staggered timing, subjective expectations that inform time-t decisions are labeled E_{t-1}.

If the perceived and actual laws of motion differ, $E_{t-1}[Y_t]$ will a sub-optimal forecast for Y_t, even though it is perceived to be optimal. Distorted component expectations $E_{t-1}[\lambda_t], E_{t-1}[\eta_t]$ are reinforced over time as decisions based on them determine future observables Y_t, π_t, i_t that are used for future inference to update component expectations $E_t[\lambda_{t+1}], E_t[\eta_{t+1}]$ and so on.

Definition 3. A **learning equilibrium** is a temporary equilibrium and a collection of subjective expectations $\{E_t[Y_{t+h}], E_t[\pi_{t+h}], E_t[i_{t+h}]\}_{h>0}$ induced by a perceived law of motion $\tilde{Y}_t, \tilde{\pi}_t, \tilde{i}_t$ and past observations $\{Y_{t-\ell}, \pi_{t-\ell}, i_{t-\ell}\}_{\ell>1}$.

To greatly simplify the model I proceed by assuming the perceived law of motion $\tilde{Y}_t, \tilde{\pi}_t, \tilde{i}_t$ is consistent with the policy rule (18). This allows us to isolate the equilibrium determination of output, which will be the focus of the remainder of the paper, from inflation and nominal interest rates. Consequently because it is commonly understood that these variables are determined by output, they provide no additional information about the components λ_t, η_t. Hence, we can treat realized output Y_t as the only observable variable agents learn from.

Consolidating equilibrium conditions into a simple aggregate demand equation, I obtain

$$Y_t \propto (1 - \beta\omega - \beta\omega\sigma\phi) \sum_{h=1}^{\infty} (\beta\omega)^h E_t[Y_{t+h}] + \varepsilon_t$$

[15]If there are lag terms, then the backward-looking expectations of lagged components, e.g. $E_{t-1}[\lambda_{t-\ell}]$ for $\ell > 1$, can be computed using the Kalman smoother. This contrasts with the simple case $\tilde{Y}_t = \lambda_t + \eta_t$ discussed above, where forward-looking component expectations, e.g. $E_t[\lambda_{t+1}]$, can be computed using the Kalman filter.

Note that Y_t is not exactly equal to the right-hand side because of the within-period general equilibrium feedback of Y_t, r_t. The constant of proportionality that I have omitted, $\beta\omega(1+\sigma\phi)$, only affects the overall level of Y_t and not the shape of its impulse response to ε_t across periods. In the following sub-sections I focus on characterizing the shape and not the overall level of impulse responses. Therefore, without loss of generality, I normalize the variance of ε_t and proceed using Equation (20) with equality. When considering policy counterfactuals in Section 5 I undo this normalization to ensure the level contribution of counterfactuals is properly accounted for.

Let $\chi := (1-\beta\omega-\beta\omega\sigma\phi)$ and permanent income $\mathcal{Y}_t := \sum_{h=1}^{\infty}(\beta\omega)^h E_t[Y_{t+h}]$. Re-writing the aggregate demand equation we obtain the following simple expression for equilibrium output.

$$Y_t = \chi\mathcal{Y}_t + \varepsilon_t \tag{20}$$

Informally, we can see that if ε_t does not exhibit inertia in the form of hump-shaped impulse responses then for Y_t to exhibit inertia $\mathcal{Y}_t$ must exhibit inertia and χ must be sufficiently large such that Y_t inherits its inertia. Given $\mathcal{Y}_t$ is the variable that summarizes the effect of future output beliefs on current realized output, I will refer to χ as the *belief multiplier*. To formalize this intuition, I proceed by considering concrete examples of perceived laws of motion to demonstrate how output inertia Y_t arises endogenously due to the effects of the shock on the evolution of beliefs.

4.3 Simple learning

As before, suppose households' perceived law of motion of output is given by the simple form

$$\tilde{Y}_t = \lambda_t + \eta_t \tag{21}$$

This perceived law of motion means households completely disregard the general equilibrium feedback of the shock when forming expectations of future output. Whenever they observe a given output realization, they simply assume it was due to direct changes in the underlying shock components. While this assumption is stark, it is useful to illustrate the consequences of perceived laws of motion which do not fully account for general equilibrium feedback. The kinds of distortions these mistaken beliefs impart on realized output dynamics will carry over to more sophisticated but still imperfect beliefs and will serve as a useful example to compare to the rational learning benchmark.

With this perceived law of motion, permanent income takes the form

$$\mathcal{Y}_t = \frac{\beta\omega\rho_\lambda}{1-\beta\omega\rho_\lambda}E_{t-1}[\lambda_t] + \frac{\beta\omega\rho_\eta}{1-\beta\omega\rho_\eta}E_{t-1}[\eta_t]$$

With $\rho_\lambda > \rho_\eta$ the same-sized belief update of $E_{t-1}[\lambda_t]$ raises expected future income $\mathcal{Y}_t$ by more than a comparable change in $E_{t-1}[\eta_t]$ because it corresponds to the belief that future income $\{E_t[Y_{t+h}]\}_{h>1}$ will be persistently higher. I denote the effective horizon of the persistent component belief as $h_\lambda := \frac{\beta\omega\rho_\lambda}{1-\beta\omega\rho_\lambda}$ and likewise for the transitory component $h_\eta := \frac{\beta\omega\rho_\eta}{1-\beta\omega\rho_\eta}$.

The equilibrium dynamics of output can be represented as a system of two equations. The first is the aggregate demand equation for output derived from Equation (20) given component

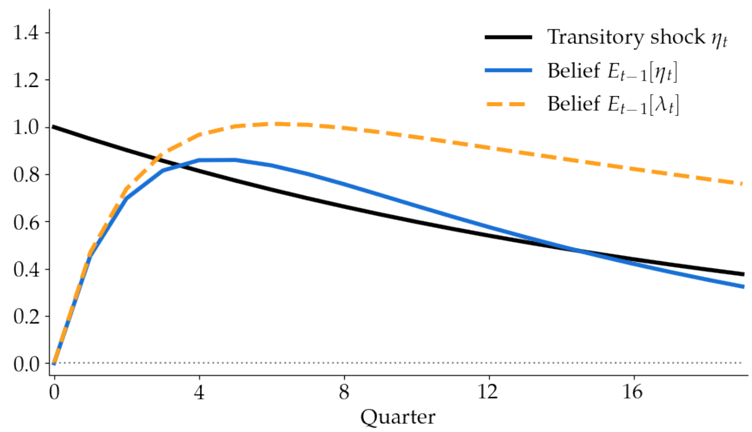


Figure 5: Component belief impulse responses to a transitory demand shock η_t

expectations. The second is the law of motion of component expectations, as a function of the output forecast error $Y_t - E_{t-1}[Y_t]$.

$$Y_t = \underbrace{\chi \mathbf{h} E_{t-1}[\boldsymbol{\varepsilon}_t]}_{\text{Belief feedback}} + \underbrace{\mathbf{1}\boldsymbol{\varepsilon}_t}_{\text{Direct shock effect}} \tag{22}$$

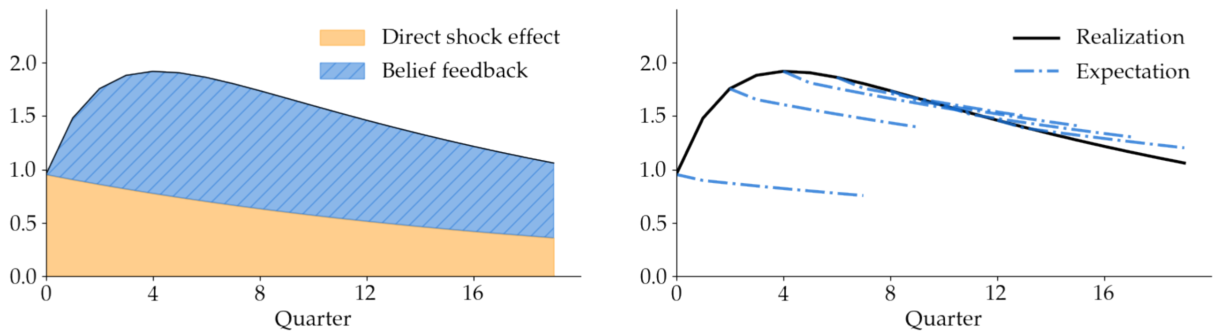
$$E_t[\boldsymbol{\varepsilon}_{t+1}] = \mathbf{F} E_{t-1}[\boldsymbol{\varepsilon}_t] + \mathbf{g}'\big(\underbrace{\chi \mathbf{h} E_{t-1}[\boldsymbol{\varepsilon}_t] + \mathbf{1}(\boldsymbol{\varepsilon}_t - E_{t-1}[\boldsymbol{\varepsilon}_t])}_{\text{Forecast error } Y_t - E_{t-1}[Y_t]}\big) \tag{23}$$

(Row) vector notation denotes the shock components $\boldsymbol{\varepsilon}_t := (\lambda_t, \eta_t)$, effective horizons $\mathbf{h} := (h_\lambda, h_\eta)$, $\mathbf{1} := (1, 1)$, autoregression persistences $\mathbf{F} := \text{diag}(\rho_\lambda, \rho_\eta)$, and steady-state Kalman gains $\mathbf{g} := (g_\lambda, g_\eta)$.

Figure 5 shows the response of beliefs about the transitory and persistent components to a transitory shock. Even though no persistent shock occurred, beliefs about both components increase because observing output does not allow households to perfectly distinguish the components in the early onset of a shock. Due to gradual learning, we see the adjustment of component beliefs exhibit inertia in response to the shock.

A distinctive and important consequence of this simple perceived law of motion is the appearance of the belief feedback term $\chi \mathbf{h} E_{t-1}[\boldsymbol{\varepsilon}_t]$ in the evolution of component beliefs in Equation (23). This contrasts with the rational learning Kalman update that I consider in the next sub-section, where the output forecast error $Y_t - E_{t-1}[Y_t]$ is simply equal to the current shock relative to beliefs $\mathbf{1}(\boldsymbol{\varepsilon}_t - E_{t-1}[\boldsymbol{\varepsilon}_t])$. This belief feedback wedge appears precisely because households are unaware that equilibrium output is determined in part by their consumption decisions and not just the direct effect of the shock. By consuming based on their distorted beliefs they alter equilibrium output, and by incorrectly updating after observing equilibrium output their beliefs become further distorted. I will refer to this self-reinforcing feedback loop as *expectations unanchoring*.

A high degree of belief feedback into output, relative to the direct effect of a shock, and component beliefs which unanchor and reinforce this feedback over time cause output Y_t to exhibit inertia as shown in Figure 6a. Because the initial shock was transitory but beliefs also



(a) Output Decomposition

(b) Output Expectation Hair Plot

Figure 6: Output impulse response to a transitory demand shock η_t

Note: The left panel decomposes the output impulse response to a transitory demand shock into the direct effect of the shock (gold solid) and the belief feedback effect (blue hatched) in Equation (22). The right panel expectation "hairs" (blue dot-dash) collect impulse response coefficients of quarter-ℓ subjective output expectations.

update toward the persistent component, output expectations are overly persistent, exhibiting systematic forecast errors which diminish gradually with time. The combination of these features permit this model to jointly explain output inertia (equivalently consumption and income in this setting) and expectations that persistently over-extrapolate from the most recent observation, as documented in Section 3 albeit to a supply shock.

Inertia and the belief multiplier χ After displaying an example of inertia arising in the simple perceived law of motion above, I now formalize the intuition for why a high belief multiplier χ induces greater output Y_t inertia. This happens both because the multiplier χ scales the relative contribution of inertial component beliefs embedded in expected permanent income $\mathcal{Y}_t$ and because it scales the belief feedback wedge that appears in the component belief law of motion, increasing the persistence of beliefs.

Definition 4. Let $X_t = \sum_{\ell=0}^{\infty}(a_\ell \lambda_{t-\ell} + b_\ell \eta_{t-\ell})$ denote the Wold representation of a variable X_t. X_t exhibits **inertia** with respect to a component shock $\lambda_{t-\ell}$ if its corresponding Wold coefficients $\{a_\ell\}$ are weakly increasing (decreasing) for $\ell \leq \bar{\ell} > 0$ and weakly decreasing (increasing) for $\ell > \bar{\ell}$[16]. Denote the impulse response period $\ell = \bar{\ell}$ as the **inertial peak**. These definitions hold symmetrically for component shock $\eta_{t-\ell}$ and its coefficients $\{b_\ell\}$.

The above definition essentially states that a variable is inertial with respect to a shock if its maximal impulse response period, which I call the inertial peak, is not the initial shock period.

Proposition 1. *For each component shock $e_t \in \{\lambda_t, \eta_t\}$ if $\chi > \underline{X}_e$, then Y_t will exhibit inertia and its inertial peak period $\bar{\ell}$ will be weakly increasing in χ.*
Proof in Appendix B.1.

Proposition 1 demonstrates the tight connection between output inertia and the belief multiplier χ, where inertia results when the multiplier is large as demonstrated in Figure 7.

[16]Given the system (22), (23), I impose regularity conditions on model parameters in Appendix B.1 that prevent the impulse responses of output and component beliefs from exhibiting oscillation, such that this definition applies.

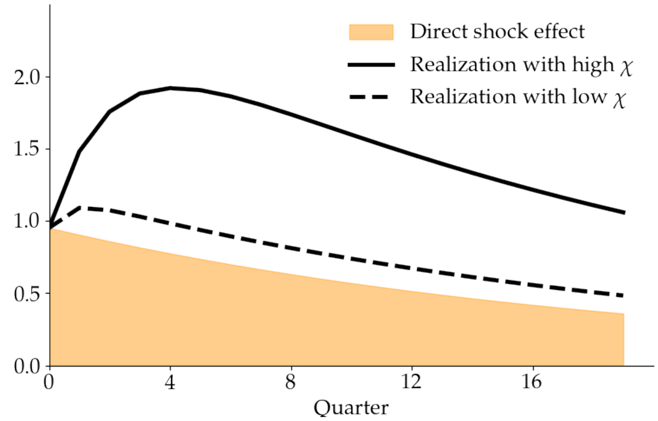


Figure 7: Output impulse responses to an η_t shock under different belief multipliers χ

Note: The black solid and dashed lines represent the output impulse response under high and low belief multiplier calibrations. The direct shock effect (gold solid) is the same under both calibrations. The gap between the direct shock effect and the black solid and dashed lines represent the size of expectation feedback under each calibration.

The intuition of the lower threshold for χ in Proposition 1 is the following. For output to exhibit inertia at all, the belief multiplier χ must be sufficiently large for the additional endogenous persistence in output Y_t contributed by the belief feedback term to exceed the exogenous decay of the direct shock effect. In other words, output inertia results when belief feedback is sufficiently amplified by χ. By increasing the multiplier χ and thus increasing the persistence of the component belief system, expectation feedback will be relatively longer lasting and contribute more to output inertia. Figure 8 demonstrates the responses of persistent and transitory component beliefs to a transitory shock in economies with different-sized belief multipliers χ.

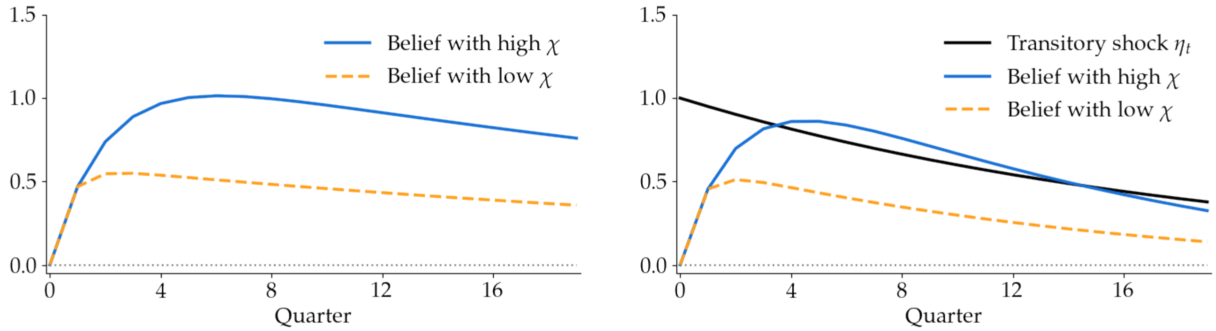
While in the initial period the criteria for inertia to arise simply requires the response of belief feedback to exceed the exogenous decay of the direct shock effect, in later periods the relevant comparison is whether the endogenous persistence of the persistent belief component $E_{t-1}[\lambda_t]$ exceeds the combined decay from the endogenous transitory belief component $E_{t-1}[\eta_t]$ and the exogenous decay of the transitory shock η_t itself.

Recall the form of the belief multiplier χ from the perpetual youth model

$$\chi := (1 - \beta\omega - \beta\omega\sigma\phi) \equiv (\text{MPC} - (1 - \text{MPC})\text{EIS}\phi)$$

It is useful to re-write the parameter values in terms of more interpretable quantities, namely the current marginal propensity to consume (MPC) and the elasticity of intertemporal substitution (EIS). We see that the multiplier χ is larger when MPC is high, which is the distinctive difference between heterogeneous-agent and representative-agent models of consumption, and when the EIS is low.

Figure 9 plots the inertial peak responses of output and the component beliefs to a transitory η_t component shock. As demonstrated in Proposition 1, output becomes more inertial when the belief multiplier is larger, in part because beliefs contribute relatively more to output than the direct shock and because beliefs themselves become more inertial. The only place the EIS enters the equilibrium output law of motion is through the belief multiplier. Thus, the left



(a) Persistent component belief $E_{t-1}[\lambda_t]$

(b) Transitory component belief $E_{t-1}[\eta_t]$

Figure 8: Component belief impulse responses to an η_t shock under different belief multipliers χ

Note: The blue solid lines represent the impulse response of component beliefs to a transitory demand shock under a high belief multiplier calibration. The gold dashed line represent the analogous responses for a low belief multiplier calibration.

panel of Figure 9 can be interpreted equivalently as how inertia changes with the EIS, where a lower EIS implies a higher multiplier and consequently more inertia. The MPC alters both the belief multiplier and the effective horizons of beliefs **h**, where a higher MPC results in a higher belief multiplier but lower effective horizons. However, as the right panel of Figure 9 demonstrates, the net effect of a higher MPC still results in more inertia.

4.4 Rational and constrained-rational learning

This sub-section contrasts the simple perceived law of motion with more sophisticated beliefs and demonstrates that inertia may be absent in certain cases of learning and present in others. The first case I consider is rational learning, where the perceived and actual laws of motion coincide. With rational learning households are able to account for the equilibrium impacts of their decisions on their own expectations and hence optimally incorporate past observations of output into their component forecasts. The rational learning equilibrium solution yields

$$Y_t \equiv \tilde{Y}_t = \left(\frac{\chi h_\lambda}{1-\chi h_\lambda} E_{t-1}[\lambda_t] + \frac{\chi h_\eta}{1-\chi h_\eta} E_{t-1}[\eta_t]\right) + \lambda_t + \eta_t \tag{24}$$

The second case, which I call "constrained-rational" learning, still restricts households' beliefs to be functions of the contemporaneous components λ_t, η_t but permits their coefficients to be optimally estimated given their economic environment. This allows me to later consider policy counterfactuals that are robust to the Lucas critique within this class of models with imperfect learning, while still retaining similar limitations and implications as simple learning. Constrained-rational learning takes the below functional form

$$\tilde{Y}_t = \tilde{a}\lambda_t + \tilde{b}\eta_t$$

Given the Wold decomposition of realized equilibrium output $Y_t = \sum_{\ell=0}^{\infty}(a_\ell u_{\lambda,t-\ell} + b_\ell u_{\eta,t-\ell})$, we can solve for the optimal coefficients $\tilde{a}, \tilde{b}$ by projecting the constrained-rational perceived

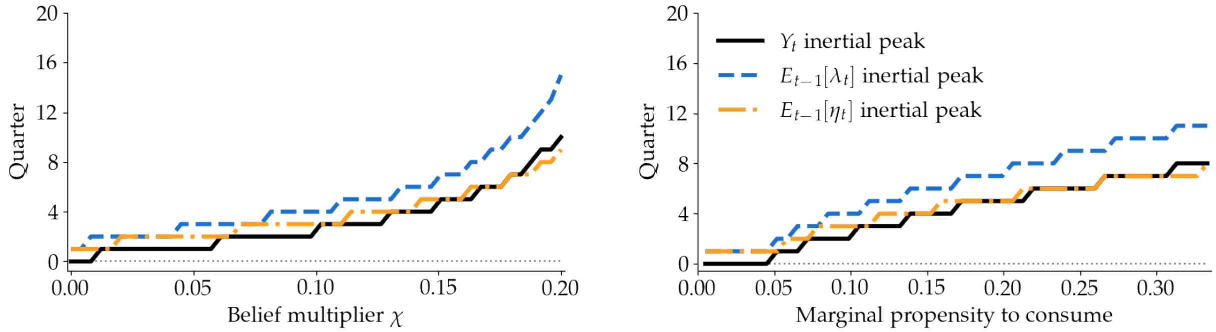


Figure 9: Inertial peaks of output and component belief impulse responses to a transitory η_t shock

Note: Each line represents the inertial peak period, defined in Definition 4, for output (black), the persistent component belief (dashed blue), and the transitory component belief (dot-dashed gold) to a transitory shock. The left panel plots the inertial peaks across values of the belief multiplier, holding the effective horizons fixed. The right panel plots the inertial peaks across values of the current marginal propensity to consume, which also alters the effective horizons.

law of motion $\tilde{Y}_t$ onto the actual one Y_t. Doing so obtains the following pair of implicit equations, which define $\tilde{a}, \tilde{b}$.

$$\tilde{a} = (1-\rho_\lambda^2)\sum_{\ell=0}^{\infty}\rho_\lambda^\ell a_\ell(\tilde{a},\tilde{b})$$
$$\tilde{b} = (1-\rho_\eta^2)\sum_{\ell=0}^{\infty}\rho_\eta^\ell b_\ell(\tilde{a},\tilde{b})$$

Let us now consider the differences between these three perceived laws of motion by analyzing their implications on agents' forward-looking reasoning. Suppose just for the present explanation that η_t is commonly known to be i.i.d, so we can simplify exposition. Consider realized output at a future horizon-$h > 0$, which takes the general form

$$Y_{t+h} = \underbrace{(\alpha_1\lambda_{t+h-1} + \alpha_2\lambda_{t+h-2} + \alpha_3\lambda_{t+h-3} + ...)}_{\text{Belief feedback } \chi E_{t+h-1}[\lambda_{t+h}]} + \underbrace{\lambda_{t+h}}_{\text{Direct shock effect}}$$

The rational learning perceived law of motion adopts this exact form and hence can fully account for the contributions of past and current shocks to time-$t+h$ output through time-$t+h-1$ beliefs. The simple perceived law of motion is on the opposite extreme, where it only accounts for the direct shock effect.

Constrained-rational learning constitutes a middle ground between these two cases. The state variable λ_{t+h} in the perceived law of motion $\tilde{Y}_{t+h}$ cannot span the infinite history of past shocks $\{\lambda_{t+h-\ell}\}_{\ell>0}$, which appear from the Wold decomposition of the belief feedback $\chi E_{t+h-1}[\lambda_{t+h}]$. However, because λ_{t+h} is an autoregressive process it correlates with the past innovations $\{u_{\lambda,t+h-\ell}\}_{\ell>0}$ and consequently the past shocks $\{\lambda_{t+h-\ell}\}_{\ell>0}$ that determine time-$t+h-1$ beliefs $E_{t-1}[\lambda_{t+h-\ell}]$. Hence the projection of these past innovations onto the shock λ_{t+h} allows $\tilde{Y}_{t+h}$ to partially capture the dependence of beliefs of past shocks due to their covariance with λ_{t+h}.

Current output Y_t is determined by households who base their consumption spending

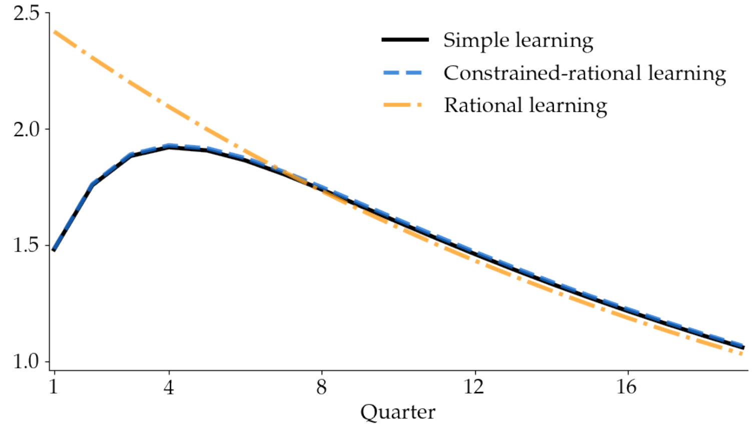


Figure 10: Output impulse responses to an η_t shock under different forms of learning

Note: Each line represents the impulse response of output to a transitory demand shock under different learning assumptions. I omit period zero because it does not contain an expectational response due to the staggered timing assumption.

on all future output expectations $\{E_{t-1}[Y_{t+h}]\}_{h>0}$. Because expectations of output at each horizon-h imperfectly captures the history-dependence of belief feedback on past shocks, the constrained-rational and simple perceived laws of motion result in a significant degree of initial dampening in realized output Y_t due to this impairment in forward reasoning. The below proposition formalizes the intuition for the dampening in constrained-rational learning versus rational learning in the special case where the transitory shocks η_t are i.i.d.

Proposition 2. *Suppose $\chi > 0$, η_t is i.i.d, and σ_η are normalized to make constrained-rational and rational Kalman gains proportional. Let Y_t^R, Y_t^{CR} denote output under rational learning and constrained-rational learning respectively.*

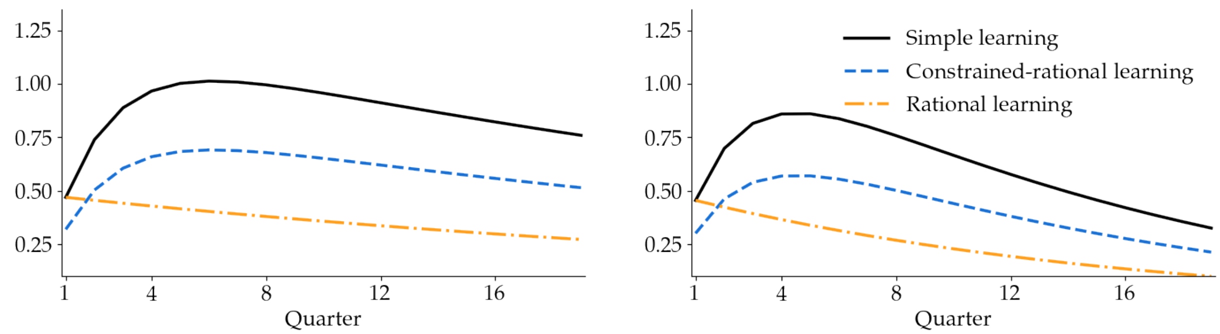
- *In response to a transitory $u_{\eta,0}$ innovation to η_0, $Y_t^R > Y_t^{CR}$ for time-$t \leq \bar{t}$, where $\bar{t} > 1$.*
- *In response to a persistent $u_{\lambda,0}$ innovation to λ_0, $Y_t^R > Y_t^{CR}$ for time-$t > 1$.*

Proof in Appendix B.2.

Figure 10 displays the response of output to a transitory η_t shock that has positive persistence $\rho_\eta > 0$ under the previously described perceived laws of motion. The response of output under rational learning is immediate and monotonically decreasing[17], similar to a typical full-information rational expectations impulse response in a standard HANK model. Additionally, it is much larger than its imperfect learning counterparts because it accounts for the full set of dynamic general equilibrium expectation feedback effects. In contrast, the response of output under constrained-rational learning is muted and closely resembles simple learning.

Eventually the output response under simple and constrained-rational learning exceeds that of rational learning. While Proposition 2 addresses a special case where η_t lacks persistence, the intuition behind the threshold $\bar{t}$ at which output under constrained-rational learning

[17]Note that output dynamics under rational learning can exhibit inertia in principle because of the gradual adjustment of beliefs in the Kalman update equation. However, quantitatively output inertia tends not to occur because the extent of inertia in component beliefs will be small relative to the decay of the direct shock effect.



(a) Persistent $E_{t-1}[\lambda_t]$

(b) Transitory $E_{t-1}[\eta_t]$

Figure 11: Component belief impulse responses to an η_t shock under different forms of learning

Note: Each line represents the impulse response of component beliefs to a transitory demand shock under different learning assumptions. I omit period zero because it does not contain an expectational response due to the staggered timing assumption.

exceeds rational learning is useful to fix ideas. Initially, output under constrained-rational learning underreacts relative to rational learning because it does not fully account for future equilibrium feedback through the persistence of component beliefs. After time-$\bar{t}$, the added endogenous persistence in output under constrained-rational learning due to the belief feedback wedge causes it to exceed output under rational learning.

While the output response is similar under both cases of imperfect learning, the behavior of component beliefs exhibit larger differences. Figure 11 displays the responses of persistent and transitory component beliefs to a transitory shock. Under simple learning, all changes in output are inferred to be due to changes in the underlying shock components. The reinforcing feedback of component beliefs into equilibrium output unanchors expectations under simple learning relative to the other perceived laws of motion which partially account for this equilibrium feedback in their belief formation. In contrast, under constrained-rational learning component beliefs respond by less and more closely resemble rational learning.

Why is the output response under constrained-rational learning more similar to simple learning, if its component beliefs are more similar to rational learning? To explain this, it is useful to consider the relative magnitude of the response of output to a given change in belief components. Contrasting simple and constrained-rational learning, we see that beliefs respond less under constrained-rational learning but deliver the same output response. As discussed previously, the degree of amplification of beliefs is much larger under constrained-rational learning because it accounts for some of the belief feedback. However, we see that the degree of amplification is still muted compared to rational learning which accounts for the full quantity of belief feedback.

The role of the belief multiplier χ Let us now consider how the belief multiplier χ affects the evolution of component beliefs in the rational and constrained-rational cases. In the rational case, the belief multiplier χ does not enter the component belief law of motion at all because there is no belief feedback wedge. The reason is expectations update optimally by accounting for belief feedback into equilibrium output irrespective of its size.

$$\mathbb{E}_t[\boldsymbol{\varepsilon}_{t+1}] = \mathbf{F}\mathbb{E}_{t-1}[\boldsymbol{\varepsilon}_t] + \mathbf{g}' \underbrace{(\mathbf{1}\boldsymbol{\varepsilon}_t - \mathbb{E}_{t-1}[\boldsymbol{\varepsilon}_t])}_{\text{Forecast error } Y_t - \mathbb{E}_{t-1}[Y_t]}$$

Given expectations $\mathbb{E}_{t-1}[\boldsymbol{\varepsilon}_t]$ are predetermined and fully accounted for in the rational learning perceived law of motion, the steady state Kalman gain $\mathbf{g}$ remains the same as in the simple learning case and is also unaffected by the multiplier χ.

Conversely, we see that in the constrained-rational case, the belief multiplier χ enters into the component belief law of motion resulting in a similar wedge to the simple learning case

$$\mathbb{E}_t[\boldsymbol{\varepsilon}_{t+1}] = \mathbf{F}\mathbb{E}_{t-1}[\boldsymbol{\varepsilon}_t] + \mathbf{g}' \underbrace{(\mathbf{1}\boldsymbol{\varepsilon}_t + \tilde{\mathbf{a}} \odot (\chi\mathbf{h} - \mathbf{1})E_{t-1}[\boldsymbol{\varepsilon}_t])}_{\text{Forecast error } Y_t - E_{t-1}[Y_t]}$$

where projection coefficients $\tilde{\mathbf{a}} := (\tilde{a}, \tilde{b})$ and $\odot$ is the element-wise (Hadamard) product.

Because the projection coefficients, which are functions of the belief multplier χ, directly load on the belief components λ_t, η_t in the constrained-rational perceived law of motion, the endogenous Kalman gains $\mathbf{g}$ are now affected by χ. This intuition for this is as follows. Households do not explicitly account for belief feedback in their perceived law of motion. Nonetheless in estimating coefficients $\tilde{\mathbf{a}}$ on components λ_t, η_t to best fit the observed dynamics of output they partially pick up the influence of expectations feedback because λ_t, η_t are functions of past component innovations embedded in expectations. This appears to them as if output fluctuations are directly more sensitive to λ_t, η_t changes than just the unit direct shock effect would suggest, causing the endogenous gain $\mathbf{g}$ to be lower when the multiplier χ is larger.

The differences in gains $\mathbf{g}$ determine the difference in the initial period response of component beliefs in Figure 11. Even though the gain $\mathbf{g}$ is lower in the constrained-rational case, which would suggest that output fluctuations should be less sensitive to changes in component beliefs, the projection coefficients $\tilde{\mathbf{a}}$ are typically larger when χ is larger, which goes in the opposite direction and amplifies the output response. These two forces largely cancel each other out, resulting in similar responses of output under simple and constrained-rational learning.

5 Policy implications of macroeconomic inertia

This section discusses novel policy considerations that arise under constrained learning and contrast them with typical policy transmission outcomes in full-information rational expectations HANK models. Under constrained learning, it is no longer desirable to be infinitely-responsive to demand-driven fluctuations because of the risk of destabilizing expectations. Gradual monetary policy approaches, in the form of a highly inertial Taylor rule, fail to stabilize output as effectively relative to the full-information rational expectations benchmark. The key reason is the inability of constrained learning to account for the future equilibrium impacts of current policy commitments. Hence, inertial monetary policy only gradually stabilizes output as past policy commitments accumulate into larger realized interest rate changes. The effects of deficit-financed fiscal stimulus are prolonged because of the same mechanism, where deferring deficits causes government debt holdings by households to accumulate, resulting in larger wealth effects.

5.1 Simple Taylor rules

The typical monetary policy prescription in response to demand shocks is to completely close the output gaps they induce, which aligns with the welfare aims of inflation stabilization in standard New Keynesian economies (Blanchard and Galí 2007). This divine coincidence will likewise hold in my setting, assuming firms expectations are the same as households. Therefore I use the discounted path of squared deviations of output from steady state as a simple welfare measure to contrast counterfactual policy rules, which do not affect the steady state.

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t Y_t^2 \tag{25}$$

I consider first the full-information rational expectations equilibrium response to a transitory shock η_t. Given the unnormalized aggregate demand equation

$$Y_t = \frac{1}{\beta\omega(1+\sigma\phi)} \left(\underbrace{(1-\beta\omega-\beta\omega\sigma\phi)}_{\text{Belief multiplier } \chi} \sum_{h=1}^{\infty} (\beta\omega)^h \mathbb{E}_t[Y_{t+h}] + \eta_t \right)$$

The equilibrium solution is given by $Y_t = b\eta_t$ where the coefficient b is

$$b = \underbrace{\frac{1}{\beta\omega(1+\sigma\phi)}}_{\to 0 \text{ as } \phi\to\infty} \underbrace{\left(1 - \frac{1-\beta\omega-\beta\omega\sigma\phi}{1+\sigma\phi}\right)^{-1}}_{\to(1-\beta\omega)^{-1} \text{ as } \phi\to\infty}$$

Because welfare is given by the discounted squared loss of output, the optimal choice of the Taylor coefficient that completely closes the output gap is for the monetary authority to be infinitely responsive $\phi \to \infty \implies b \to 0$. Further, welfare loss strictly decreases as ϕ increases for any finite ϕ. This shows that in the standard full-information rational expectations setting, a counterfactual policy that is more responsive to demand shocks is always welfare-improving.

However, when agents form expectations with simple and constrained-rational learning the optimal policy prescription differs. Simple learning yields the aggregate demand equation

$$Y_t = \underbrace{\frac{1-\beta\omega-\beta\omega\sigma\phi}{1+\sigma\phi}}_{\to-\beta\omega \text{ as } \phi\to\infty} \left(\frac{\rho_\lambda}{1-\beta\omega\rho_\lambda} E_{t-1}[\lambda_t] + \frac{\rho_\eta}{1-\beta\omega\rho_\eta} E_{t-1}[\eta_t] \right) + \underbrace{\frac{1}{\beta\omega(1+\sigma\phi)}}_{\to 0 \text{ as } \phi\to\infty} (\lambda_t + \eta_t)$$

In the infinitely-responsive ϕ limit, output is perfectly stabilized at the steady state if component beliefs are fully anchored $E_{-1}[\lambda_0] = E_{-1}[\eta_0] = 0$. Unlike in the full-information rational expectations case, if ϕ is not taken fully to the infinite limit but instead is finite and sufficiently large then the component beliefs $E_t[\varepsilon_{t+1}]$ and consequently output Y_t itself will be destabilized. Higher monetary responsiveness is therefore effective only up to a point, a limitation that is similarly demonstrated in Eusepi et al. (2024).

Figure 12 demonstrates that this behavior also holds in the constrained-rational learning case. The shared reason in both simple and constrained-rational learning is the expectation feedback wedge that appears in the belief component law of motion is increasingly negative as ϕ increases. The intuition is by consuming more on the basis of positive component beliefs,

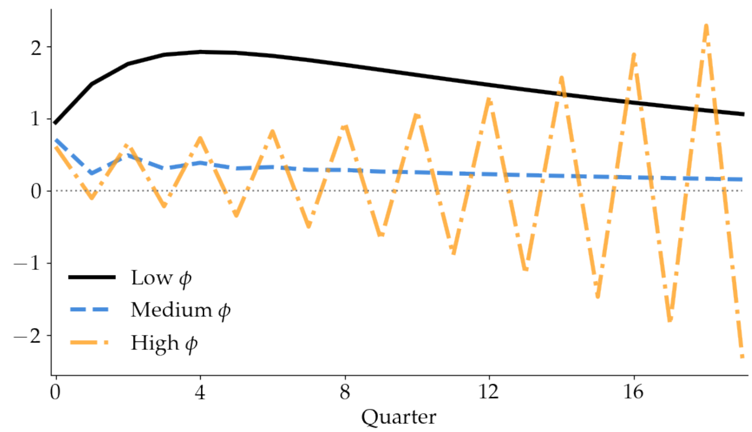


Figure 12: Output (de-)stabilization under different Taylor rule ϕ

Note: Each line represents the impulse response of output to a transitory demand shock in an economy with different Taylor rule coefficients ϕ.

the monetary authority raises interest rates sufficiently to induce a contraction in output. This causes households to mistakenly infer that the shock components were actually realized to be negative and larger in magnitude than previously anticipated. The result is an increasingly unstable negative feedback loop, resulting in the explosive oscillation of the high ϕ case shown in Figure 12.

However, with mildly elevated responsiveness in the medium ϕ we see not only a reduced level response of output to the shock but also the absence of output inertia. Even though there is an expectation wedge in the component belief law of motion that in principle contributes to inertia, by choosing a Taylor coefficient ϕ that sets the belief multiplier χ close to zero enables the monetary authority to shut down the expectation feedback loop that induces inertia through endogenous expectations unanchoring.

In choosing an optimal level of responsiveness to demand shocks, a monetary authority facing households with learning constraints should not respond as forcefully as in rational benchmarks because of the risk of destabilizing expectations.

5.2 Inertial Taylor rules and monetary policy gradualism

A popular Taylor rule specification includes a lagged or "inertial" term

$$r_t = \rho r_{t-1} + \phi Y_t$$

Early justifications for this approach were based on observed inertia in interest rate policy (Clarida et al. 1998). However whether the policy rules themselves are inertial or are simply responding to inertial economic conditions was subject to debate (Rudebusch 2005). Other justifications for inertial policy rules include uncertainty about the effects of policy Sack (1998) and their ability to implement optimal allocations when forward-looking agents understand the dynamic implications of policy commitments as in (Woodford 1999).

I expand briefly on this latter reason by demonstrating the inability of constrained-rational and simple learning to fully understand dynamic policy commitments. The intuition for this closely resembles the expectation feedback mechanism discussed in the previous section. Just as constrained households are unable to fully internalize the equilibrium feedbacks of future expected output changes on current output, so too are they unable to internalize the effects of current policy commitments on future expected output which in turn has equilibrium consequences for current output.

The aggregate demand equation for output Y_t for arbitrary subjective expectations E_t, where households understand the inertial form of the policy rule and its parameters yields

$$Y_t = \underbrace{-\frac{\sigma\bar{\phi}}{1+\sigma\bar{\phi}}\left(\sum_{\ell=1}^{\infty}\rho^{\ell}Y_{t-\ell}\right)}_{\text{Policy commitments in } r_{t-1}} + \frac{1-\beta\omega-\beta\omega\sigma\bar{\phi}}{1+\sigma\bar{\phi}}\sum_{h=1}^{\infty}(\beta\omega)^{h-1}E_t[Y_{t+h}] + \frac{1}{\beta\omega(1+\sigma\bar{\phi})}\varepsilon_t$$

The "effective" Taylor coefficient $\bar{\phi} = \frac{\phi}{1-\beta\omega\rho}$ demonstrates that whether policy responds contemporaneously via ϕ or with a delay via ρ, there is a way to equate their effective contribution toward current equilibrium output Y_t conditional on future expectations $\{E_t[Y_{t+h}]\}_{h>0}$. Hence, in response to an unanticipated shock at time-0, absent pre-existing policy commitments $r_{-1} = 0$ and fixing a given path of future expectations $\{E_t[Y_{t+h}]\}_{h>0}$ the response of time-0 output Y_0 should be the same for a continuum of regimes (ρ, ϕ) that induce the same effective $\bar{\phi}$.

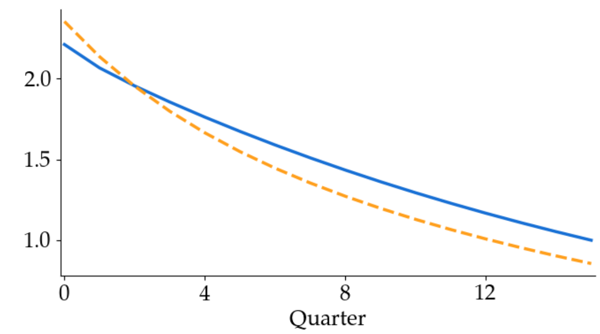
However, the crucial step in the above consideration is that the path of future expectations was held fixed. Consider if households correctly perceived time-$t+h$ output used to inform time-t consumption which determines time-t output in equilibrium.

$$\tilde{Y}_{t+h} = \underbrace{-\frac{\sigma\bar{\phi}}{1+\sigma\bar{\phi}}\left(\sum_{\ell=1}^{\infty}\rho^{\ell}\tilde{Y}_{t+h-\ell}\right)}_{\text{Function of } \{\varepsilon_{t+h-\ell}\}_{\ell>0}} + \frac{1-\beta\omega-\beta\omega\sigma\bar{\phi}}{1+\sigma\bar{\phi}}\underbrace{\sum_{j=1}^{\infty}(\beta\omega)^{j-1}E_{t+h}[Y_{t+h+j}]}_{\text{Function of } \{\varepsilon_{t+h-\ell}\}_{\ell\geq 0}} + \frac{1}{\beta\omega(1+\sigma\bar{\phi})}\varepsilon_{t+h}$$

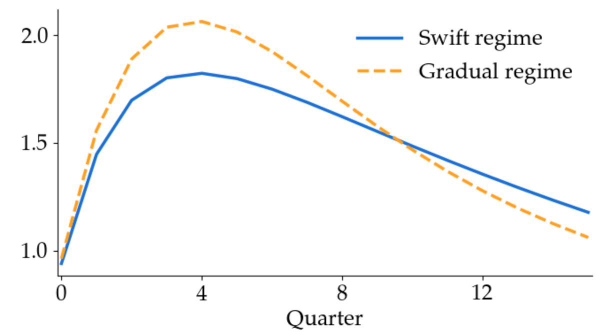
By correctly perceiving future output at time-$t+h$, households understand that current policy decisions which respond to current shocks will persist into time-$t+h$ with persistence ρ. Hence, two regimes with the same effective $\bar{\phi}$ would exhibit different equilibrium responses of output at time-0 if households perceptions correctly detected that the regime with higher policy rule persistence ρ would continue to respond more forcefully in future periods to the time-0 shock.

However, when learning rules are restricted to load on contemporaneous shocks as in the constrained-rational and simple learning cases where $\tilde{Y}_{t+h}$ can only be a function of ε_{t+h}, we see both expectation feedback in $\{E_{t+h}[Y_{t+h+j}]\}_{j>0}$ and lasting effects of current policy commitments on future output $\{\tilde{Y}_{t+h-\ell}\}_{\ell>0}$ must be partially ignored. This is again because contemporaneous shocks ε_{t+h} cannot span the space of all past shocks $\{\varepsilon_{t+h-\ell}\}_{\ell>0}$.

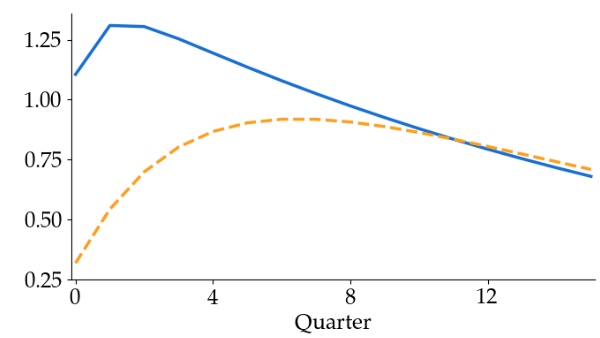
To demonstrate the consequences for welfare, I utilize the discounted squared output loss $\mathcal{L}$ in Equation (25) from before and consider two policy regimes. I call the first policy regime the "swift" policy regime, $(\rho^{\mathrm{S}}, \phi^{\mathrm{S}})$, and the second one the "gradual" policy regime, $(\rho^{\mathrm{G}}, \phi^{\mathrm{G}})$, where the swift regime exhibits less inertia $\rho^{\mathrm{S}} < \rho^{\mathrm{G}}$ and greater contemporaneous responsiveness $\phi^{\mathrm{S}} > \phi^{\mathrm{G}}$. I choose these regimes to equate their welfare loss from a transitory shock under a full-information rational expectations benchmark.



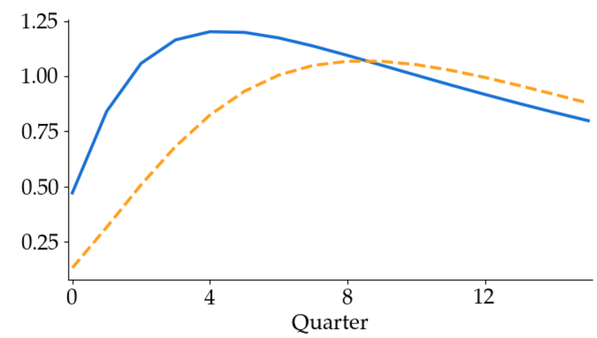
(a) Output response under FIRE



(b) Output response under learning



(c) Interest rate response under FIRE



(d) Interest rate response under learning

Figure 13: Output and interest rate impulse responses to an η_t shock across monetary regimes

Note: The top panels represent the impulse responses of output to a transitory demand shock and the bottom panels the analogous responses for the real interest rate. The left column plots the impulse responses under full-information rational expectations (FIRE), and the right column under constrained learning.

Figure 13 demonstrates the responses to a transitory shock of output in the top panel and interest rates in the bottom panel for the swift and gradual regimes. The left column of responses are under the full-information, rational expectations (FIRE) benchmark and the right column are with constrained-rational learning. Focusing first on the FIRE case, we see that under both regimes output responds immediately, where interest rates in the swift regime rise by more initially bringing output down by more. However, this difference only lasts for two quarters before output in the gradual regime crosses the swift regime even though interest rates in the gradual regime have yet to catch up to the swift regime. This occurs because households rationally understand the accumulated effects of interest rate changes will persist for a long time and reduce consumption spending accordingly.

In contrast, under learning we see that output continues to be less contained in the gradual than in the swift regime until gradual regime interest rates exceed those of the swift regime. Figure 14 demonstrates the difference in current-period discounted welfare $-\beta^t Y_t^2$ between the two regimes. Because these regimes were chosen to equate total welfare loss under full-information rational expectations, the area under the blue curve in Figure 14 integrates to 1. We see that the swift regime achieves higher initial welfare for the first two quarters by responding more forcefully but the gradual regime slowly makes up for the welfare difference in the long-run. However, with learning because the gradual regime is less effective at containing output in the short-run due to frictions households face in forward reasoning, the initial differences in welfare loss are too large to be offset by persistently higher rates and a smaller output response in later periods.

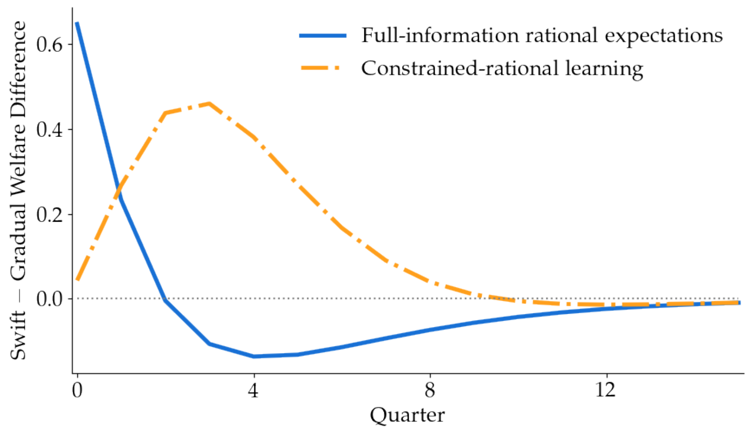


Figure 14: Monetary regime welfare differences under different models of expectation formation

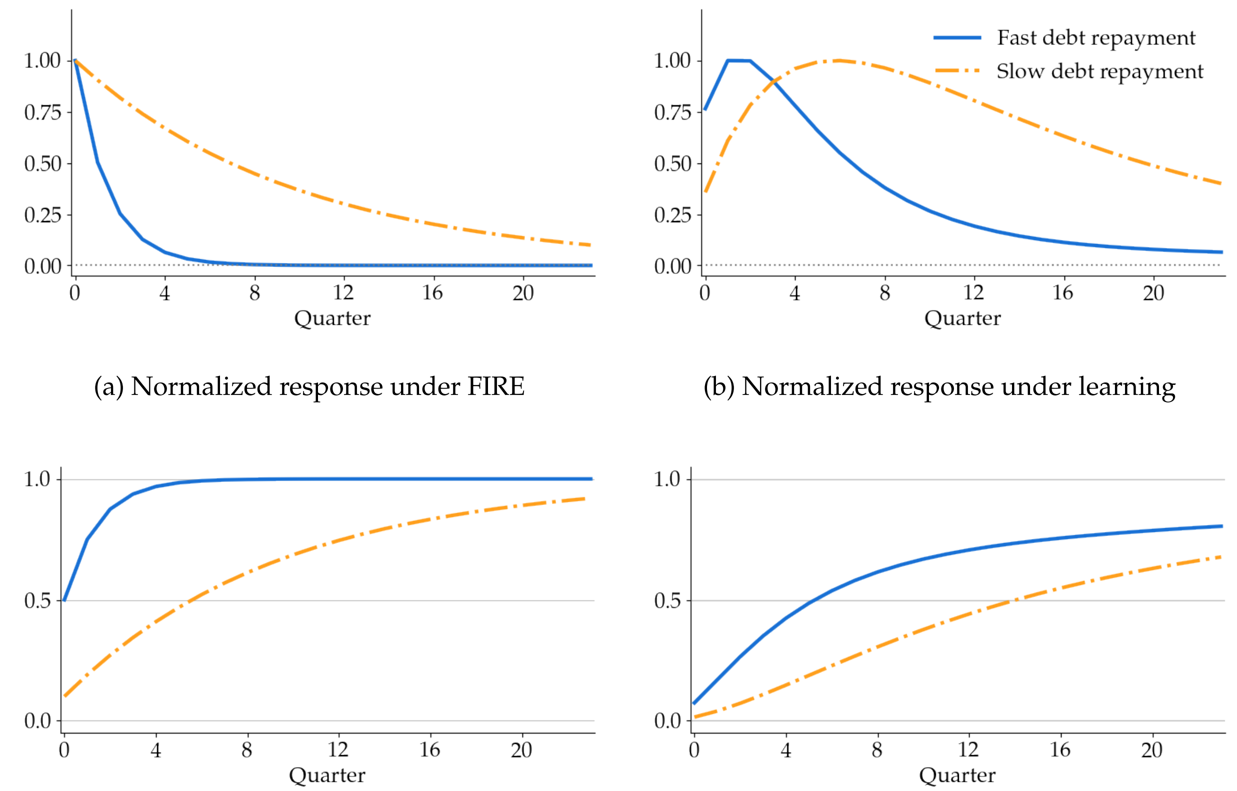
Note: The measure of welfare loss is the discounted squared output deviation each quarter. Each line corresponds to the difference in welfare loss incurred between the swift and gradual policy regime under a different model of expectation formation. When the lines exceed zero, the swift regime incurred a lower discounted welfare loss in that quarter and vice versa. The regimes were chosen such that the area between the blue curve and zero integrates to zero.

5.3 Deferred financing and the delayed impacts of fiscal stimulus

One of the key differences in the policy implications of heterogeneous-agent versus representative-agent macroeconomic models are their responses to fiscal stimulus (Auclert et al. 2024, Angeletos et al. 2023). Representative-agent models have Ricardian equivalence and hence the effects of government spending or transfer policies are unaffected by the timing of financing. Conversely, deficit-financed spending and transfers induce large and immediate output responses in heterogeneous-agent models. Further, the magnitude of these responses increase the further financing is delayed.

When households face learning constraints, as considered in this paper, there are additional implications of financing delays for the time profile of output in response to fiscal stimulus. In particular, deferred financing can delay the peak response to fiscal stimulus, potentially reducing its effectiveness if the policymaker desires immediate results. Prolonging deficits additionally stretches out the cumulative response of output across a much longer horizon, running the risk that stimulus will last longer than intended and beyond the initially desired period of fiscal support. The propagation channels that induce these effects are similar to those in governing expectation feedback and prior monetary policy commitments.

I consider a simple form of fiscal policy that resembles the setting in Angeletos et al. (2023). The government issues real-valued debt B_t that is financed by a lump-sum tax T_t. Lump-sum taxes adjust to repay debt gradually, where the speed of repayment is given by $\delta \in (0, 1)$. The



(a) Normalized response under FIRE

(b) Normalized response under learning

(c) Discounted cumulative response under FIRE

(d) Discounted cumulative response under learning

Figure 15: Output impulse responses to a transfer shock ζ_t across debt repayment speed regimes

Note: The top panel displays the output impulse response under each debt repayment regime, where the peak response is normalized to one. The bottom panel displays the share of the cumulative impulse response of output, discounted by the inverse gross interest rate.

linearized government budget constraint and tax rule are given by

$$B_t = \frac{1}{\beta}(B_{t-1} - T_t)$$
$$T_t = \delta B_{t-1} - (1-\delta)\zeta_t$$

I assume there is no outstanding government debt in steady state $B = 0$, and ζ_t denotes an i.i.d deficit shock which I will use in the following policy exercise. By assuming debt is real-valued, I omit the possibility that surprise inflation erodes the real value of debt through nominal revaluation so we can still maintain our focus on real output alone. In addition, we now need to enforce asset market clearing between household wealth and government debt.

$$A_t = B_t$$

Given this setting we can again derive an aggregate demand equation that determines equilibrium output that is analogous to Equation (13) in Angeletos et al. (2023) but without the real interest rate peg.

$$Y_t = \frac{1}{1-\chi}\left(\frac{(1-\beta\omega)(1-\omega)(1-\delta)}{1-\omega(1-\delta)}(B_{t-1}+\zeta_t) + \chi\sum_{h=1}^{\infty}(\beta\omega)^h E_t[Y_{t+h}] + \varepsilon_t\right)$$

where the belief multiplier $\chi := (1-\beta\omega-\beta\omega\sigma\phi)$ as before.

Figure 15 displays the response of output to a time-0 deficit shock ζ_0 under fast (large δ) and slow (small δ) debt repayment regimes. The top panel displays the response of output with its peak period normalized to one, and the bottom panel displays the cumulative response of output, discounted by the inverse steady state gross interest rate $(1+r)^{-1} \equiv \beta$ with a total response normalized to one. The left column displays the responses under a full-information rational expectations (FIRE) benchmark and the right column under constrained-rational learning.

In the FIRE case, we see that the initial response of output to a one-time fiscal transfer is peaked on impact and monotonically decreasing. This is because the full dynamic effects of higher debt holdings and slower debt repayment are internalized on impact by household consumption decisions. A large share of the discounted cumulative response of output $\sum_{t=0}^{\infty}(1+r)^{-t}Y_t$, which is a commonly-used measure of the size of fiscal stimulus (Mountford and Uhlig 2009), also occurs at relatively short horizons. In the FIRE case, half of the discounted cumulative output response occurs immediately under the fast debt repayment regime and after five quarters in the slower repayment regime. In contrast, with constrained-learning there is a difference in the peak output response of one year between regimes and ten quarters for the discounted cumulative response, more than double the gap under FIRE.

6 Conclusion

This paper demonstrates that canonical heterogeneous-agent models are not just consistent with aggregate consumption inertia but fundamentally contribute to its emergence. I first show that the minimal structure imposed by these models yield model-implied impulse responses that closely resemble observed consumption inertia. I then adopt a learning model that matches the over-extrapolation bias displayed in the expectations data to identify the features of expectations that cause inertia to arise. An interaction between over-extrapolation bias and the belief multiplier, a key model quantity representing the size of equilibrium amplification, determines whether inertia emerges and how protracted it is.

The learning frictions that result in inertial amplification yield additional costs of monetary policy gradualism and the deferred financing of fiscal deficits. Under rational expectations these policy approaches can produce a large and immediate consumption response because of households ability to accurately reason far into the future. In the learning model I consider, where households far-horizon reasoning is constrained, delayed policy action induces expectations to unanchor which can reduce policy effectiveness and lengthens policy transmission lags.

References

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

Aiyagari, S Rao, "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 1994, *109* (3), 659–684.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective models of the macroeconomy: Evidence from experts and representative samples," *The Review of Economic Studies*, 2022, *89* (6), 2958–2991.

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

__ , Chen Lian, and Christian K Wolf, "Can Deficits Finance Themselves?," Technical Report, National Bureau of Economic Research 2023.

__ , Fabrice Collard, and Harris Dellas, "Business-cycle anatomy," *American Economic Review*, 2020, *110* (10), 3030–3070.

__ , Zhen Huo, and Karthik A Sastry, "Imperfect macroeconomic expectations: Evidence and theory," *NBER Macroeconomics Annual*, 2021, *35* (1), 1–86.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models," *Econometrica*, 2021, *89* (5), 2375–2408.

__ , Matthew Rognlie, and Ludwig Straub, "Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model," Technical Report, National Bureau of Economic Research 2020.

__ , __ , and __ , "The Intertemporal Keynesian Cross," *Journal of Political Economy*, 2024.

Bardóczy, Bence and Joao Guerreiro, "Unemployment insurance in macroeconomic stabilization with imperfect expectations," *Manuscript, April*, 2023.

Barnichon, Regis and Geert Mesters, "Identifying modern macro equations with old shocks," *The Quarterly Journal of Economics*, 2020, *135* (4), 2255–2298.

Bayer, Christian, Benjamin Born, and Ralph Luetticke, "Shocks, frictions, and inequality in US business cycles," *American Economic Review*, 2024, *114* (5), 1211–1247.

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

Best, Michael Carlos, James S Cloyne, Ethan Ilzetzki, and Henrik J Kleven, "Estimating the elasticity of intertemporal substitution using mortgage notches," *The Review of Economic Studies*, 2020, *87* (2), 656–690.

Bewley, Truman, "Stationary monetary equilibrium with a continuum of independently fluctuating consumers," *Contributions to mathematical economics in honor of Gérard Debreu*, 1986, *79.*

Blanchard, Olivier and Jordi Galí, "Real wage rigidities and the New Keynesian model," *Journal of money, credit and banking*, 2007, *39*, 35–65.

Blanchard, Olivier J, "Debt, deficits, and finite horizons," *Journal of Political Economy*, 1985, *93*, 223–247.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Diagnostic expectations and credit cycles," *The Journal of Finance*, 2018, *73* (1), 199–227.

__ , __ , Yueran Ma, and Andrei Shleifer, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–2782.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko, "Communication and the beliefs of economic agents," Technical Report, National Bureau of Economic Research 2020.

Carroll, Christopher D, Edmund Crawley, Jiri Slacalek, Kiichi Tokuoka, and Matthew N White, "Sticky expectations and consumption dynamics," *American economic journal: macroeconomics*, 2020, *12* (3), 40–76.

Chetty, Raj and Adam Szeidl, "Consumption commitments and habit formation," *Econometrica*, 2016, *84* (2), 855–890.

Christiano, Lawrence and Yuta Takahashi, "Anchoring Inflation Expectations," Technical Report, Mimeo 2020.

Christiano, Lawrence J, Martin Eichenbaum, and Benjamin K Johannsen, "Slow Learning," 2024.

__ , __ , and Charles L Evans, "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 2005, *113* (1), 1–45.

Clarida, Richard, Jordi Galı, and Mark Gertler, "Monetary policy rules in practice: Some international evidence," *European economic review*, 1998, *42* (6), 1033–1067.

Crump, Richard K, Stefano Eusepi, Emanuel Moench, and Bruce Preston, "The term structure of expectations," in "Handbook of economic expectations," Elsevier, 2023, pp. 507–540.

da Silveira, Rava Azeredo, Yeji Sung, and Michael Woodford, "Optimally imprecise memory and biased forecasts," *American Economic Review*, 2024, *114* (10), 3075–3118.

Dynan, Karen E, "Habit formation in consumer preferences: Evidence from panel data," *American Economic Review*, 2000, *90* (3), 391–406.

Eusepi, Stefano and Bruce Preston, "Expectations, learning, and business cycle fluctuations," *American Economic Review*, 2011, *101* (6), 2844–2872.

— , **Marc Giannoni, and Bruce Preston**, "The Short-run Policy Constraints of Long-run Expectations," *Journal of Political Economy*, 2024.

Evans, George W and Seppo Honkapohja, "Learning dynamics," *Handbook of macroeconomics*, 1999, *1*, 449–542.

Farmer, Leland E, Emi Nakamura, and Jón Steinsson, "Learning about the long run," *Journal of Political Economy*, 2024, *132* (10), 000–000.

Fuhrer, Jeffrey C, "Habit formation in consumption and its implications for monetary-policy models," *American economic review*, 2000, *90* (3), 367–390.

Gabaix, Xavier, "Behavioral inattention," in "Handbook of behavioral economics: Applications and foundations 1," Vol. 2, Elsevier, 2019, pp. 261–343.

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

Gagliardone, Luca and Mark Gertler, "Oil prices, monetary policy and inflation surges," 2023.

Galí, Jordi, *Monetary Policy, Inflation, and the Business Cycle*, Princeton University Press, 2015.

Guerreiro, Joao, "Belief disagreement and business cycles," *Northwestern University manuscript*, 2023.

Havranek, Tomas, Marek Rusnak, and Anna Sokolova, "Habit formation in consumption: A meta-analysis," *European economic review*, 2017, *95*, 142–167.

Huggett, Mark, "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of economic Dynamics and Control*, 1993, *17* (5-6), 953–969.

Imrohoroğlu, Ayşe, "Cost of business cycles with indivisibilities and liquidity constraints," *Journal of Political economy*, 1989, *97* (6), 1364–1383.

Jentsch, Carsten and Kurt G Lunsford, "The dynamic effects of personal and corporate income tax changes in the United States: Comment," *American Economic Review*, 2019, *109* (7), 2655–2678.

Kamdar, Rupal and Walker Ray, "The effects of news shocks and supply-side beliefs," 2023.

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

Kaplan, Greg and Giovanni L Violante, "The marginal propensity to consume in heterogeneous agent models," *Annual Review of Economics*, 2022, *14* (1), 747–775.

Kosar, Gizem and Cormac O'Dea, "Expectations data in structural microeconomic models," in "Handbook of Economic Expectations," Elsevier, 2023, pp. 647–675.

Kurmann, André and Eric Sims, "Revisions in utilization-adjusted TFP and robust identification of news shocks," *Review of Economics and Statistics*, 2021, *103* (2), 216–235.

Lewis, Daniel J and Karel Mertens, "Dynamic Identification Using System Projections and Instrumental Variables," 2022.

Luo, Yulei, "Consumption dynamics under information processing constraints," *Review of Economic dynamics*, 2008, *11* (2), 366–385.

Maćkowiak, Bartosz and Mirko Wiederholt, "Business cycle dynamics under rational inattention," *The Review of Economic Studies*, 2015, *82* (4), 1502–1532.

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

Manski, Charles F, "Measuring expectations," *Econometrica*, 2004, *72* (5), 1329–1376.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson, "The power of forward guidance revisited," *American Economic Review*, 2016, *106* (10), 3133–3158.

Mertens, Karel and Morten O Ravn, "The dynamic effects of personal and corporate income tax changes in the United States," *American economic review*, 2013, *103* (4), 1212–1247.

Milani, Fabio, "Expectation shocks and learning as drivers of the business cycle," *The Economic Journal*, 2011, *121* (552), 379–401.

Molavi, Pooya, "Simple models and biased forecasts," *arXiv preprint arXiv:2202.06921*, 2022.

Mountford, Andrew and Harald Uhlig, "What are the effects of fiscal policy shocks?," *Journal of applied econometrics*, 2009, *24* (6), 960–992.

Nagel, Stefan, "Leaning Against Inflation Experiences," Technical Report 2024.

Negro, Marco Del and Stefano Eusepi, "Fitting observed inflation expectations," *Journal of Economic Dynamics and control*, 2011, *35* (12), 2105–2131.

Newey, Whitney K and Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 1994, *4*, 2111–2245.

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

Ring, Marius Alexander Kalleberg, "Wealth taxation and household saving: Evidence from assessment discontinuities in Norway," *Review of economic studies*, 2024.

Romer, Christina D and David H Romer, "A new measure of monetary shocks: Derivation and implications," *American economic review*, 2004, *94* (4), 1055–1084.

Rozsypal, Filip and Kathrin Schlafmann, "Overpersistence bias in individual income expectations and its aggregate implications," *American Economic Journal: Macroeconomics*, 2023, *15* (4), 331–371.

Rudebusch, Glenn D, "Monetary policy inertia: fact or fiction?," *FRB of San Francisco Working Paper*, 2005, (2005-19).

Sack, Brian P, "Uncertainty, learning, and gradual monetary policy," *Available at SSRN 121439*, 1998.

Silva, Tim De and Pierfrancesco Mei, "Selective Inattention," Technical Report 2024.

Sims, Christopher A, "Implications of rational inattention," *Journal of monetary Economics*, 2003, *50* (3), 665–690.

Stock, James H and Mark W Watson, "Disentangling the Channels of the 2007-2009 Recession," Technical Report, National Bureau of Economic Research 2012.

Williams, Noah, "Adaptive learning and business cycles," *Manuscript, Princeton University*, 2003.

Woodford, Michael, "Optimal monetary policy inertia," *The Manchester School*, 1999, *67*, 1–35.

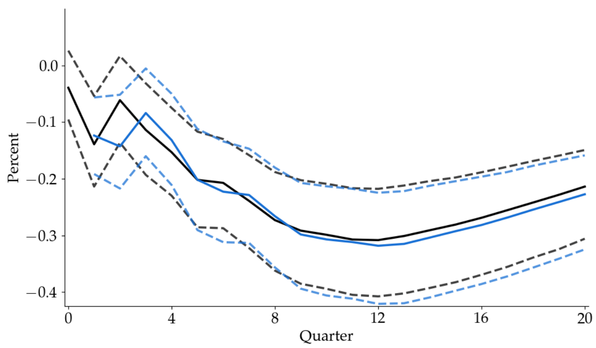
__ , "Imperfect common knowledge and the effects of monetary policy," 2001.

__ , "Macroeconomic analysis without the rational expectations hypothesis," *Annu. Rev. Econ.*, 2013, *5* (1), 303–346.

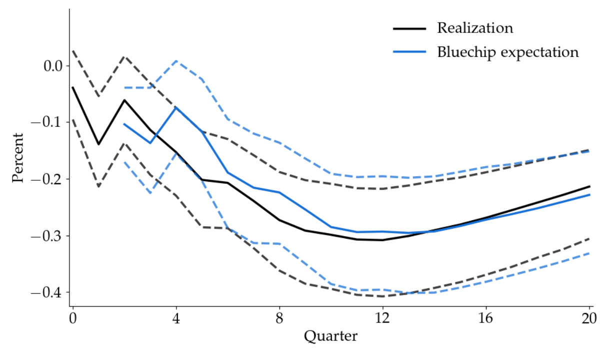
Yaari, Menahem E, "Uncertain lifetime, life insurance, and the theory of the consumer," *The Review of Economic Studies*, 1965, *32* (2), 137–150.

A Additional impulse response results

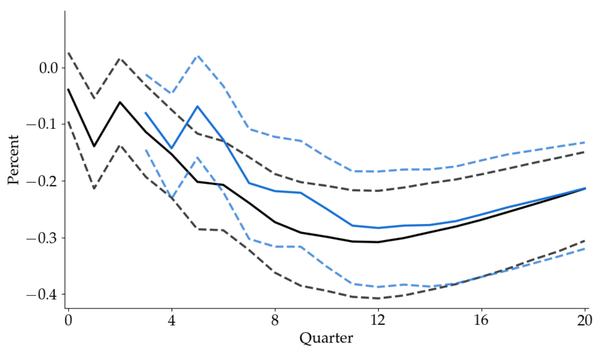
A.1 Impulse response comparisons with confidence bands



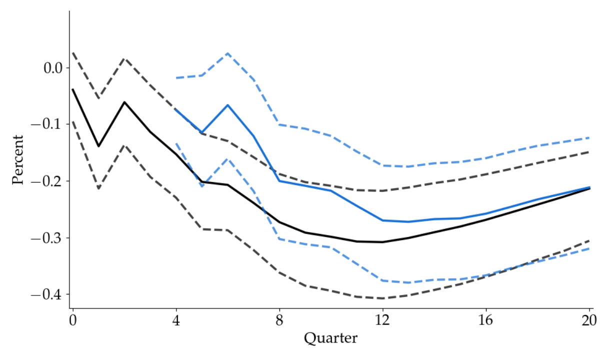
(a) One-quarter ahead Bluechip expectation



(b) Two-quarters ahead Bluechip expectation



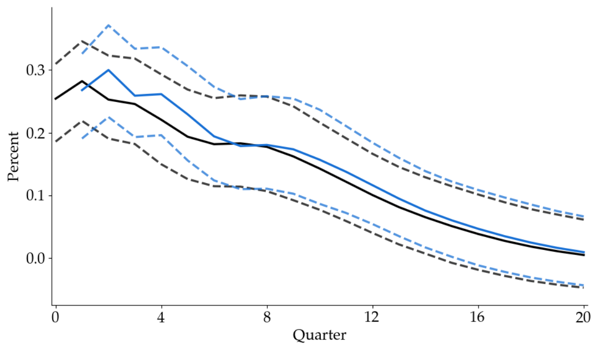
(c) Three-quarters ahead Bluechip expectation



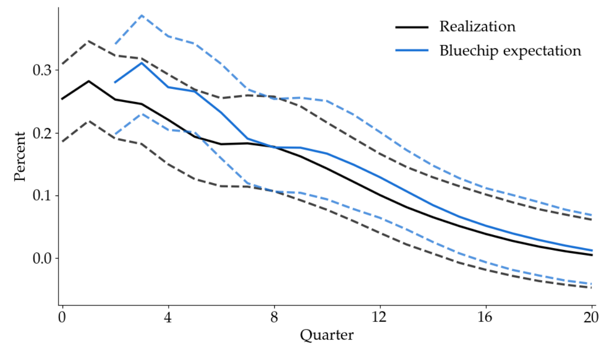
(d) Four-quarters ahead Bluechip expectation

Figure 16: Real disposable income impulse responses to a Känzig (2021) oil shock

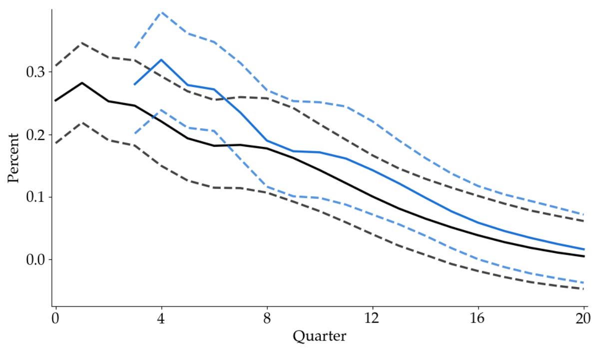
Note: each panel contains an impulse response function of realizations (black) and a fixed horizon-h (blue) forecast from Bluechip survey expectations data to a positive Känzig (2021) oil price news shock. The dashed lines are 68% confidence bands produced using moving block bootstrap by Jentsch and Lunsford (2019).



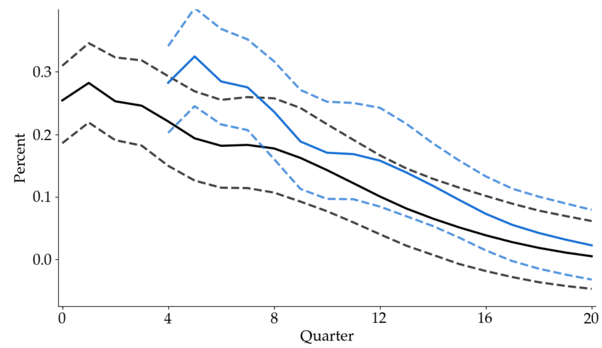
(a) One-quarter ahead Bluechip expectation



(b) Two-quarters ahead Bluechip expectation



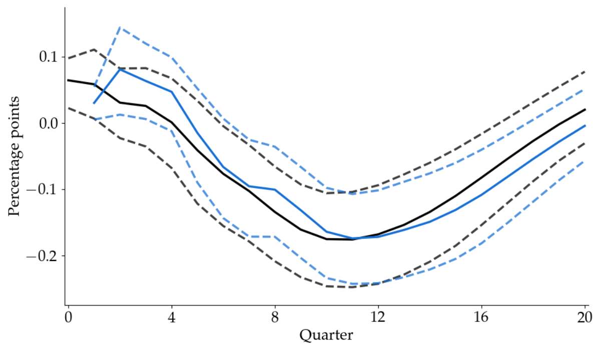
(c) Three-quarters ahead Bluechip expectation



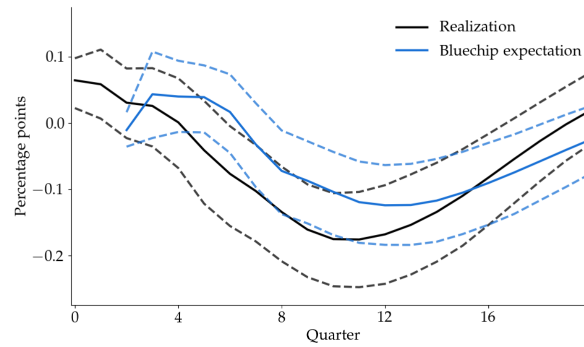
(d) Four-quarters ahead Bluechip expectation

Figure 17: CPI inflation impulse responses to a Känzig (2021) oil shock

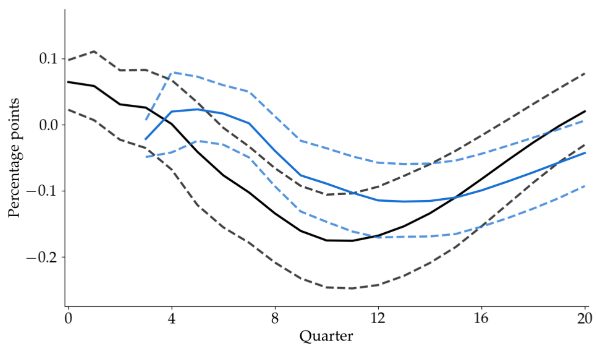
Note: each panel contains an impulse response function of realizations (black) and a fixed horizon-h (blue) forecast from Bluechip survey expectations data to a positive Känzig (2021) oil price news shock. The dashed lines are 68% confidence bands produced using moving block bootstrap by Jentsch and Lunsford (2019).



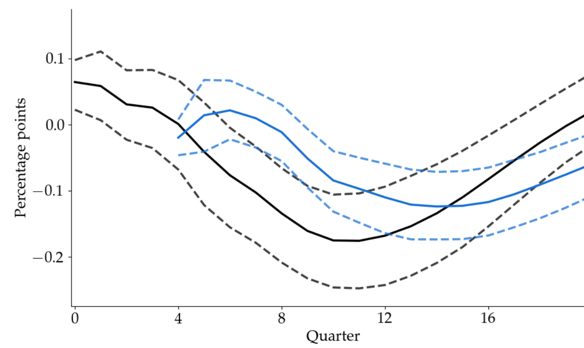
(a) One-quarter ahead Bluechip expectation



(b) Two-quarters ahead Bluechip expectation



(c) Three-quarters ahead Bluechip expectation



(d) Four-quarters ahead Bluechip expectation

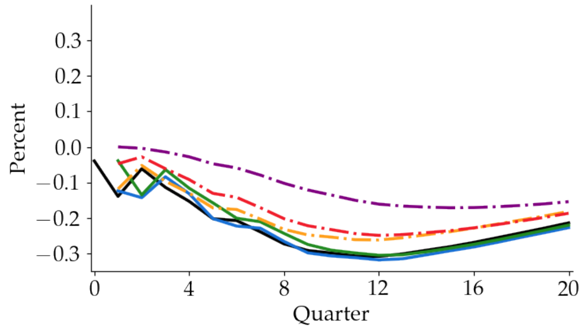
Figure 18: Nominal federal funds rate impulse responses to a Känzig (2021) oil shock

Note: each panel contains an impulse response function of realizations (black) and a fixed horizon-h (blue) forecast from Bluechip survey expectations data to a positive Känzig (2021) oil price news shock. The dashed lines are 68% confidence bands produced using moving block bootstrap by Jentsch and Lunsford (2019).

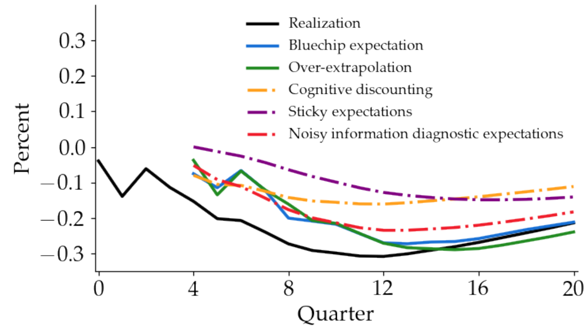
A.2 Other models of expectation formation

Figure 19 illustrates the difficulty that many existing models of expectation formation have matching impulse responses of expectations data. The left column plots the impulse response functions of each variables' realization, the one-quarter ahead Bluechip survey expectation and the same expectation implied by models of expectation formation. The right column plots the impulse response functions for the four-quarter ahead expectations. The horizons of each expectational impulse response is (vertically) aligned to the period it is forecasting.

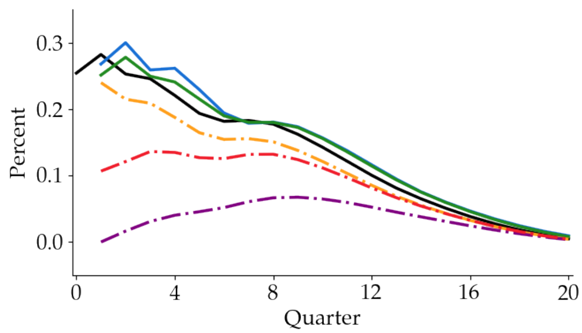
The over-extrapolation model is the model from Equation (16) and whose hair plot is displayed in Figure 2. As shown earlier, the over-extrapolation model is able to rationalize the Bluechip expectations impulse responses across variables, time, and horizons. We can similarly construct impulse responses implied by a number of other models, treating the realized impulse response as the relevant full-information rational expectations (FIRE) benchmark.



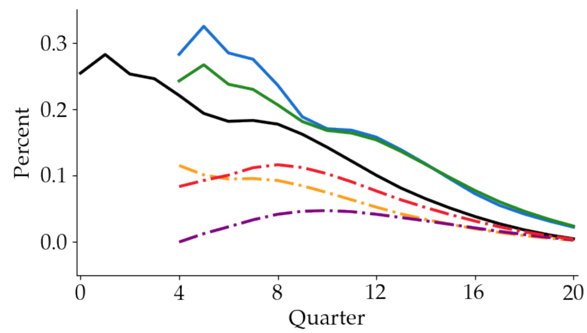
(a) Real disposable income, one-quarter ahead exp.



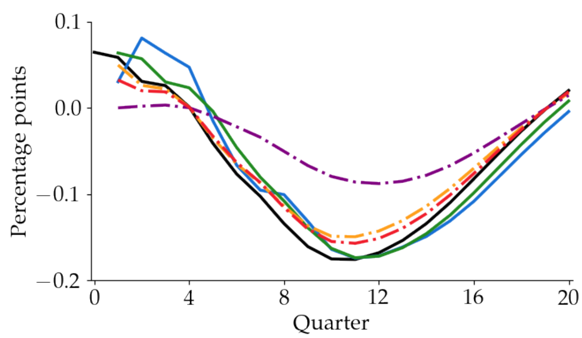
(b) Real disposable income, four-quarters ahead exp.



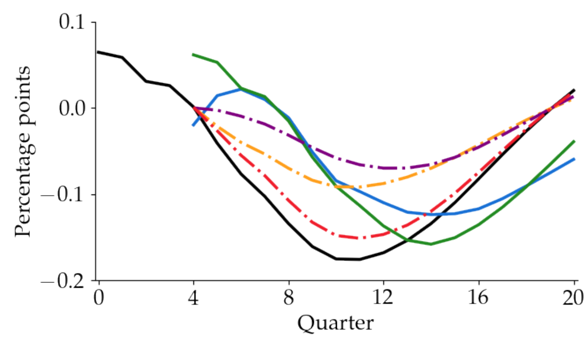
(c) CPI inflation, one-quarter ahead exp.



(d) CPI inflation, four-quarters ahead exp.



(e) Nominal Fed funds rate, one-quarter ahead exp.



(f) Nominal Fed funds rate, four-quarters ahead exp.

Figure 19: Expectation data and model impulse response comparisons

Note: each panel contains impulse response functions of realizations (black) and expectations data or model expectations (color) to a positive Känzig (2021) oil price news shock. The left column contains impulse response functions of one-quarter ahead expectations, and the right column contains analogous responses of four-quarter ahead expectations.

The impulse response implied by the Gabaix (2019) model of cognitive discounting with cognitive discount parameter θ is

$$\Psi(E_t^{\mathrm{CD}}[Y_{t+h}], \varepsilon_{t-\ell}) = \theta^h \underbrace{\Psi(Y_{t+h}, \varepsilon_{t-\ell})}_{\text{Full-information rational expectation IR}}$$

Cognitive discounting implies uniform under-reaction relative to FIRE, where the degree of under-reaction increases with the horizon. Hence, the expectation impulse response under cognitive discounting (gold dot-dashed) under-reacts by more for the four-quarter ahead expectation (right panel) than the one-quarter ahead expectation (left panel). While we see that this under-reaction is largely consistent with Bluechip expectations of real disposable income, it is inconsistent with Bluechip expectations of CPI inflation. Figure 19 uses $\theta = 0.85$ from Gabaix (2019).

The impulse response implied by the Carroll et al. (2020) model of sticky expectations with parameter θ for horizon $h > 0$ is

$$\Psi(E_t^{\text{SE}}[Y_{t+h}], \varepsilon_{t-\ell}) = (1 - \theta^{\ell+1}) \underbrace{\Psi(Y_{t+h}, \varepsilon_{t-\ell})}_{\text{Full-information rational expectation IR}}$$

Sticky expectations implies uniform under-reaction relative to FIRE, where the degree of under-reaction decreases with the time elapsed since the initial shock. While this model of expectation formation struggles to match the CPI inflation expectations for this reason, similarly to cognitive discounting, it also implies too much under-reaction at early periods of the impulse response for each horizon. In contrast to this model, the Bluechip expectations do not exhibit more pronounced under- or over-reaction in early impulse response periods. Figure 19 uses $\theta = 0.935$ estimated in Auclert et al. (2020).

The impulse response implied by the Bordalo et al. (2020) model of dispersed noisy information and diagnostic expectations with parameters θ, τ, the diagnosticity and signal-to-noise precision ratio, is

$$\Psi(E_t^{\text{NIDE}}[Y_{t+h}], \varepsilon_{t-\ell}) = \begin{cases} (1+\theta)\left(\frac{1}{\tau+1}\right)\Psi(Y_{t+h}, \varepsilon_t) & \text{for } \ell = 0 \\ \left((1+\theta)\left(\frac{\ell+1}{\tau+\ell+1}\right) - \theta\left(\frac{\ell}{\tau+\ell}\right)\right)\Psi(Y_{t+h}, \varepsilon_{t-\ell}) & \text{for } \ell > 0 \end{cases}$$

This is obtained by first considering the diagnostic expectation relative to the noisy information rational expectation benchmark, denoted $\tilde{\mathbb{E}}_t$

$$\mathbb{E}_t^{\text{DE}}[Y_{t+h}] = \tilde{\mathbb{E}}_t[Y_{t+h}] + \theta(\tilde{\mathbb{E}}_t[Y_{t+h}] - \tilde{\mathbb{E}}_{t-1}[Y_{t+h}])$$

All agents receive a signal s_t each period about the exogenous impulse response shock $\varepsilon_{t-\ell}$ of the form $s_t = \varepsilon_{t-\ell} + \nu_t$, as in the Appendix of Auclert et al. (2020). Let τ denote the ratio of (constant) signal precision (inverse standard deviation of ν_t) to the precision of $\varepsilon_{t-\ell}$. Then the impulse response of the horizon-h noisy information rational expectation can be written as

$$\Psi(\tilde{\mathbb{E}}_t[Y_{t+h}], \varepsilon_{t-\ell}) = \frac{\ell+1}{\tau+\ell+1}\Psi(Y_{t+h}, \varepsilon_{t-\ell})$$

On the initial shock impact period when $\ell = 0$, past expectations $\mathbb{E}_{t-1}$ are still anchored at 0, hence the $\ell = 0$ case in the above impulse response of the noisy information, diagnostic expectation. However, after $\ell > 0$, the full-information rational expectation of the prior referenced period will fully adjust, hence $\mathbb{E}_t = \mathbb{E}_{t-1}$ for $\ell > 0$.

I use estimated values of θ, τ from Bordalo et al. (2020) for the noisy information, diagnostic expectation of variables plotted in Figure 19. For real disposable income expectations, I use the

		Consumption-Savings Models		
Extrapolation	Parameter	Perpetual Youth	Standard Incomplete Markets	Rep. agent
AR(2)	EIS	0.08	0.09	0.00
	MPC	0.04	0.05	0.005
AR(1)	EIS	0.11	0.07	0.00
	MPC	0.05	0.07	0.005
AR(1) of AR(1)s	EIS	0.06	0.07	0.00
	MPC	0.05	0.07	0.005
"Over-extrapolation"	EIS	0.01	0.10	0.00
	MPC	0.03	0.07	0.005

Table 4: Estimated parameters across missing-horizon extrapolation models

Note: Each panel contains of estimated parameters for each consumption-savings model under different extrapolation methods for missing expectations data horizons. The parameter estimates enforce that the steady state assets-to-income ratio is equal to the initial calibration target.

estimated values in Bordalo et al. (2020) for Bluechip real GDP growth expectations. They estimate θ, τ for CPI inflation and the nominal Federal funds rate expectations from the Bluechip, so I use exactly those values for these variables.

While in principle diagnostic expectations can produce over-reaction, due to the diagnosticity parameter θ, in Bordalo et al. (2020) the values of τ are sufficiently large that the average expectation does not display over-reaction, as it would with pure diagnostic expectations as in Bordalo et al. (2018).

A.3 Extrapolating missing horizons of expectations data

In Section 3.4, I discussed the need to extrapolate the missing horizons of expectations data. While the Bluechip expectations data has forecasts for a finite number of horizons, we need an infinite set of horizons of expectations to evaluate model-implied consumption. To obtain these missing horizons, I estimate an auxiliary, parametric model on the existing horizons and impulse response periods and use it to extrapolate the missing horizons.

Using two-stage least squares where moments in Equation (10) are targeted with weights given by the inverse covariance matrix of the sequence of instruments $\{z_{t-\ell}\}$, I estimate the following auxiliary models, which result in the parameter estimates I report in Table 4. I also impose an additional penalty on each model to ensure that the far-horizon expectations implied by each model are stationary, that is

$$\lim_{h\to\infty} \mathbb{E}[F_t[W_{t+h}; \boldsymbol{\vartheta}] z_{t-\ell}] = 0$$

Note that the top panel of Table 4 is the baseline extrapolation I choose in the main set of results reported in Table 3 in the main body of the paper.

The AR(2) and AR(1) are the standard univariate autoregressive processes, with two and one period lags respectively. The "AR(1) of AR(1)s" is a single lag autoregressive process, whose innovation term is also an AR(1) process. This functional form choice is motivated by the functional form of the equilibrium law of motion of output Y_t in Section 4, where the belief component law of motion follows a vector equivalent of a "AR(1) of AR(1)s". Finally, the "over-extrapolation" model is the one given by Equation (16) and displayed in Figure 2.

B Proofs and derivations

B.1 Proof of Proposition 1

Let the persistent shock component λ_t and the transitory shock component η_t each follow AR(1) processes

$$\begin{aligned}\lambda_t &= \rho_\lambda \lambda_{t-1} + u_{\lambda,t} \\ \eta_t &= \rho_\eta \eta_{t-1} + u_{\eta,t}\end{aligned}$$

The perceived law of motion $\tilde{Y}_t$ and equilibrium law of motion of output Y_t are given by

$$\begin{aligned}\tilde{Y}_t &= \lambda_t + \eta_t \\ Y_t &= \chi(h_\lambda E_{t-1}[\lambda_t] + h_\eta E_{t-1}[\eta_t]) + \lambda_t + \eta_t\end{aligned}$$

where $h_\lambda := \frac{\rho_\lambda}{1-\beta\omega\rho_\lambda}$ and h_η is analogously defined with respect to ρ_η. Beliefs about each shock component evolve according to the Kalman update equation

$$\begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \begin{bmatrix} \rho_\lambda & 0 \\ 0 & \rho_\eta \end{bmatrix} \begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} + \begin{bmatrix} g_\lambda \\ g_\eta \end{bmatrix} (Y_t - E_{t-1}[Y_t])$$

where the subjective expectation $E_{t-1}[Y_t] = E_{t-1}[\lambda_t] + E_{t-1}[\eta_t]$ is the conditional expectation of output induced by the perceived law of motion, given the history of past output observations $\{Y_{t-\ell}\}_{\ell>0}$, and g_λ, g_η are the steady state Kalman gains under the perceived law of motion.

Evaluating and re-organizing terms in the belief component law of motion, we obtain

$$\begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \underbrace{\left(\begin{bmatrix} \rho_\lambda & 0 \\ 0 & \rho_\eta \end{bmatrix} + \begin{bmatrix} g_\lambda(\chi h_\lambda - 1) & g_\lambda(\chi h_\eta - 1) \\ g_\eta(\chi h_\lambda - 1) & g_\eta(\chi h_\eta - 1) \end{bmatrix} \right)}_{\text{Let } \mathbf{A}:=} \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix} + \underbrace{\begin{bmatrix} g_\lambda & g_\lambda \\ g_\eta & g_\eta \end{bmatrix}}_{\text{Let } \mathbf{G}:=} \begin{bmatrix} \lambda_t \\ \eta_t \end{bmatrix} \tag{26}$$

where I require the eigenvalues of $\mathbf{A}$ to be within the unit circle, such that the belief component law of motion is stationary.

Inertia at time-1 for Y_t to exhibit inertia with respect to an innovation to a component shock, the net increase in belief feedback at time-1 must exceed the decay from the direct effect of the component shock.

Consider a time-0 positive innovation to a component shock $e_0 > 0$, where $e_0 \in \{\lambda_0, \eta_0\}$. At time-0, only the direct shock effect occurs so $Y_0 = e_0$, given beliefs prior to time-0 are zero in steady state. At time-1, we have

$$\begin{bmatrix} E_0[\lambda_1] \\ E_0[\eta_1] \end{bmatrix} = \begin{bmatrix} g_\lambda \\ g_\eta \end{bmatrix} e_0$$

Evaluating Y_1 and solving for the threshold $Y_1 > Y_0$, we obtain the lower bound

$$\chi > \frac{1-\rho_e}{h_\lambda g_\lambda + h_\eta g_\eta} > 0 \tag{27}$$

Let us denote this lower threshold for χ as $\underline{X}_{e,0}$. Given the range of permissible parameters, where $\beta, \omega, \rho_e \in (0,1)$, it must be that this threshold is strictly positive, i.e. $\underline{X}_{e,0} > 0$. Thus, if $\chi > \underline{X}_{e,0}$, then Y_t will exhibit inertia with the inertial peak period $\bar{\ell} \geq 1$.

Regularity conditions on A A necessary condition for Y_t to be increasing with time-t up until an inertial peak period $\bar{\ell}$, is for the belief feedback term, $\chi(h_\lambda E_{t-1}[\lambda_t] + h_\eta E_{t-1}[\eta_t])$, to be increasing with time-t. I now derive some restrictions on parameters, or equivalently regularity conditions on $\mathbf{A}$, that ensures that the belief feedback term $\chi(h_\lambda E_{t-1}[\lambda_t] + h_\eta E_{t-1}[\eta_t])$ is positive for all time-t. For times-t leading up to the inertial peak period $\bar{\ell}$ this is required for the necessary condition to hold[18].

Positive, real eigenvalues of A

For the eigenvalues of $\mathbf{A}$ to be real (note that $\mathbf{A}$ is not positive semi-definite), the discriminant of its characteristic polynomial must be positive. This can be simplified to checking whether the following expression is greater than zero

$$(\rho_\lambda - \rho_\eta)^2 + (\theta_\lambda + \theta_\eta)^2 + 2(\rho_\lambda - \rho_\eta)(\theta_\eta - \theta_\lambda) > 0$$

where $\theta_\lambda := g_\lambda(1 - \chi h_\lambda)$ and likewise for θ_η. Given our definitions of $\chi := (1 - \beta\omega - \beta\omega\sigma\phi)$ and the range of permissible parameters, it must be that $\theta_\lambda, \theta_\eta > 0$. Hence, simplifying the above expression in terms of lower bound on χ, we obtain

$$\chi \geq \frac{g_\lambda - g_\eta}{h_\lambda g_\lambda - h_\eta g_\eta} \tag{28}$$

which ensures the eigenvalues of $\mathbf{A}$ are real.

For the eigenvalues to be positive, we have

$$\begin{aligned} w_1 + w_2 &= \rho_\lambda + \rho_\eta - \theta_\lambda - \theta_\eta \\ w_1 w_2 &= \rho_\lambda \rho_\eta - \rho_\lambda \theta_\eta - \rho_\eta \theta_\lambda \end{aligned}$$

[18] This is potentially stronger than necessary for bounding the absolute value of the Wold coefficients of Y_t for times-$t > \bar{\ell}$ to be less than the time-$\bar{\ell}$ coefficient. However, another way to justify the strength of these regularity conditions for periods after $\bar{\ell}$ is if we desire Y_t to return to its steady state at zero without "over-shooting" and becoming negative in response to a positive component shock. If so, these regularity conditions will ensure this.

The following lower bounds on χ ensure that the eigenvalues will be positive

$$\chi \geq \frac{g_\lambda - \frac{1}{2}\rho_\lambda}{h_\lambda g_\lambda}, \quad \chi \geq \frac{g_\eta - \frac{1}{2}\rho_\eta}{h_\eta g_\eta} \tag{29}$$

with at least one holding as a strict inequality.

Effective horizons, gains $\mathbf{h}, \mathbf{g}'$ and eigenvectors of $\mathbf{A}$

Given the χ lower bound in (27) and $\theta_\lambda, \theta_\eta > 0$, the matrix $\mathbf{A}$ will have positive diagonal entries and negative off-diagonal entries. Due to this, the eigenvector v_1 corresponding to the dominant eigenvalue $w_1 > w_2$ will have one positive $v_{11} > 0$ and one negative entry $v_{12} < 0$. The non-dominant eigenvector v_2 will solely have positive entries, i.e. $v_{21}, v_{22} > 0$.

Consider the following expression, where we can unwind the recursion in Equation (26) given the initial component shock e_0

$$\begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \sum_{\ell=0}^{t} \mathbf{A}^{t-\ell} \mathbf{g}' \rho_e^\ell e_0$$

Left-multiplying by $\mathbf{h}$ to compute the contribution of belief feedback and expanding the eigen-decomposition of $\mathbf{A} = \mathbf{V}\mathbf{W}\mathbf{V}^{-1}$, we obtain

$$\mathbf{h} \begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \mathbf{h}\mathbf{V} \sum_{\ell=0}^{t} \mathbf{W}^{t-\ell} \mathbf{V}^{-1} \mathbf{g}' \rho_e^\ell e_0$$

With the conditions

$$\frac{h_\lambda}{h_\eta} > -\frac{v_{12}}{v_{11}}, \quad \frac{g_\lambda}{g_\eta} < \frac{v_{22}}{v_{21}} \tag{30}$$

the above expression constitutes the sum of strictly positive, bilinear forms which itself must be positive, hence we will have as desired

$$\mathbf{h} \begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} > 0, \quad \forall t$$

Without condition (30), we cannot ensure the belief feedback term is positive for any time-t.

Inertia at time-t I now proceed with the induction to prove that the inertial peak $\bar{\ell}$ is (weakly) increasing in χ, assuming conditions (27), (28), (29), (30) hold.

Suppose $Y_{t-\ell} > Y_{t-\ell-1}$ for $\ell \in \{0, t\}$, and at time-$t+1$ we have $Y_{t+1} = Y_t$, placing the inertial peak $\bar{\ell} = t$. The equality $Y_{t+1} = Y_t$ can be written as

$$\chi(h_\lambda \Delta E_t[\lambda_{t+1}] + h_\eta \Delta E_t[\eta_{t+1}]) = (1 - \rho_e)\rho_e^t e_0 \tag{31}$$

where $\Delta E_t[\lambda_{t+1}] := E_t[\lambda_{t+1}] - E_{t-1}[\lambda_t]$ and likewise for η.

Our goal is to demonstrate that as χ increases the left-hand side exceeds the right-hand side which is invariant to χ, thus shifting the inertial peak to $\bar{\ell} = t+1$. If Equation (31) held with an inequality $<$, then a marginal increase in χ would not shift the peak, hence this result only implies $\bar{\ell}$ is weakly increasing in χ.

Differentiating the left-hand side of Equation (31), we obtain

$$\overbrace{h_\lambda \Delta E_t[\lambda_{t+1}] + h_\eta \Delta E_t[\eta_{t+1}]}^{=\chi^{-1}(1-\rho_e)\rho_e^t e_0 > 0} + \chi(h_\lambda \partial_\chi \Delta E_t[\lambda_{t+1}] + h_\eta \partial_\chi \Delta E_t[\eta_{t+1}])$$

Given $\chi > 0$ by Equation (27), it suffices to verify $h_\lambda \partial_\chi \Delta E_t[\lambda_{t+1}] + h_\eta \partial_\chi \Delta E_t[\eta_{t+1}] > 0$.

Differentiating the component beliefs in Equation (26) with respect to χ, we obtain

$$\begin{bmatrix} \partial_\chi E_t[\lambda_{t+1}] \\ \partial_\chi E_t[\eta_{t+1}] \end{bmatrix} = \mathbf{A} \begin{bmatrix} \partial_\chi E_{t-1}[\lambda_t] \\ \partial_\chi E_{t-1}[\eta_t] \end{bmatrix} + \mathbf{g}'\mathbf{h} \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix}$$

Differencing, unwinding the recursion, and left-multiplying again by $\mathbf{h}$ we obtain

$$\mathbf{h} \begin{bmatrix} \partial_\chi \Delta E_t[\lambda_{t+1}] \\ \partial_\chi \Delta E_t[\eta_{t+1}] \end{bmatrix} = \mathbf{h}\mathbf{V}\mathbf{W}^{t-1}\mathbf{V}^{-1}\mathbf{g}'\mathbf{h} \begin{bmatrix} E_0[\lambda_1] \\ E_0[\eta_1] \end{bmatrix} + \mathbf{h}\mathbf{V} \sum_{\ell=1}^{t-1} \mathbf{W}^{t-1-\ell}\mathbf{V}^{-1}\mathbf{g}'\mathbf{h} \begin{bmatrix} \Delta E_t[\lambda_{t+1}] \\ \Delta E_t[\eta_{t+1}] \end{bmatrix} > 0$$

Finally, given the $\chi > 0$ by Equation (27) and the above expression, we have

$$\chi(h_\lambda \partial_\chi \Delta E_t[\lambda_{t+1}] + h_\eta \partial_\chi \Delta E_t[\eta_{t+1}]) > 0$$

which implies the left-hand side of Equation (31) is strictly increasing in χ.

B.2 Proof of Proposition 2

Given η_t is i.i.d, the output law of motion under rational learning is given by

$$Y_t^R = \frac{\chi h_\lambda}{1 - \chi h_\lambda} E_{t-1}[\lambda_t] + \lambda_t + \eta_t \tag{32}$$

and the output law of motion under constrained-rational learning is given by

$$Y_t^{CR} = \underbrace{\chi h_\lambda \, (1-\rho_\lambda^2) \sum_{\ell=0}^{\infty} \rho^\ell a_\ell}_{\text{Denote } \tilde{a} :=} \, E_{t-1}[\lambda_t] + \lambda_t + \eta_t \tag{33}$$

a_ℓ denote the Wold coefficients of Y_t^{CR} with respect to persistent component innovations $u_{\lambda,t-\ell}$.

Given η_t is i.i.d, the component belief law of motion for λ_t under rational learning is given by

$$E_t[\lambda_{t+1}] = \underbrace{(\rho_\lambda - g_\lambda)}_{\text{Denote } f_\lambda :=} E_{t-1}[\lambda_t] + g_\lambda(\lambda_t + \eta_t) \tag{34}$$

and under constrained-rational learning

$$E_t[\lambda_{t+1}] = \underbrace{(\rho_\lambda - \tilde{g}_\lambda \tilde{a}(1 - \chi h_\lambda))}_{\text{Denote } \tilde{f}_\lambda :=} E_{t-1}[\lambda_t] + \tilde{g_\lambda}(\lambda_t + \eta_t) \tag{35}$$

To simplify Equation (35), compare the Kalman gains $g_\lambda, \tilde{g}_\lambda$ under the two different perceived state space models. The perceived law of motion under rational learning yields the measurement equation

$$Y_t^R = \frac{\chi h_\lambda}{1-\chi h_\lambda} E_{t-1}[\lambda_t] + \lambda_t + \eta_t$$

Under constrained-rational learning, the measurement equation is

$$Y_t^{CR} = \tilde{a}\lambda_t + \eta_t$$

Given $E_{t-1}[\lambda_t]$ is pre-determined in Y_t^R and the state transition equations under rational and constrained-rational learning are known to be the same, the steady state variance of the one-step ahead prediction errors p is the same. Applying the steady-state Kalman gain formula under constrained-rational learning and setting $\sigma_\eta^2 = \frac{\tilde{\sigma}_\eta^2}{\tilde{a}}$ obtains $\tilde{g}_\lambda \tilde{a} = g_\lambda$.

Denote the Wold decompositions of Y_t^R, Y_t^{CR} as

$$Y_t^R = \sum_{\ell=0}^{\infty} a_\ell^R u_{\lambda,t-\ell} + b_\ell^R u_{\eta,t-\ell}$$
$$Y_t^{CR} = \sum_{\ell=0}^{\infty} a_\ell^{CR} u_{\lambda,t-\ell} + b_\ell^{CR} u_{\eta,t-\ell}$$

Y_t responses to a transitory innovation $u_{\eta,0}$

Iterating Equations (34), (35) forward and plugging into (32), (33), we obtain the expressions for the Wold coefficients with respect to a transitory innovation t-periods ago

$$b_t^R = \frac{\chi h_\lambda}{1-\chi h_\lambda} g_\lambda f_\lambda^{t-1}$$
$$b_t^{CR} = \chi h_\lambda g_\lambda \tilde{f}_\lambda^{t-1}$$

The inequality $b_t^R > b_t^{CR}$ can be written as

$$\frac{1}{1-\chi h_\lambda}\left(\frac{f_\lambda}{\tilde{f}_\lambda}\right)^{t-1} > 1 \tag{36}$$

Therefore, in the initial period (time-1) in which beliefs respond the inequality will hold because $(1-\chi h_\lambda)^{-1} > 1$, given $1-\chi h_\lambda \in (0,1)$ for $\chi > 0$. However, as $t \to \infty$, the left-hand side approaches zero because the ratio $\frac{f_\lambda}{\tilde{f}_\lambda} \in (0,1)$. This indicates that at some positive period $\bar{t} > 1$ the inequality will no longer hold.

Y_t responses to a persistent innovation $u_{\lambda,0}$ Iterating Equations (34), (35) forward and plugging into (32), (33), we obtain the expressions for the Wold coefficients with respect to a persistent innovation t-periods ago

$$a_t^R = \frac{\chi h_\lambda}{1-\chi h_\lambda} g_\lambda \sum_{\ell=0}^{t-1} f_\lambda^{t-1-\ell}\rho_\lambda^\ell + \rho_\lambda^t$$
$$a_t^{CR} = \chi h_\lambda g_\lambda \tilde{a} \sum_{\ell=0}^{t-1} \tilde{f}_\lambda^{t-1-\ell}\rho_\lambda^\ell + \rho_\lambda^t$$

The expression for a_t^{CR} is implicit, since $\tilde{a}$ contains Wold coefficients $\{a_\ell\}_{\ell\geq 0}$. Unwinding the implicit expression in $\tilde{a}$ and solving out the following sum

$$\sum_{\ell=0}^{t-1} f_\lambda^{t-1-\ell}\rho_\lambda^\ell = \frac{\rho_\lambda^t - f_\lambda^t}{\rho_\lambda - f_\lambda}$$

and likewise for $\tilde{f}_\lambda$, we obtain

$$a_t^R = \frac{\chi h_\lambda}{1-\chi h_\lambda} g_\lambda \left(\frac{\rho_\lambda^t - f_\lambda^t}{\rho_\lambda - f_\lambda}\right) + \rho_\lambda^t$$
$$a_t^{CR} = \chi h_\lambda g_\lambda \left(\frac{\rho_\lambda}{1-\tilde{f}_\lambda\rho_\lambda}\chi h_\lambda g_\lambda + 1\right)\left(\frac{\rho_\lambda^t - \tilde{f}_\lambda^t}{\rho_\lambda - \tilde{f}_\lambda}\right) + \rho_\lambda^t$$

Setting the inequality $a_t^R > a_t^{CR}$ and simplifying we obtain

$$\frac{\rho_\lambda^t - f_\lambda^t}{\rho_\lambda^t - \tilde{f}_\lambda^t} > 1 + \frac{\rho_\lambda}{1-\tilde{f}_\lambda\rho_\lambda}\chi h_\lambda g_\lambda$$

Given $0 < f_\lambda < \tilde{f}_\lambda < \rho_\lambda$, the left-hand side is positive and increasing in time-t. Checking the inequality at time-1, I obtain

$$\frac{1}{1-\chi h_\lambda} > \frac{\rho_\lambda}{1-\tilde{f}_\lambda\rho_\lambda}\chi h_\lambda g_\lambda + 1$$

which further simplifies to $1 > \rho_\lambda^2$, which always holds for the cases we consider $\rho_\lambda \in (0,1)$. Therefore, $Y_t^R > Y_t^{CR}$ for all times-t.