Explaining the Macroeconomic Inertia Puzzle

Michael Cai

July 2025

Click here for the most recent version

Abstract

Benchmark macroeconomic models require additional frictions to explain the sluggish response of aggregate variables to sudden shocks or changes in policy. I show that standard heterogeneous-agent (HA) models—the Blanchard (1985) perpetual youth and Bewley (1986) incomplete markets models—are consistent with aggregate consumption inertia without the use of habit preferences or any specific model of expectation underreaction to dampen the responsiveness of consumption-savings decisions. I instead replicate observed consumption inertia in standard HA models by directly substituting survey expectations of income and interest rates for agents' expectations. I propose a new theory of macroeconomic inertia that rationalizes the observed extrapolation bias in survey expectations by embedding an unobserved components model of expectations into a tractable HA general equilibrium environment. Inertia results when expectations imperfectly account for the equilibrium amplification of shocks, which is large in HA economies. This imperfect inference causes expectations to gradually unanchor as agents repeatedly misattribute large responses of equilibrium outcomes simply to larger shocks. This theory also illustrates 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 due to their decreased effectiveness at anchoring expectations.

*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 and Society for Economic Dynamics 2025 meeting 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 display a sluggish response to sudden shocks or 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, the benchmark model for policy analysis, cannot rationalize these slow responses and instead predicts that policy actions should have immediate effects. Theories of adjustment frictions and bounded rationality have been proposed to limit the agents' foresight and the responsiveness of their decisions, two features preventing the textbook model from generating inertia. However, no consensus has emerged among these theories.

This paper focuses on explaining the observed inertia in aggregate consumer spending because it plays a central role in facilitating policy transmission. I first adopt a semi-structural approach to analyze consumption-savings models without imposing restrictive assumptions on expectation formation or altering preferences. My goal is to identify whether the basic structure of these models is consistent with aggregate consumption inertia without taking a stand on how expectations are formed. To do so, I first compute the empirical impulse responses of survey expectations and realizations of income and interest rates to an identified shock. I then evaluate 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 standard heterogeneous-agent models[2] can match observed aggregate consumption inertia as long as agents' average expectations of income and interest rates align with average survey expectations of these variables. In contrast, 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 (MPCs).

The consistency of heterogeneous-agent models with aggregate consumption inertia does not explain the underlying causes of inertia. Additionally, patterns of extrapolation bias that I document in survey expectations are not easily explained by many existing models of expectation formation. I therefore proceed to propose a new theory of macroeconomic inertia that rationalizes expectations data by embedding an unobserved components model of expectations into a tractable heterogeneous-agent equilibrium environment.

My second main result explains why high MPCs in heterogeneous-agent models contribute to the emergence of inertia. Agents learn by observing equilibrium output (income) but cannot perfectly separate the contributions of the direct effect of an exogenous demand shock from the ensuing Keynesian amplification, which is large when the average MPC is large. As agents overweight the direct impact of the shock and optimistically consume more, equilibrium output is further magnified, which partially confirms their initial beliefs. Inertia arises due to the repeated misattribution and delayed amplification that results from this feedback loop. As a consequence of imperfect learning, policy regimes which delay action and rely on far-horizon reasoning through equilibrium outcomes become less effective.

[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]Perpetual youth, overlapping generations (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 expectation formation 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 bounded rationality that uniformly predict under- or overreaction relative to the full-information rational expectations cannot explain 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. If the path of income displays a hump-shaped profile, over-extrapolation entails initial underreaction and eventual overreaction. This over-extrapolation can also 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

[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 4.4 and also in Appendix A.3.

formation, disciplined to the expectations data, and a standard New Keynesian equilibrium environment. Expectations are formed via 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 from the current observation.

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). 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 solving and estimating structural models following Manski (2004), Manski (2018). Piazzesi and Schneider (2009) use expectations of housing market and business conditions to characterize boom-bust cycles in a search model of housing transactions. 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.

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.

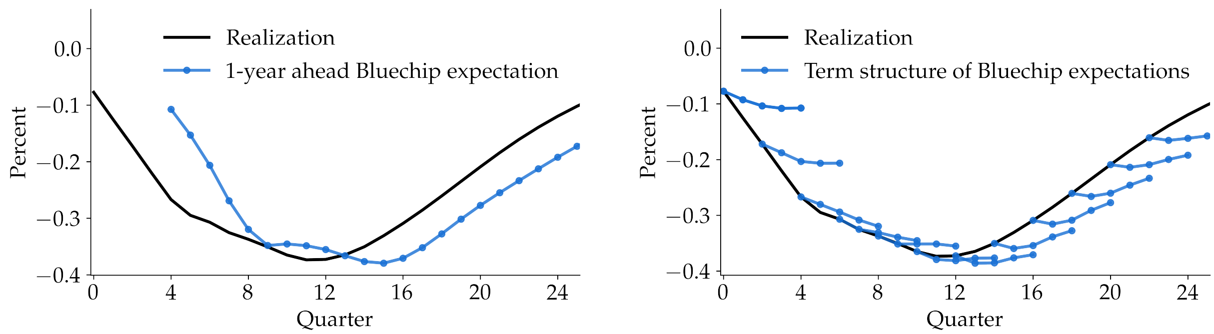


Figure 1: Realized and expected real output responses to a Känzig (2021) oil shock

Note: Both panels plot the realized real output impulse response (black) to a positive Känzig (2021) oil supply news shock, which raises real oil prices by ten percent on impact. The left panel plots the impulse response function of the consensus Bluechip one-year ahead expectation of real output (blue). The right panel plots the impulse responses of the one-quarter through one-year ahead expectations (blue dots), connected across horizons as opposed to time.

2 Impulse responses of measured expectations

I begin by documenting empirical patterns of impulse responses of measured expectations from survey data and their systematic biases. I demonstrate how existing theories that deviate from the full-information rational expectations (FIRE) benchmark struggle to rationalize these patterns. This observation highlights the usefulness of the semi-structural estimation method that I propose in the next two sections, where I directly condition on the path of measured expectations in place of a model of expectation formation to assess the fit of structural models of consumption-savings against consumption data. I then offer supportive evidence for a form of extrapolation bias, which can rationalize the empirical patterns of measured expectations.

As discussed in Kučinskas and Peters (2022) and Angeletos et al. (2021), impulse responses of measured expectations can be useful moments to diagnose biases in expectation formation. In the left panel of Figure 1, I plot the impulse response of realized and one-year ahead expected real output to an oil supply news shock from Känzig (2021), using a similar structural VAR specification and realized data as in Känzig (2021) and augmented with consensus survey expectations data from the Bluechip Economic Indicators and Financial Forecasts. I postpone discussion of the details of the data and the relevance of this particular choice of shock until Section 4.

Real output displays an inertial contraction reaching its peak response after nearly twelve quarters in response to an oil supply news shock that raises real oil prices on impact. One-year ahead expectations initially underreact, with the impact period one-year ahead output expectation falling to -0.1 percent compared to the ex-post realization of -0.3 percent. However, similarly to the response of other macroeconomic variables to other shocks as in Angeletos et al. (2021), output expectations here eventually overreact after twelve quarters.

This pattern of delayed overreaction holds for the responses of expectations across various horizons, as shown in the right panel of Figure 1, where each blue dot extending from the black line represents the impulse response coefficient of the one-quarter through one-year ahead real output expectations.

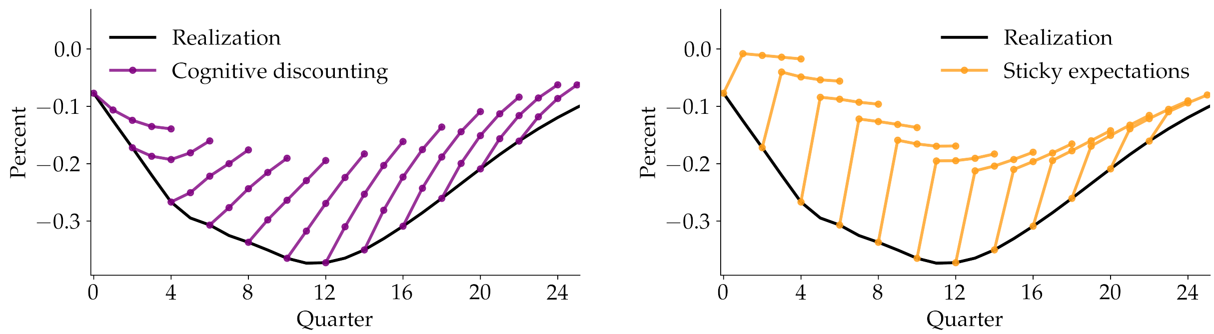


Figure 2: Systematic underreaction in expected real output responses to a Känzig (2021) oil shock

Note: In addition to the realized real output impulse response (black), both figures plot the one-quarter to one-year term structure of output expectations implied by two models of expectation underreaction, cognitive discounting (purple) from Gabaix (2020) and sticky expectations (gold) from Carroll et al. (2020).

Models of expectation underreaction have frequently been employed to rationalize inertia in realized outcomes. The economic rationale is that underreaction by definition dampens the initial responsiveness of economic agents, for example the consumption responses of households at the onset of a contractionary supply shock. However, over time as the contractionary output surprises continue this causes households to further spend down their wealth, causing output fall to its eventual trough.

Let us define the realized (linear) impulse response function of a variable Y_t as

$$\Psi(Y_t; \varepsilon_{t-\ell}) = \mathbb{E}[Y_t | \varepsilon_{t-\ell} = 1, \mathbf{X}_{t-\ell}] - \mathbb{E}[Y_t | \varepsilon_{t-\ell} = 0, \mathbf{X}_{t-\ell}]$$

Likewise we can define the response of a subjective expectation $E_t[Y_{t+h}]$ as

$$\Psi(E_t[Y_{t+h}]; \varepsilon_{t-\ell}) = \mathbb{E}[E_t[Y_{t+h}] | \varepsilon_{t-\ell} = 1, \mathbf{X}_{t-\ell}] - \mathbb{E}[E_t[Y_{t+h}] | \varepsilon_{t-\ell} = 0, \mathbf{X}_{t-\ell}]$$

If $E_t[Y_{t+h}]$ is the rational expectation $\mathbb{E}_t[Y_{t+h}]$ then by the law of iterated expectations, we have

$$\Psi(Y_{t+h}; \varepsilon_{t-\ell}) = \Psi(\mathbb{E}_t[Y_{t+h}]; \varepsilon_{t-\ell})$$

Figure 2 plots the behavior of the term structure of expectations implied by two common models of expectation underreaction: cognitive discounting as in Gabaix (2020) and sticky expectations as in Carroll et al. (2020). These expectations were produced by applying each behavioral bias to the realized impulse response path of real output, which we treat as the rational expectations benchmark [8].

Because cognitive discounting and sticky information are constructed off of a rational expectations benchmark, where cognitive discounting anchors more to zero across horizons-h

[8]This relies on the assumption that the estimated impulse response from our empirical specification is the efficient linear conditional expectation of real output given the observable variables included in the specification and the realization of the exogenous shock. If the data were generated from a structural model, which yielded a given structural moving average representation with respect to the exogenous shocks, then this assumption relies on our empirical specification being able to accurately recover this representation.

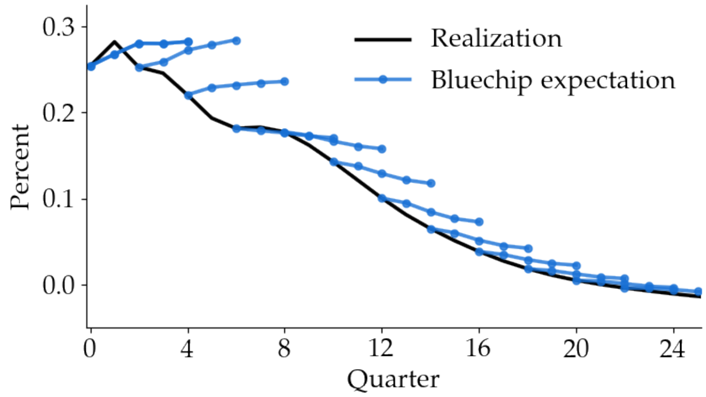


Figure 3: Realized and expected consumer price response to a Känzig (2021) oil shock

Note: This figure displays the consumer price response (black) to a positive Känzig (2021) oil supply news shock, which raises real oil prices by ten percent on impact, and the impulse responses of the one-quarter through one-year ahead expectations (blue dots), connected across horizons.

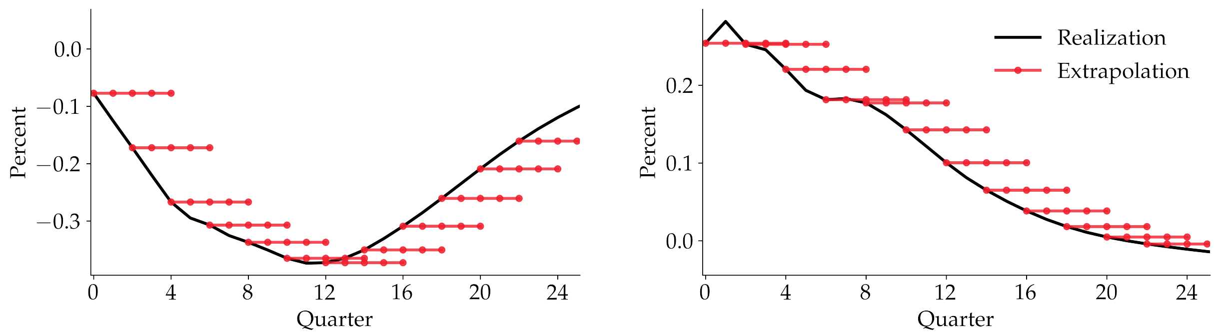
and sticky expectations anchors less to zero as the time-ℓ following the shock passes, we can manipulate the time-t rational expectation of $t+h$ impulse response, or equivalently the $t+h$ realization impulse response, to construct the response of expectations implied by these models of underreaction. The impulse responses of subjective expectations from these models are then given by

$$\Psi(E_t^{\text{CD}}[Y_{t+h}]; \varepsilon_{t-\ell}) = \theta^h \Psi(Y_{t+h}; \varepsilon_{t-\ell})$$
$$\Psi(E_t^{\text{SE}}[Y_{t+h}]; \varepsilon_{t-\ell}) = (1 - \vartheta^{\ell+1}) \Psi(Y_{t+h}; \varepsilon_{t-\ell})$$

where Figure 2 uses a cognitive discount $\theta = 0.85$ as in Gabaix (2020) and the sticky expectation $\vartheta = 0.935$ as estimated in Auclert et al. (2020).

While both of these models of expectation underreaction are capable of generating inertia in realized outcomes when embedded in a standard DSGE model, they are unable to explain key patterns of bias in measured expectations shown in Figure 1, most notably the eventual overreaction in measured expectations. Existing models that have been proposed to explain delayed overreaction combine elements of models of pure overreaction, such as baseline diagnostic expectations (Bordalo et al. 2018), with models of underreaction, where the underreaction bias disappears over time. Examples include combinations of noisy information and diagnostic expectations (Bordalo et al. 2020, Bardóczy and Guerreiro 2023) or over-persistence bias (Angeletos et al. 2021). These combined models of underreaction and overreaction, however, are unable to replicate patterns of pure overreaction in the response of other variables, such as the consumer prices plotted in Figure 3, which has also independently been documented in experimental settings (Afrouzi et al. 2023).

While differing degrees of under- and overreaction across variables and over time are difficult to rationalize with many existing models of expectation formation, models that imply a form of extrapolation bias are able to explain observed patterns of forecast errors. Figure 4 illustrates the explanatory power of extrapolation bias, displaying the realized real output and



(a) Real output

(b) Consumer price index

Figure 4: Extrapolation in the response of expectations to a Känzig (2021) oil shock

Note: The panels display the real output and consumer price response (black) to a positive Känzig (2021) oil supply news shock, which raises real oil prices by ten percent on impact. Each panel additionally plots the time-t one-quarter through one-year ahead expectations generated by extrapolating forward the time-t realized impulse response (red).

consumer price impulse responses alongside a subjective expectation constructed by mechanically extrapolating forward the current realization response, that is

$$\Psi(E_t^{\text{Ex}}[Y_{t+h}]; \varepsilon_{t-\ell}) = \Psi(Y_t; \varepsilon_{t-\ell}) \text{ for all } h > 0$$

The key insight to replicating both delayed overreaction in real output expectations and pure overreaction in consumer price expectations rests on the shape of the underlying realized impulse response. For hump-shaped realized responses, extrapolation results in delayed overreaction, but for realized responses that are peaked on impact and monotonically declining thereafter, extrapolation results in pure overreaction.

Hence, existing theories of expectation formation, which relied purely on underreaction to generate inertia in realized outcomes, may have done so at the cost of missing important features of measured expectations, namely extrapolation bias. Instead of taking a stand on a particular theory of expectation formation, I next demonstrate how one can evaluate the aggregate consumption path implied by a given structural model of consumption-savings by conditioning directly on the paths of measured expectations and realizations of income and interest rates.

3 Model-implied impulse responses using expectations data

In this section I construct moment conditions, which represent the distance between unrestricted, empirical impulse responses and structural model-implied impulse responses to an externally identified exogenous shock. The main innovation relative to existing approaches, such as Euler equation estimation in Hansen and Singleton (1982), Barnichon and Mesters (2020), Lewis and Mertens (2022), is the use of a linearized, aggregate, sequential representation of the structural model, which removes the need to impose rational expectations or any other theory of expectation formation. In lieu of a model of expectation formation, I directly use expectations data of the decision-relevant variables, such as income and interest rates for consumption, to evaluate the moment condition. I first illustrate this approach in the familiar representative-agent example and then proceed to the heterogeneous-agent case.

3.1 Representative-agent example

Consider a representative household consumption-savings problem given arbitrary, subjective expectations $E_t[\cdot]$ of future income and interest rates Y_s, r_s for $s > t$.

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

The household consumes C_t and saves in a one-period, risk-free asset A_t, taking income and interest rates as given. Additionally, their choices are subject to an exogenous discount factor shock ζ_t, which serves as an unobserved source of endogeneity from the perspective of the econometrician, who is aiming to estimate the structural elasticity of substitution parameter, σ.

Taking first-order conditions and linearizing around the steady-state $\beta(1 + r) = 1$, we obtain the subjective expectations consumption function, where $W_t := r_{t-1}A + (1 + r)A_{t-1}$ denotes financial wealth, $\gamma := \sigma\beta - (1 - \beta)A$ denotes the net interest rate elasticity, and ε_t is the exogenous "demand shock", which is a function of the discount factor shock ζ_t.

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

I 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 structural model given in Equation (1) to match the unrestricted, empirical impulse response of consumption C_t to an externally-identified, exogenous shock z_t, obtained from a local projection or vector auto-regression specification.

One classic approach in the literature, as used in Christiano et al. (2005), is "impulse response matching", which imposes three key assumptions to estimate Equation (1). First, assume the structure of the rest of the surrounding general equilibrium environment: the firm problem, policy rules, market clearing, and the other exogenous shocks. Second, assume that the identified shock from the empirical specification has a model counterpart in the equilibrium model. Third, assume full-information rational expectations, or more generally some other model of expectation formation such as sticky expectations in Auclert et al. (2020). With these three assumptions, one can simulate the shock in the model to replicate the targeted responses produced by the empirical specification.

To avoid the specification bias that may arise due to the first two assumptions in impulse response matching, single-equation estimation approaches to estimating structural impulse responses, as in Barnichon and Mesters (2020), Lewis and Mertens (2022), can be employed. However, a recursive, aggregate representation of a structural model such as an Euler equation which is used in these previous papers cannot generically be obtained when relaxing the full-information, rational expectations assumption, such as in models of dispersed, private information and a lack of common knowledge (Angeletos and Lian 2018).

To remain agnostic to the correct model of expectation formation and move away from the rational expectations assumption, we must work directly with the sequential representation of the model. Denote the net present value of subjective expected income and rates as

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

Suppose we have imperfectly measured expectations data for income and interest rates for all future horizons, denoted by the expectations operator $E_t^{\text{data}}[\cdot]$. 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 a measurement error term $\mu_t := \mathcal{Y}_t - \mathcal{Y}_t^{\text{data}} + \mathcal{R}_t - \mathcal{R}_t^{\text{data}}$, we obtain

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

Given the potential endogeneity of ε_t, μ_t to $\mathcal{Y}_t^{\text{data}}, \mathcal{R}_t^{\text{data}}$, we need to find a valid instrumental variable z_t that satisfies the standard assumptions[9]

$$\begin{aligned} \text{Exogeneity} \quad & \mathbb{E}[\varepsilon_t z_t] = \mathbb{E}[\mu_t z_t] = 0 \\ \text{Relevance} \quad & \mathbb{E}\left[\mathcal{R}_t^{\text{data}} z_t\right] \neq 0 \end{aligned}$$

The net interest rate elasticity γ contains the structural elasticity of intertemporal substitution, henceforth the EIS, σ, which is the only free parameter to estimate.

What instruments will satisfy these assumptions? As in Barnichon and Mesters (2020) and Lewis and Mertens (2022), I will use contemporaneous values and lags of suitable, externally identified structural shocks as instruments, denoted $\{z_{t-\ell}\}$. For example, to instrument for a demand shock ε_t and satisfy the exogeneity condition, a natural candidate for $z_{t-\ell}$ is a supply shock. Additionally, to satisfy exogeneity with respect to the measurement error μ_t, the response of the expectations data to the shock $z_{t-\ell}$ that I use to construct proxy measures $\mathcal{Y}_t^{\text{data}}, \mathcal{R}_t^{\text{data}}$ must resemble that of the response of true, subjective expectations of households $\mathcal{Y}_t, \mathcal{R}_t$.

A nice interpretation from using structural shocks as instruments is the instrumented version of Equation (2) can be interpreted as an ordinary least-squares "regression on impulse responses". That is, post-multiplying and taking expectations we see that the instrumental variables moment condition can be interpreted as a comparison of impulse responses[10].

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

The left-hand side of the above expression can be interpreted as the unrestricted impulse re-

[9] $\mathcal{Y}_t^{\text{data}}$ does not show up in the relevance condition because the discount factor $\beta = (1+r)^{-1}$ is a known quantity.

[10] The explanatory and response variables in the below expression are unnormalized estimands of a simple local projection onto the shock $z_{t-\ell}$, which is the approach used by Barnichon and Mesters (2020). Lewis and Mertens (2022) demonstrate how to obtain an analogous representation using standard LP/VAR specifications with lags and controls.

sponse of consumption and the right-hand side as the representative-agent, model-implied impulse response of consumption to the shock $\{z_{t-\ell}\}_{\ell \geq 0}$.

A consumption function of histories of realizations and expectations The aggregate consumption function written in Equation (1) had a single dynamic state variable, aggregate financial wealth W_t, which contains incoming period assets A_{t-1}. To write a more general representation of the consumption function that will provide intuition for the heterogeneous-agent case, let us re-consider Equation (2). Model-implied consumption is the sum of responses due to endogenous and exogenous variables, denoted $\tilde{\mathcal{C}}_t, \tilde{\mathcal{E}}_t$ respectively

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

Using the household budget constraint $C_t + A_t = Y_t + (1 + r_{t-1})A_{t-1}$, we can recursively substitute the asset state A_{t-1} out of Equation (3) to obtain an alternate representation

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

Equipped with a set of externally identified structural shocks as instrumental variables as before, we can derive the "regression on impulse responses" moment condition for the general consumption function representation in Equation (4)

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

when $\sigma = \sigma_0$, the true parameter value.

While representations (3) and (4) resulted from a simple algebraic manipulation, they are connected by a useful intuition. Current financial wealth represents the *total* accumulated effect on current consumption of past consumption-savings decisions, which were based on past expectations and realizations of income and interest rates. By integrating out wealth as the dynamic state variable, we can explicitly consider the *individual* contribution of each past expectation and realization in isolation, say $E_\tau^{\text{data}}[Y_{\tau+h}]$ at some time period $\tau < t$, on past consumption and savings C_τ, A_τ, and in turn how that change in outgoing savings A_τ propagated forward to eventually impact current consumption C_t. This chain of events is summarized by the coefficient of $\mathcal{C}_t$ on $E_\tau^{\text{data}}[Y_{\tau+h}]$ in Equation (4).

Dynamic state variables like W_t are often low-dimensional in representative-agent models, which allows for this backward algebraic substitution to derive Equation (4) in explicit form. However, when dynamic state variables are function-valued distributions as in heterogeneous-agent models, this procedure becomes analytically intractable. I will show in the following section, with a few additional assumptions, that one can obtain an analogous aggregate consumption function representation to Equation (4) in heterogeneous-agent models with distributional state variables, like incomplete markets models with idiosyncratic risk and borrowing constraints.

3.2 Heterogeneous-agent consumption functions and moment conditions

I now demonstrate how a class of heterogeneous-agent models that include idiosyncratic risk, incomplete markets, and borrowing constraints (Bewley 1986, İmrohoroğlu 1989, Huggett 1993, Aiyagari 1994) admit analogous moment conditions to Equation (5)[11].

Dynamic programming problem setup Suppose individuals-i have time-t subjective expectations $E_{i,t}$. 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

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

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 one $\mathbf{X}_t$. Individuals-i take aggregates $\mathbf{X}_t$ as given[12], which represent either aggregate exogenous shocks or aggregate endogenous variables determined in equilibrium. I deviate from the usual definition of the state $\mathbf{S}_{i,t}$, which includes the full information set of expectation $E_{i,t}$, to separate variables which directly influence the decision problem in Equation (6) via current constraints and flow payoffs, denoted $\mathbf{S}_{i,t}$, from those that may also influence expectation formation, which I leave implicit in the operator $E_{i,t}$. 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 whose functional form is common and commonly known across individuals-i.

Because I will focus on the effects of measured biases in average survey expectations of aggregate realizations, I reduce the generality of this problem setup by only considering deviations from full-information, rational expectations about aggregate variables. By Assumption 1, subjective expectations of functions of only exogenous idiosyncratic states are therefore taken with respect to their true probability 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) **Dynamic consistency** $E_{i,t}[v_{i,t+h}] = E_{i,t}[v(E_{i,t+h}[v_{i,t+h+1}], \mathbf{S}_{i,t+h}; \boldsymbol{\theta})]$ for $h > 1$

b) **Law of iterated expectations** $E_{i,t}[E_{i,t'}[X_{i,T}]] = E_{i,t}[X_{i,T}]$ for all $t < t' < T$ and stochastic variables $X_{i,t}$

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

Assumptions 2a) and 2b) build on the idea of individual rationality from Assumption 1. Assumption 2a) requires agents to be aware of their dynamic programming problem as written in Equation (6) and to form expectations in a manner that is consistent with its recursive structure. Assumption 2b) is a typical assumption requiring agents' own perceptions of their subjective expectations to be efficient and hence not systematically subject to future revision. Notably, this is not the same thing as assuming law of iterated expectations of the cross-sectional average expectation $\int_{i \in \mathcal{I}} E_{i,t}$, which will not hold generically, for example when relaxing the common knowledge assumption as in Angeletos and Lian (2018).

[11] The dynamic programming setup that leads to the following consumption function derivation is also outlined in Auclert et al. (2020), albeit applied to sticky expectations and other parametric models of expectation formation, and in Bardóczy and Guerreiro (2023), who derive a similar representation but in terms of forecast errors and revisions.

[12] 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.

Assumption 2c) is the main substantive assumption, restricting all forms of correlated heterogeneity in expectations with individual characteristics or states. An example of a model of expectation formation where this assumption holds is sticky information as in Mankiw and Reis (2002), where the prior information update period is independent of all individual characteristics and states even though it is heterogeneous across the population. Hence, the individual state variable encoded in the subjective expectation $E_{i,t}$ would be the prior information update period $\tau \leq t$, which is assumed to be independent of individual states $\mathbf{s}_{i,t}$ such as income. An example of a violation of this assumption is if individual attention correlated with the incidence of aggregate income on individual income, as in Guerreiro (2023).

While this assumption is not innocuous, I would need additional data on the covariances of individual-i subjective expectations and characteristics, which are not be available in the survey expectations data that I use. Nonetheless, given that a large class of models of imperfect information and/or non-rational expectations abides by Assumptions 1 and 2[13] and that my focus is to explain deviations of cross-sectional average expectations about aggregate variables, I believe it is useful to operate under these commonly maintained assumptions.

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

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

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

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

for a scalar, aggregated variable Y_t.

Definition 1. A **steady state** $(\{v_i, \mathbf{s}_i\}_{i\in\mathcal{I}}, \mathbf{D}, \mathbf{X}, Y)$ is a 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 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 expressed 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 an aggregate consumption function for this class of models under the same limited set of assumptions on expectation formation.

Consider the heterogeneous-agent aggregate consumption function implied by Equation

[13] Some examples include diagnostic expectations Bordalo et al. (2018) applied to aggregate variables, sticky information Mankiw and Reis (2002) and sticky expectations Carroll et al. (2020), exogenous noisy and dispersed information Angeletos et al. (2021).

(9) and the coefficients on its typical inputs: income, interest rates, and demand shocks[14].

$$C_t = \sum_{\tau \le t} \sum_{h \ge 0} \mathcal{F}^Y_{t-\tau,h} E_\tau[Y_{\tau+h}] + \mathcal{F}^r_{t-\tau,h} E_\tau[r_{\tau+h}] + \mathcal{F}^\varepsilon_{t-\tau,h} E_\tau[\varepsilon_{\tau+h}] \tag{10}$$

As in the representative-agent case in Equation (4), aggregate consumption here is also a function of both current and past (cross-sectional average) expectations of income, interest rates, and demand shocks. The sum across horizons-h of the terms $\sum_{h\ge0} \mathcal{F}^Y_{t-\tau,h} E_\tau[Y_{\tau,h}]$ is equivalent to the net present value of subjective expected income in the representative-agent model, previously denoted $\mathcal{Y}_\tau$. A single coefficient $\mathcal{F}^Y_{t-\tau,h}$ can be interpreted as the contribution of the distribution of *individuals-i* consumption-savings responses at time-τ, based on a change in their average, aggregate income expectation (or realization if $h = 0$), to time-t aggregate consumption, holding beliefs and realizations in other periods fixed. These time-τ consumption-savings responses propagate forward and affect time-t aggregate consumption because of their effects on the evolution of wealth in the intervening periods.

The reason why coefficients in the matrices $\mathcal{F}^Y, \mathcal{F}^r$ represent the responses of aggregate consumption to a change in arbitrary, average subjective expectations relies principally on certainty equivalence. As long as agents have certainty equivalence, then the response of their decisions to changes in their expectations at a given time-τ are the same, regardless of whether those expectations are rational or not.

If households collectively expect future income will be higher, the effect of that average belief[15] on aggregate consumption is the same irrespective of whether the belief is accurate or distorted. Likewise, when the next period arrives and realized income is lower or higher than expected, the change in aggregate consumption can be decomposed into two components: the response of aggregate consumption to an unanticipated aggregate income shock (without any change in future average expectations), and the response of aggregate consumption to a potentially revised set of average expectations. As before, the effect of the latter component on the aggregate consumption response will be the same regardless of whether the average beliefs are accurate or distorted.

Moment conditions As in the previous section, we can proceed to construct moment conditions from the aggregate consumption function in Equation (10). For simplicity, I proceed by following the notation and distributed lag specification in Barnichon and Mesters (2020), but the same procedure applies for more general VAR/LP specifications as in Lewis and Mertens (2022), which I will use later in the actual estimation[16].

Suppose we have a vector of current and lagged structural shocks $\mathbf{z}_t = \{z_{t-\ell}\}_{\ell=0,\dots,N_\ell}$ to use as instruments.

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

[14] 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.

[15] 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.

[16] The moment conditions in this case would be written with all variables written as differences from their projection onto the set of controls, i.e. $Y^\perp := Y - \mathbb{E}[Y|X]$ for controls X.

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

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

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 $E_t[\cdot]$ and a vector of instruments $\mathbf{z}_t$, apply Equation (9), where $\mathbf{W} \subseteq \mathbf{X}$ denotes the subset of aggregate inputs $\mathbf{X}$ that are observable to the econometrician to 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 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Moment conditions with missing data for distant horizons 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 moment conditions involving the full term structure of expectations, we need to extrapolate the missing horizons in the expectations data.

Let $F_t[\mathbf{W}_{t+h}; \boldsymbol{\vartheta}]$ denote an auxiliary model of the subjective expectation $E_t[\mathbf{W}_{t+h}]$, with parameter vector $\boldsymbol{\vartheta}$ to be used for extrapolation. We must then make an assumption that this auxiliary model produces an unbiased fit to the impulse response of expectations data for farther horizons.

Assumption 5. **(Shape restriction)**

$$\mathbb{E}[(E_t^{\text{data}}[W_{t+h}] - F_t[W_{t+h}; \boldsymbol{\vartheta}_0])z_{t-\ell}] = 0, \text{ for } h \leq H, W_{t+h} \in \mathbf{W}_{t+h},\, z_{t-\ell} \in \mathbf{z}_t \tag{11}$$

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 $z_{t-\ell}$.

Defined expectations data with missing horizons extrapolated from the auxiliary model as

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

Then as before, assuming that the expectations data is an unbiased fit to the true model's average, subjective expectation

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 our desired moment conditions

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

Equation (12) 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 (11). Following that one can evaluate the moment condition to estimate the structural parameters of interest $\boldsymbol{\theta}$. Applying Equation (12) to the case of aggregate consumption, I construct the following set of moment conditions, which encompass both representative- and heterogeneous-agent models, that I estimate in the following section.

$$\mathbb{E}\left[\left(C_t - \sum_{\tau=t-\ell}^{t} \sum_{h \geq 0} \mathcal{F}^Y_{t-\tau,h}(\boldsymbol{\theta}_0) E_\tau[Y_{\tau+h}; \boldsymbol{\vartheta}_0] + \mathcal{F}^r_{t-\tau,h}(\boldsymbol{\theta}_0) E_\tau[r_{\tau+h}; \boldsymbol{\vartheta}_0]\right) z_{t-\ell}\right] = 0, \quad \text{for } z_{t-\ell} \in \mathbf{z}_t \tag{13}$$

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. 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 (11), (12). Relying on the typical assumption that impulse response functions are smooth (Barnichon and Brownlees 2019) provides some additional justification for focusing on low-dimensional, smooth auxiliary models for extrapolation.

4 Semi-structural estimation of model-implied impulse responses

The aim of this section is to estimate and evaluate the fit of a standard set of consumption-savings models against consumption data using moment conditions described in Equation (13), substituting assumptions on a model of expectation formation with expectations data.

4.1 Consumption functions from structural consumption-savings models

The structural models of consumption-savings I consider are the standard representative-agent model and two heterogeneous-agent models.

The first heterogeneous-agent model I consider is the perpetual youth, overlapping generations model because of its analytical tractability. I use the variation of the model developed in Angeletos et al. (2023), which builds on the original Yaari (1965), Blanchard (1985). The consumption function from the model solution linearized around steady state $\beta(1+r) = 1$ is

$$C_t = (1 - \beta\omega) \left(\sum_{h=0}^{\infty} (\beta\omega)^h E_t[Y_{t+h}] + W_t \right) - \gamma \sum_{h=0}^{\infty} (\beta\omega)^h E_t[r_{t+h}] \tag{14}$$

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

This model is attractive because it closely mirrors its representative-agent counterpart, albeit with an 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 as measured by the current marginal propensity to consume (MPC) out of unearned income. As in the representative-agent model, the discount factor β is pinned down by the steady state real interest rate r and is therefore not a degree of freedom for estimation.

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

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

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

As mentioned in the previous section, all models can be represented by a linearized aggregate consumption function of the functional form

$$C_t = \sum_{\tau \leq t} \sum_{h \geq 0} \mathcal{F}^Y_{t-\tau,h} E_\tau[Y_{\tau+h}] + \mathcal{F}^r_{t-\tau,h} E_\tau[r_{\tau+h}] \tag{15}$$

where the coefficients $\mathcal{F}^Y, \mathcal{F}^r$ vary across each model.

All linearized consumption functions are local approximations about a steady state, which I calibrate following McKay et al. (2016). The discount factor β in the standard incomplete markets model must be calibrated to hit the asset-to-disposable income calibration target of 1.4. This leaves at most two parameters available for estimation, which are listed in the lower panel of Table 1: the elasticity of intertemporal substitution and the hazard rate in the perpetual youth model.

Calibrated parameters	Description	Value
r	Real interest rate	0.005
β	Discount factor	See note below
ρ_e	Idiosyncratic productivity persistence	0.966
σ_e^2	Idiosyncratic productivity variance	0.504
Estimated parameters		
σ	Elasticity of intertemporal substitution (EIS)	
ω	Perpetual youth hazard rate	

Table 1: Model parameters

Note: The real interest rate and disposable income values are calibrated to a quarterly frequency. The discount factor β is calibrated to $(1+r)^{-1}$ in the representative-agent and perpetual youth, overlapping generations model and to hit the asset-to-disposable income ratio of 1.4 in the standard incomplete markets model. The hazard rate ω is only available for estimation in the perpetual youth overlapping generations model.

4.2 Data

To evaluate the consumption functions of all models requires 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 weakly positive wealth, a savings rate is the best analog to the model interest rate. Due to this, the federal funds rate is a reasonable proxy for the model-relevant savings rate, given savings rates 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 (expectations) of interest rates, I 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 one short-horizon forecast, typically one-year ahead, and one longer five-to-ten year horizon forecast. Because of the need to extrapolate horizons it is also 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 amount to a violation of 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.

4.3 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 $\tilde{\mathbf{s}}_1 = \mathbf{s}_1 / s_{1,1}$ 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.

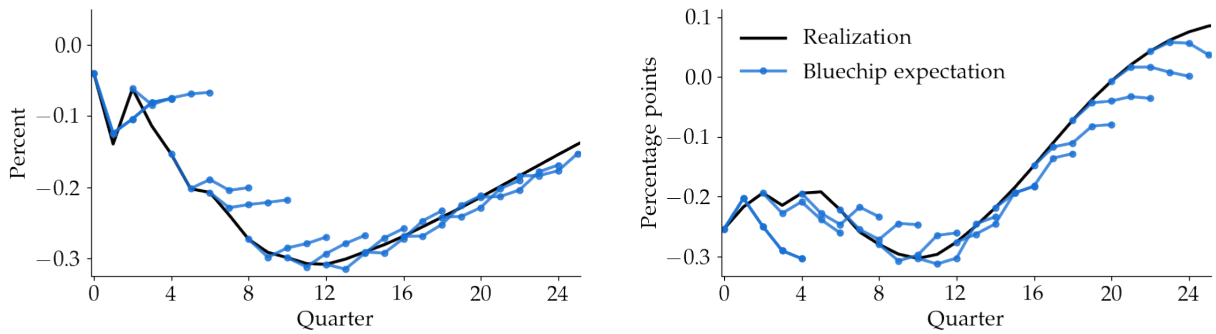
4.4 Impulse responses of income and interest rates to an oil shock

Figure 5 displays impulse responses of realizations and Bluechip expectations of real income and real interest rates, the two main input variables into consumption-savings decisions in standard models. In response to an inflationary oil shock, realized real income and real interest rates (black lines) exhibit a prolonged decline. Counter to results using a time sample with an earlier start date (Bernanke et al. 1997, Gagliardone and Gertler 2023), real interest rates decline as a result of the oil supply news shock due to a largely accommodative nominal federal funds rate response.

I consider a battery of simple parametric models fit to the observed term structure of expectations to extrapolate missing horizons. 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.

4.5 Empirical vs. model-implied impulse responses of consumption

The expectation impulse responses plotted in Figure 5 correspond to the impulse response estimands reported in the model-implied consumption moment condition (12). Figure 6 displays the model-implied impulse responses of consumption from the estimated representative-agent, perpetual youth overlapping generations, and standard incomplete markets models.



(a) Real Disposable Income

(b) Real Federal Funds Rate

Figure 5: Income and interest rate 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, which raises real oil prices by ten percent on impact, and the impulse responses of the one-quarter through one-year ahead expectations (blue dots), connected across horizons. The real federal funds rate is the nominal federal funds rate deflated by consumer price inflation.

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 5b 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 to amplify the negative income effects from lower rates.

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 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.

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 and more positive 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 6 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

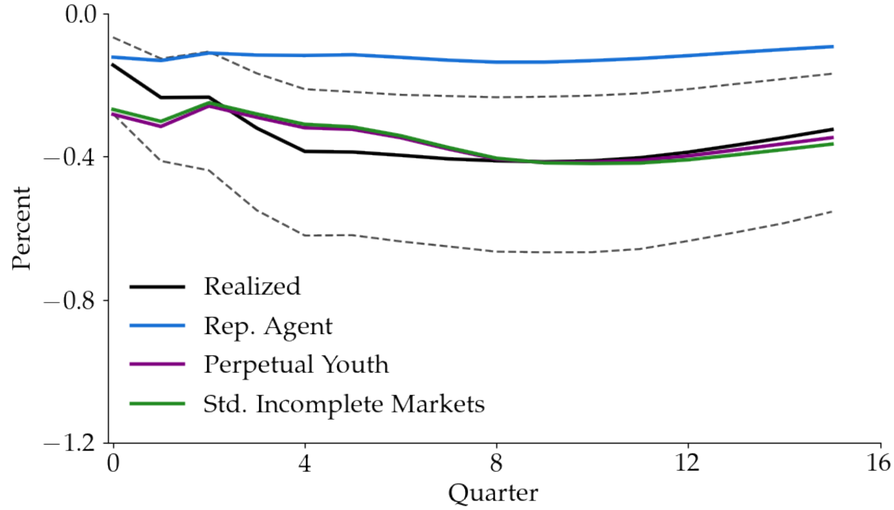


Figure 6: Estimated model-implied consumption impulse responses to a Känzig (2021) oil shock

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).

horizons for income smoothing, the representative-agent models' response to the shock is less severe.

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 7, 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

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 2: 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.

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 (14). Given the large, prolonged decline in realized and expected real interest rates in response to the shock shown in Figure 5, 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 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 6 and 7 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. If the true data generating process for observed real income and interest rates is an economy where agents' expectations are biased, as demonstrated by the Bluechip expectations, then one cannot answer this counterfactual by simply setting the impulse response of expectations equal to the realized response of income and interest rates, i.e. the "rational expectation".

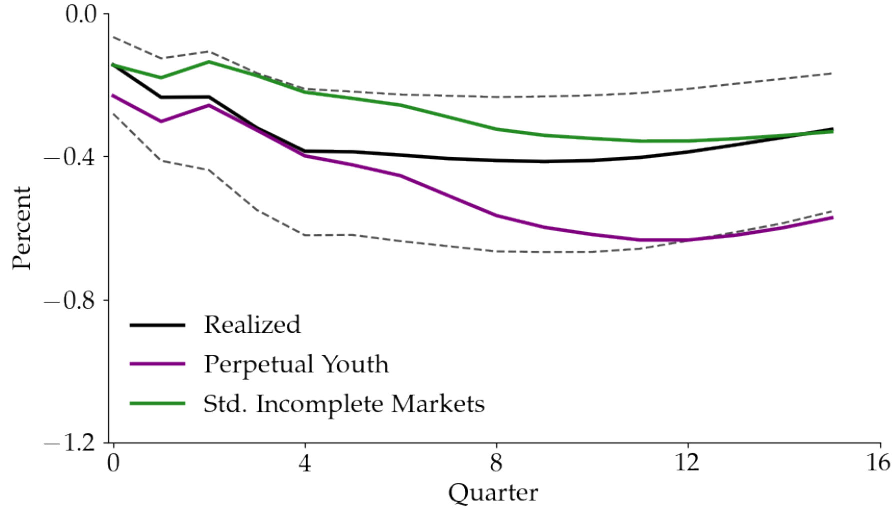


Figure 7: Calibrated model-implied consumption impulse responses to a Känzig (2021) oil shock

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).

5 An equilibrium model with extrapolation in expectations and inertia in realizations

To better understand the determinants of the documented biases in expectation formation and how they result in inertia in aggregate consumption, I proceed to analyze a specific model of expectation formation embedded in a general equilibrium environment.

5.1 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. One can consider the earlier mentioned consumption functions from Section 4.1 as the aggregate demand block of the temporary equilibrium environment, given we will not consider investment for the sake of simplicity. As in Woodford (2013), I immediately resort to using the linearized equilibrium[17], 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 rest of the equilibrium environment closely follows Angeletos et al. (2023), although I simplify along a few dimensions that are not central to my analysis. I only 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.

[17]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.

Households and firms The household sector is exactly the same as the Angeletos et al. (2023) specification of the perpetual youth overlapping generations model. 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.

$$\pi_t = \kappa Y_t + \beta E_t[\pi_{t+1}] \tag{16}$$

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

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

$$i_t - E_t[\pi_{t+1}] \equiv r_t = \phi Y_t \tag{17}$$

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 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 also analyze the dynamics of subjective inflation expectations.

Market clearing in the goods market is given by

$$C_t = Y_t \tag{18}$$

Definition 2. A (linearized) **temporary equilibrium** consists of sequences of prices $\{i_t, \pi_t\}$ and quantities C_t, Y_t that satisfy (14), (16), (17), (18) 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, requiring households to hold zero net wealth in equilibrium. Given output, or equivalently aggregate income, is equal to consumption in equilibrium, moving forward I will simply refer to output as the real quantity of interest.

5.2 General equilibrium with learning about unobserved components

Closing the temporary equilibrium defined in Section 5.1, I assume households form expectations with a particular class of learning models. These models have been shown to be consistent with multiple dimensions of empirical evidence on expectation formation, including in the cross-section of households and forecasters (Nagel 2024, Chen and Liu 2025), experiments (Afrouzi et al. 2023), and unconditional time-series (Crump et al. 2023, Farmer et al. 2024).

In these models, agents observe aggregate variables such as output, perceiving their dynamics to be driven by the sum of a persistent and a transitory component[18], also referred to as

[18]Some of the cited literature adopt the convention that the persistent component is a non-stochastic, long-run mean

"trend" and "cycle" components. Because they cannot immediately distinguish these components, they gradually update their beliefs about each component by solving a filtering problem given the history of observations.

For example, consider an unobserved components model, where households form beliefs about realized output Y_t via their perceived law of motion for output $\tilde{Y}_t$, which is given by

$$\tilde{Y}_t = \tilde{\lambda}_t + \tilde{\eta}_t \tag{19}$$

where $\tilde{\lambda}_t, \tilde{\eta}_t$ are two independent AR(1) processes, parameterized by persistence parameters $\tilde{\rho}_\lambda, \tilde{\rho}_\eta$ and standard deviations $\tilde{\sigma}_\lambda, \tilde{\sigma}_\eta$ with Gaussian innovations. Given observed output Y_t, households update their mean estimates of each component by the standard Kalman update equation

$$\begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \mathbf{F} \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix} + \mathbf{g}(Y_t - E_{t-1}[Y_t]) \tag{20}$$

where $\mathbf{F}$ collects $\tilde{\rho}_\lambda, \tilde{\rho}_\eta$ into a diagonal matrix, $\mathbf{g}$ is the stationary Kalman gain, and $E_{t-1}[Y_t] = E_{t-1}[\lambda_t] + E_{t-1}[\eta_t]$ is the expectation induced by the perceived law of motion in Equation (19).

However, instead of treating Y_t as exogenous and given by a simple functional form like an AR(1), as is commonly done in the forecasting literature on unobserved components models, I close the previously defined temporary equilibrium with this unobserved components model of expectation formation. Therefore, the dynamics of Y_t are endogenous to households' component beliefs $E_t[\lambda_{t+1}], E_t[\eta_{t+1}]$ because of the equilibrium feedback of these beliefs into output, inflation, and interest rates via households' decisions.

To simplify the learning representation and clarify the key model mechanism that results in both extrapolation in expectations and inertia in realizations I make two further assumptions.

I assume that the policy rule in Equation (17) is common knowledge and therefore that the perceived laws of motion for inflation and rates, $\tilde{\pi}_t, \tilde{i}_t$, are consistent with the rule. This allows us to isolate the determination of equilibrium output from inflation and nominal interest rates. Additionally, I assume that the only exogenous shock that directly alters consumption and consequently output Y_t is a demand shock ε_t, which is itself the sum of two AR(1) components

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

which is itself the sum of two AR(1) components with persistence parameters $\rho_\lambda > \rho_\eta$ and mean-zero Gaussian i.i.d innovations with variances $\sigma^2_\lambda, \sigma^2_\eta$.

Consequently, because the perceived and realized real rate are the only ways that inflation and the nominal rate affect output and because these real rates are determined only by perceived and realized output, inflation and the nominal rate provide no additional information about the components λ_t, η_t. Hence, we can treat realized output Y_t as the only observable agents learn from to infer the unobserved components, $\tilde{\lambda}_t, \tilde{\eta}_t$ in the example above, via the perceived law of motion $\tilde{Y}_t$.

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

parameter, which agents are nonetheless uncertain about. The inference problem is similar to the case I study here and results in similar forms of extrapolation that lie at the core of my analysis.

generic equilibrium definition, which nests certain special cases that I will study further.

Definition 3. A **learning equilibrium** is a temporary equilibrium and a collection of subjective expectations $\{E_{t-1}[Y_{t+h}]\}_{h>0}$, which are induced by a perceived law of motion $\tilde{Y}_t$, itself a function of a vector of latent states, and the history $\{Y_{t-\ell}\}_{\ell>1}$.

This equilibrium definition is generic in that I have not imposed any particular consistency criterion between the perceived law of motion $\tilde{Y}_t$ that generates subjective expectations and the resulting equilibrium process for realized Y_t. In the following sections I will consider different perceived laws of motion, which range in the consistency criterion, and study their implications for whether inertia in Y_t results as an equilibrium outcome.

Consolidating equilibrium conditions into a single equation determining real output

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

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, $\beta\omega(1+\sigma\phi)$, affects the overall level of Y_t but 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 denoting Equation (21) with equality for convenience. When considering policy counterfactuals in Section 6 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 the subjective expectation of permanent income $\mathcal{Y}_t := \sum_{h=1}^{\infty}(\beta\omega)^h E_{t-1}[Y_{t+h}]$. Re-writing the aggregate demand equation we obtain the following compact expression for realized output.

$$Y_t = \chi\mathcal{Y}_t + \varepsilon_t \tag{21}$$

In the simple unobserved components model described above, the full set of equilibrium conditions determining realized Y_t can be summarized by Equations (19), (20), (21). We can see that if ε_t does not exhibit inertia in the form of a hump-shaped impulse response, 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 shape. 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 consider different perceived laws of motion to demonstrate how Y_t inertia arises endogenously due to the evolution of beliefs.

5.3 Simple learning

As a tractable starting point, suppose households' perceived law of motion of output is given by the simple form

$$\tilde{Y}_t = \lambda_t + \eta_t \tag{22}$$

where the parameters of persistent component λ_t and transitory component η_t, the additive components of the exogenous demand shock ε_t, are known. This perceived law of motion implies that households fully understand the partial equilibrium or direct effect of the shocks on output but disregard the general equilibrium[19] or indirect effect of the shocks on output through the shocks' feedback into future beliefs, future spending, and back into current spending[20]. While households perceive the output process to be exogenous, realized output is in fact endogenous to households' beliefs through this general equilibrium feedback.

Given the perceived law of motion in Equation (22), we can describe the evolution of component beliefs given by Equation (20) by

$$\begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \mathbf{F} \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix} + \mathbf{g} \left(\overbrace{\underbrace{\chi \mathcal{Y}_t}_{\text{Misspecified } \tilde{Y}_t \text{ wedge}} + \underbrace{\mathbf{1}' \left(\begin{bmatrix} \lambda_t \\ \eta_t \end{bmatrix} - \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix} \right)}_{\text{Standard (rational) Kalman update}}}^{\text{Forecast error } Y_t - E_{t-1}[Y_t]} \right) \tag{23}$$

where $\mathbf{1}$ is a (2×1) vector of ones.

Permanent income implied by Equation (22) takes the form

$$\mathcal{Y}_t = \frac{\beta\omega\rho_\lambda}{1-\beta\omega\rho_\lambda} E_{t-1}[\lambda_t] + \frac{\beta\omega\rho_\eta}{1-\beta\omega\rho_\eta} E_{t-1}[\eta_t] \equiv \mathbf{h}' \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix}$$

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. Hence, I will denote the coefficients preceding component beliefs $E_{t-1}[\lambda_t], E_{t-1}[\eta_t]$ as the "effective horizons" $\mathbf{h}' := [h_\lambda, h_\eta]$ of the current beliefs.

The key difference in the evolution of component beliefs relative to rational learning, i.e. $Y_t = \tilde{Y}_t$, is the wedge $\chi\mathcal{Y}_t$ that stems from the misspecified perceived law of motion $\tilde{Y}_t$. This wedge can be interpreted as households misattributing the portion of a given realized forecast error $Y_t - E_{t-1}[Y_t]$ that was actually due to the endogenous feedback of component beliefs into output instead to perceived, larger exogenous shock realizations λ_t, η_t. This causes a larger component belief revision, if $\chi\mathcal{Y}_t > 0$, than would be warranted under rational learning.

One can obtain an equivalent but informative interpretation by collecting terms $\tilde{\mathbf{F}} := \mathbf{F} + \chi\mathbf{g}\mathbf{h}'$, noting that all entries of $\tilde{\mathbf{F}} > \mathbf{F}$ as long as the belief multiplier χ is positive. The rearranged expression in Equation (23) is

$$\begin{bmatrix} E_t[\lambda_{t+1}] \\ E_t[\eta_{t+1}] \end{bmatrix} = \tilde{\mathbf{F}} \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix} + \mathbf{g}\,\mathbf{1}' \left(\begin{bmatrix} \lambda_t \\ \eta_t \end{bmatrix} - \begin{bmatrix} E_{t-1}[\lambda_t] \\ E_{t-1}[\eta_t] \end{bmatrix} \right)$$

where the effective persistence of beliefs $\tilde{\mathbf{F}}$ is higher (and correlated across components, since $\tilde{\mathbf{F}}$ is no longer diagonal) because of the endogenous feedback from the misspecification wedge $\chi\mathcal{Y}_t$. Figure 8 displays the evolution of component beliefs given a shock to the transitory component η_t. Note that in either interpretation the only reason why there are larger belief revisions or effectively more belief persistence is because of the general equilibrium feedback of beliefs

[19] Bastianello and Fontanier (2025) study this idea of "partial equilibrium thinking" in an application to learning from endogenously determined asset prices.

[20] This feedback loop is also known as the intertemporal Keynesian cross, as in Auclert et al. 2024, whereas here I am also incorporating the counterveiling effects of interest rates into

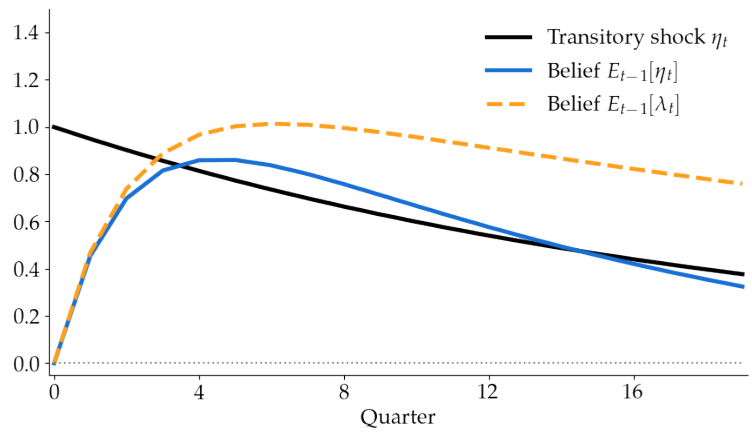


Figure 8: Component belief impulse responses to a transitory demand shock η_t

into realizations governed by χ.

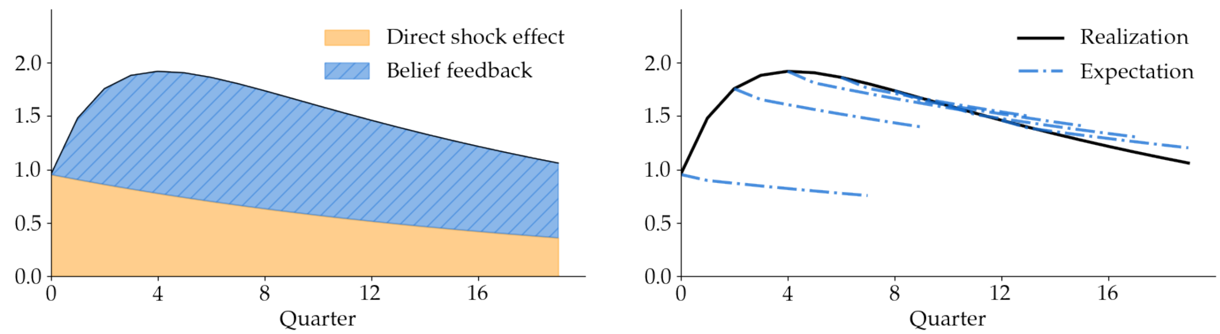
The size of general equilibrium belief feedback $\chi\mathcal{Y}_t$ influences not only the evolution of component beliefs by increasing their effective persistence but also the extent to which changes in component beliefs alter realized outcomes Y_t. In simple learning, the exact same term $\chi\mathcal{Y}_t$ amplifies both effective persistence as described above and realized output in Equation (21).

The left panel of Figure 9 visualizes the intuition stated earlier that realized output Y_t displays inertia only if belief feedback is sufficiently large. Notice also that in the effective persistence of past beliefs, $\tilde{\mathbf{F}}$, the upper row is larger than the bottom row, because the persistent component belief $E_{t-1}[\lambda_t]$ has a larger effective horizon $h_\lambda > h_\eta$. As belief feedback increases the size of belief revisions due to the misattribution of equilibrium effects to further shocks, it also shifts the composition of beliefs toward the persistent component belief, $E_{t-1}[\lambda_t]$, over time. Extrapolation bias arises as a consequence of this compositional shift. Hence, delayed overreaction arises endogenously when the response of output is inertial and as persistent component beliefs become further reinforced as shown in the right panel of Figure 9. I will refer to the two-way feedback loop of beliefs into equilibrium output as "**unanchoring**", given the propensity for an initial shock to trigger an exaggerated and protracted response of beliefs and outcomes relative to the rational learning benchmark, which I will analyze further in the next section.

Inertia and the belief multiplier χ Figure 9 suggests that inertia in realized output and extrapolation bias in expectations arises when the belief multiplier χ is sufficiently large. Here, I formalize this argument and expound further on the intuition.

Definition 4. Consider the moving average representation of $Y_t = \sum_{\ell=0}^{\infty}(a_\ell\lambda_{t-\ell} + b_\ell\eta_{t-\ell})$. Y_t exhibits **inertia** with respect to a component shock $\lambda_{t-\ell}$ if its corresponding coefficients $\{a_\ell\}$ are weakly increasing (decreasing) for $\ell \leq \bar{\ell} > 0$ and weakly decreasing (increasing) for $\ell > \bar{\ell}$[21]. 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 $\{b_\ell\}$.

[21] Given the system (**??**), (**??**), 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.



(a) Output Decomposition

(b) Output Expectation Hair Plot

Figure 9: Output impulse response to a transitory demand shock η_t

Note: Both panels plot the realized output Y_t impulse response to a transitory η_t demand shock in black. The left panel decomposes the output response into the direct effect of the shock (gold solid) and the belief feedback effect (blue hatched) in Equation (21). The right panel expectation "hairs" (blue dot-dash) collect impulse response coefficients of output expectations across horizons.

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 Y_t inertia and the belief multiplier χ. For Y_t to exhibit inertia at all, the belief multiplier χ must be sufficiently large, such that the endogenous amplification and persistence contributed by equilibrium belief feedback exceeds the exogenous decay of the direct shock effect. In the initial period the criteria for inertia to arise simply requires the response of belief feedback to exceed initial exogenous decay of the direct shock effect. However, 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 transitory belief component $E_{t-1}[\eta_t]$, when this belief starts reverting to zero, and the transitory shock η_t itself.

Figure 10 displays the shape of the baseline output response under a high belief multiplier χ in the solid black line and the response under a counterfactual economy with a lower multiplier in the dashed line. We can see that not only is the overall level of the output response lower, but the time profile of the response is also shifted earlier (to the left), since the amplification from belief feedback only occurs gradually through the dynamic unanchoring of beliefs.

What role does the "heterogeneous-agent" side of the economy play in generating inertia? 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 rewrite the parameter values in terms of more interpretable quantities, namely the current marginal propensity to consume (MPC) and the elasticity of intertemporal substitution

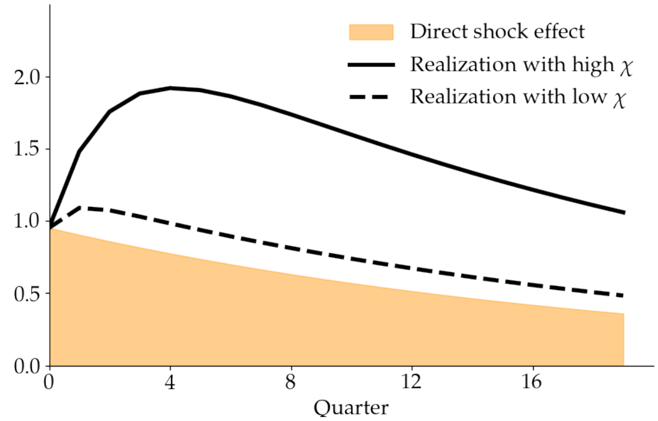


Figure 10: Output impulse responses to an η_t shock under different belief multipliers χ

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.

(EIS). We see then that the belief multiplier χ is large when MPC is high, which is the key difference between heterogeneous and representative-agent models of consumption, and when the EIS is low.

Figure 11 shows how the inertial peaks of output and the component beliefs change as the belief multiplier and the MPC change. The only place the EIS enters the equilibrium output law of motion is through the belief multiplier. Thus, the left panel of Figure 11 can be interpreted equivalently as how inertia changes with the EIS, where a lower EIS implies a higher multiplier and consequently more inertia. The reason is that when the EIS is low, equilibrium output is more insensitive to the counterveiling response of real interest rates via the monetary policy rule. Hence, a boom caused by an initial transitory shock is more likely to cause beliefs to unanchor and result in inertia in realized output. A higher MPC results in a higher belief multiplier but a lower effective horizon for component beliefs, which makes its contribution toward inertia ambiguous in principle. However in practice, as the right panel of Figure 11 demonstrates, the net effect of a higher MPC still results in more inertia.

This same intuition for how Y_t inertia arises also applies to a persistent component shock λ_t. However, the main difference in that case, as documented in Figure 12, is that expectations underreact at all horizons and time periods following a persistent λ_t shock. Even though the composition of beliefs still shift toward the persistent component belief, the portion of beliefs attributed to the transitory component generate underreaction that diminishes over time. This effect resembles the standard mechanisms that arise in models, such as sticky or noisy information, which also exhibit diminishing underreaction over time.

5.4 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 while present in others. The first case I consider is rational learning, where the perceived and actual laws of motion

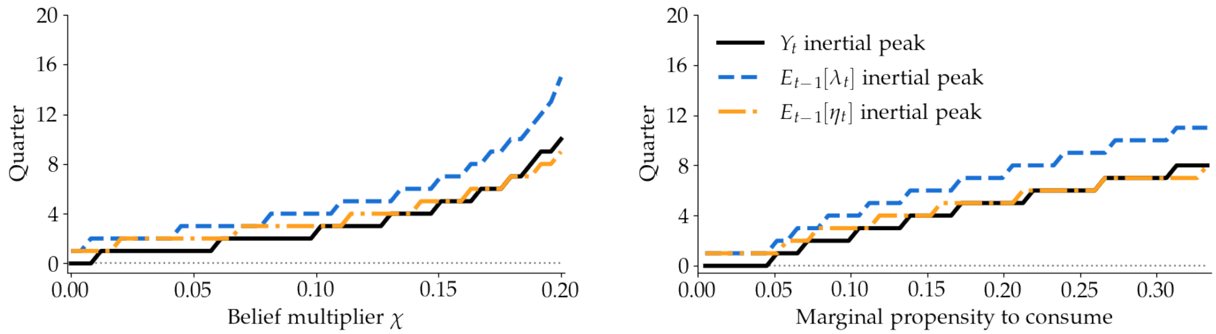


Figure 11: Inertial peaks of output and component belief impulse responses to a transitory η_t shock

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.

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 is given by

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

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 moving average representation of realized 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 $\tilde{a}, \tilde{b}$ as the linear projection coefficients of the constrained-rational perceived law of motion $\tilde{Y}_t$ onto the actual one Y_t. The projection yields the following pair of implicit equations, which define $\tilde{a}, \tilde{b}$ as the solution to their fixed point.

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

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

Let us now consider the differences between these simple, constrained-rational, and rational learning by analyzing their implications on agents' "forward-looking reasoning". I have started by writing down "backward-looking" or "statistical" learning rules that repre-

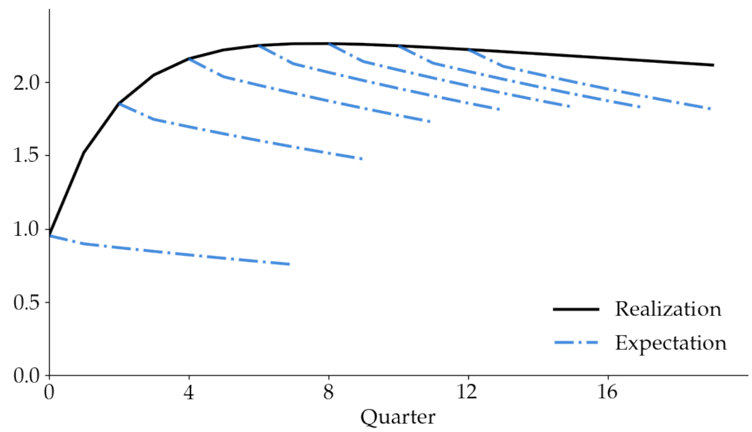


Figure 12: Realized and expected output response to a persistent demand shock λ_t

Note: The black line represents the realized output Y_t impulse response to a persistent λ_t demand shock. The expectation "hairs" (blue dot-dash) collect impulse response coefficients of output expectations across horizons.

sent households' perceived laws of motion, as opposed to starting with a perceived "forward-looking" or "structural" model of the economy, whose reduced form solution is given by these learning rules. A standard notion of "reasoning" within a structural model is an agents' ability to understand the dynamic equilibrium implications of a particular exogenous shock occurring today on current and future outcomes. This typically consists of the shock's forward propagation, e.g. through the exogenous persistence of the shock itself, and then backward propagation, e.g. through the endogenous response of output (consumption) today understanding future output (income) will rise due to the persistence of the shock into the future—that is, the intertemporal Keynesian cross.

These equilibrium feedbacks, both forward and backward, are encoded in the reduced form, or statistical, representation of the perceived structural model as the coefficient on the particular shock, e.g. the moving average coefficient (or loading) a_ℓ of the shock innovation $u_{\lambda,t-\ell}$ on realized output Y_t. Hence, the assumption of a particular "backward-looking" learning rule, such as simple learning where $\tilde{Y}_t = \lambda_t + \eta_t$ with the loading $\tilde{a}_\ell = \rho_\lambda^\ell$, implies that households beliefs behave *as-if* they are not "reasoning" at all about the general equilibrium feedback of a λ_t shock into future output (income) and its effects on current output (consumption) when conducting inference on λ_t, η_t after observing realized output Y_t. In other words, assuming learning rules are restricted, in the sense that they are too low-dimensional to capture the true stochastic process for output Y_t, can be treated analogously as an assumption that agents' general equilibrium reasoning is distorted.

Consider realized output at a future horizon-$h > 0$, $Y_{t+h} = \chi \mathcal{Y}_{t+h} + \varepsilon_{t+h}$. The rational learning equilibrium is the solution to the fixed point of $Y_t = \tilde{Y}_t$, where subjective expectations of future output $E_{t-1}[Y_{t+h}]$ fully account for the amplification from anticipated future belief feedback $\chi \mathcal{Y}_{t+h}$ and in turn from further future belief feedback $\chi \mathcal{Y}_{t+h'}$ for $h' > h$ encoded in $\chi \mathcal{Y}_{t+h}$ and so on. The simple perceived law of motion is on the opposite extreme, where it contemplates that future output is only affected by the direct impact of the shock itself.

Constrained-rational learning constitutes a middle ground between these two polar cases.

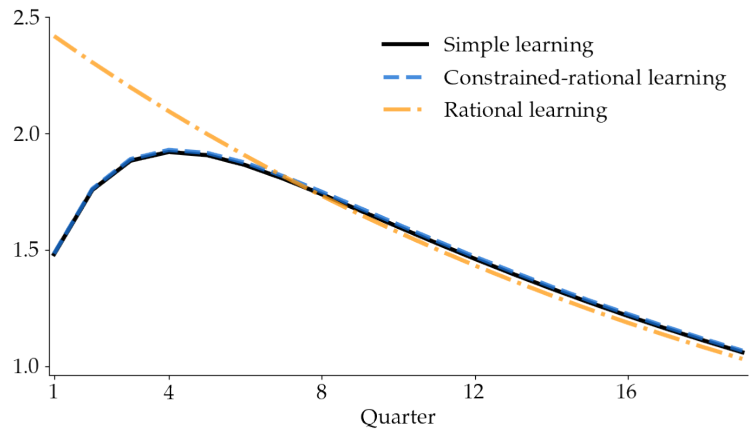


Figure 13: Output impulse responses to an η_t shock under different forms of learning

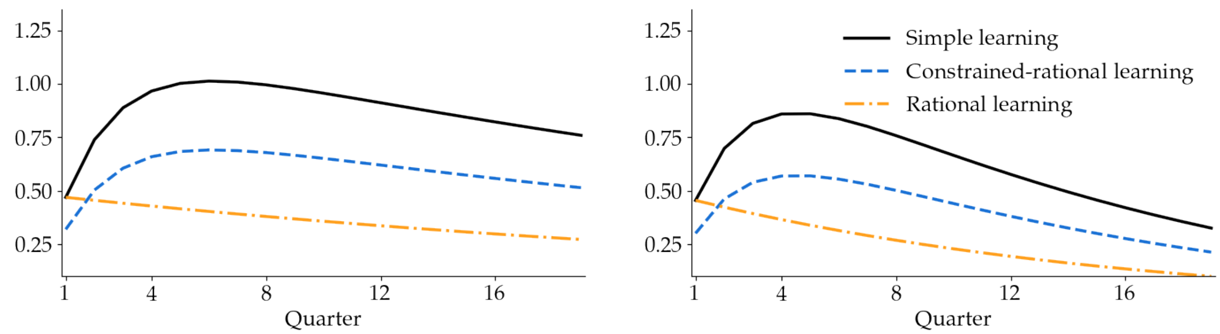
Note: Each line represents the impulse response of output to a transitory demand shock under different learning assumptions. I omit period zero because it does not contain an expectational response due to the staggered timing assumption.

While the persistence of the components in the constrained-rational learning rule are still fixed to be the actual, exogenous persistences of the shocks ρ_λ, ρ_η, this learning rule allows for the level impact of the shock to differ from the direct effect via the coefficients $\tilde{a}, \tilde{b}$. Hence, while constrained-rational learning does not account for the additional endogenous persistence in output due to equilibrium belief feedback, it can partially account for the equilibrium amplification of shocks due to this feedback.

Figure 13 displays the impulse responses of realized output Y_t to a transitory η_t shock under counterfactual economies with the three different learning rules just described. Notably, the impulse response under rational learning does not display inertia. This is not a generic feature of rational learning, which in principle alone could generate inertia, but it will depend ultimately on the persistences and variances of the underlying shock components. However, it is worth emphasizing that inertia is more likely to arise under simple and constrained-rational learning for a given pre-defined set of shock components because of the additional amplification and endogenous persistence in beliefs due to the feedback wedge $\chi \mathcal{Y}_t$ as in Equation (23).

The output response under rational learning is initially much larger than under simple or constrained-rational learning, due to the belief feedback amplification, but it eventually decays below their responses. In other words, the time profile of amplification of rational learning over simple learning is not uniform, where constrained rational learning seems more closely resemble simple learning rather than rational learning.

What accounts for the absence of amplification in the realized output response under constrained-rational learning relative to the simple learning benchmark? Figure 14 displays the impulse responses of component beliefs under the three different forms of learning and provides some insight to address this question. First, recall that under all three forms of learning the direct shock effect is the same. Therefore, if the realized output responses are similar this must mean that the size and shape of the belief feedback effects are similar.



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

Figure 14: 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.

Figure 14 shows that the responses of component beliefs under constrained-rational learning are smaller than under simple learning. This means that the degree of equilibrium amplification in realized output Y_t for a given level of belief $E_{t-1}[\lambda_t], E_{t-1}[\eta_t]$ must then be larger under constrained-rational learning for the overall belief feedback effect to be equal the feedback effect under simple learning. In turn, if we treat the response of beliefs under rational learning as the rational benchmark, then we see that the degree of unanchoring of beliefs under constrained-rational learning is substantially smaller than under simple learning because of the ability of beliefs to partially account for equilibrium feedback. Nonetheless, there is still some unanchoring of beliefs, and it is the resultant amplification from this unanchoring that eventually causes the output response under constrained-rational learning to exceed that of rational learning. I consider this idea more formally in Proposition 2 in a special case with an i.i.d transitory shock η_t.

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.

Beliefs under constrained-rational learning exhibit partial unanchoring, which makes the intuition from the simple learning example I developed earlier a helpful analogy. Unlike simple learning, constrained-rational beliefs still maintain a degree of consistency to realized output dynamics in being the best approximation in the class of learning rules that are linear in the shock components. Given beliefs are partly responsive to their surrounding economic environment, constrained-rational learning is a useful illustrative model of expectation formation that delivers inertia in realizations and extrapolation bias in expectations, while still satisfying a limited notion of the Lucas critique. Therefore, we proceed to the next section to consider some simple policy counterfactuals and how alternative policies can alter the degree of inertia in realizations via their ability to anchor expectations.

6 Policy implications of macroeconomic inertia

This section discusses novel policy considerations that arise under constrained-rational learning and contrast them with typical policy transmission outcomes in models with full-information rational expectations. Under constrained-rational learning, it is no longer desirable to be infinitely-responsive to demand-driven fluctuations because of the risk of destabilizing expectations. Gradual policy approaches to monetary policy, as in a highly inertial Taylor rule, fail to stabilize output as effectively as they would under the full-information rational expectations benchmark. In addition, the stimulus effects of a deficit-financed fiscal transfer in a heterogeneous-agent New Keynesian economy may no longer be front-loaded.

6.1 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 also holds 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 in an economy with solely demand shocks.

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

I consider first the full-information rational expectations equilibrium response to a transitory demand 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 $E_t[\lambda_{t+1}], E_t[\eta_{t+1}]$ and consequently output Y_t itself can become destabilized. Higher monetary responsiveness is therefore effective only up to a point, a limitation that is also demonstrated in Eusepi et al. (2024).

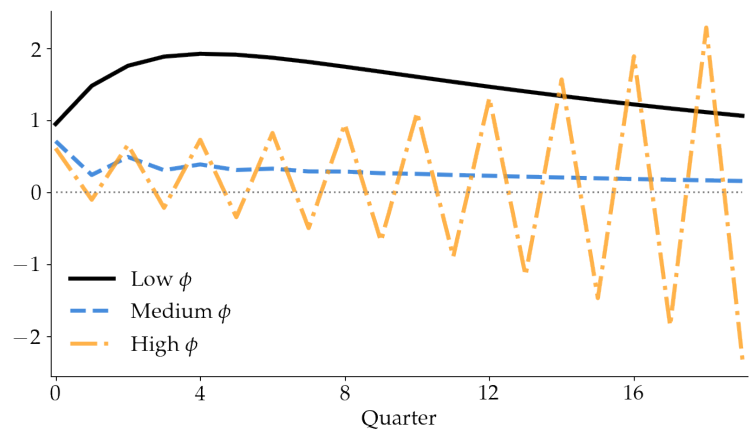


Figure 15: Output (de-)stabilization under different Taylor rule ϕ

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

Figure 15 demonstrates that this behavior also holds in the constrained-rational learning case. The shared reason in both simple and constrained-rational learning is that the expectation feedback wedge that appears in the belief component law of motion is increasingly negative as ϕ increases. When households spend more due to optimistic beliefs about demand shocks, the monetary authority raises interest rates so significantly that it triggers a contraction in output. This causes households to mistakenly infer that the realized shocks were actually negative and larger in magnitude than they had previously anticipated. The result is an increasingly unstable negative feedback loop, resulting in the explosive oscillation of the high ϕ case shown in Figure 15.

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 the expectation wedge still appears in the component belief law of motion, which in principle still contributes to inertia, a monetary authority that chooses a Taylor coefficient ϕ that sets the belief multiplier χ sufficiently close to zero effectively sets the wedge $\chi\mathcal{Y}_t$ equal to zero. This stops the expectation feedback loop that would induce inertia from starting in the first place.

In choosing an optimal level of responsiveness to demand shocks ϕ, a monetary authority facing households with constrained-rational learning should not respond as forcefully as in the rational benchmark because of the risk of destabilizing expectations and therefore may not be able to fully shut down demand-driven fluctuations.

6.2 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 (Woodford 1999).

I expand briefly on this latter reason by demonstrating the inability of constrained-rational and simple learning to map the effects of dynamic policy commitments to perceived output. Just as constrained households are unable to fully internalize the equilibrium feedbacks of future expected output changes on current output, which we discussed in the previous section, so too are they unable to internalize the effects of current policy commitments on future expected output.

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

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

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, one can equate their contribution to dampening the level of equilibrium output Y_t due to their feedback from 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 $\bar{\phi}$.

The crucial step in the above explanation was 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.

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

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. Two regimes with the same effective $\bar{\phi}$ would exhibit different equilibrium responses of output at time-0 if households correctly perceived that the regime with higher policy rule persistence ρ would continue to respond more forcefully in future periods and that the future equilibrium output change would propagate backward to time-0 output via the intertemporal Keynesian cross.

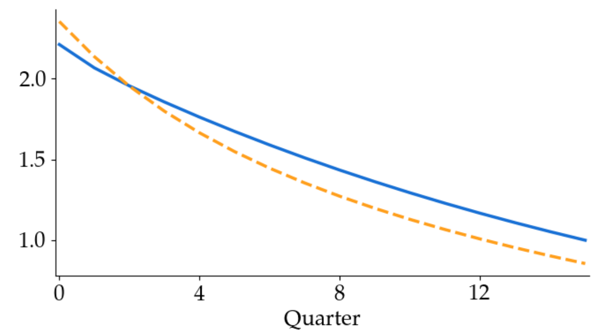
However, when learning rules are restricted to load on contemporaneous shocks as in the constrained-rational and simple learning cases where $\tilde{Y}_{t+h}$ can only be a function of ε_{t+h}, we see that both belief feedback in $\{E_{t+h}[Y_{t+h+j}]\}_{j>0}$ and effects of continued policy commitments on future output $\{\tilde{Y}_{t+h-\ell}\}_{\ell>0}$ are imperfectly incorporated into beliefs. This is again because simple linear functions of contemporaneous shocks ε_{t+h} cannot capture the equilibrium dynamics of the contributions of past shocks $\{\varepsilon_{t+h-\ell}\}_{\ell>0}$ through belief feedback and policy commitments.

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 the full-information rational expectations benchmark, since there is a continuum of regimes (ρ, ϕ) that obtain the same welfare loss.

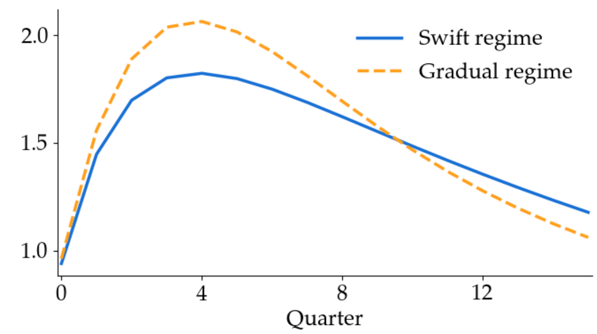
Figure 16 displays the impulse responses of output and interest rates to a transitory demand shock η_t under swift and gradual regimes. The left column are the responses under the full-information, rational expectations (FIRE) benchmark and the right column are the responses under constrained-rational learning. Focusing first on the FIRE case, we see that output responds immediately under both regimes, even though the peak response of interest rates is not immediate under either regime. This is because time-0 beliefs instantly incorporate the equilibrium effects of future implied interest rate changes on future output, which feed back into time-0 output.

Because the gradual regime has a higher degree of policy inertia, we see the path of interest rates rising more slowly than in the swift regime. Nonetheless, the path of output under the gradual regime falls by more than under the swift regime only a couple of quarters after the initial shock even though it takes interest rates nearly twelve quarters to finally exceed the rate path under the swift regime. In contrast, under learning we see that output gap between the gradual and swift regimes is still positive until the interest rate gap between regimes also switches signs.

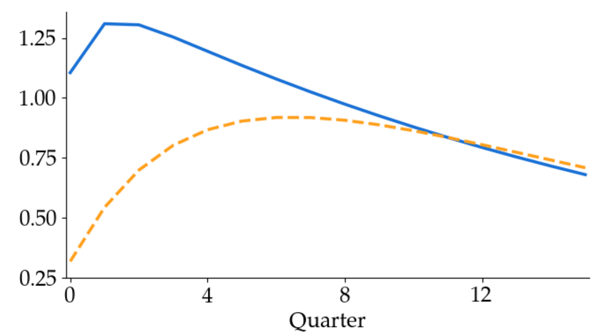
Figure 17 demonstrates the difference in current-period discounted welfare $-\beta^t Y_t^2$ between the two regimes. Because these regimes were chosen to equate total welfare loss under full-information rational expectations, the area under the blue curve in Figure 17 integrates to 1. We see that the 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, because the gradual regime is less effective at containing output under learning due to households limits in 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.



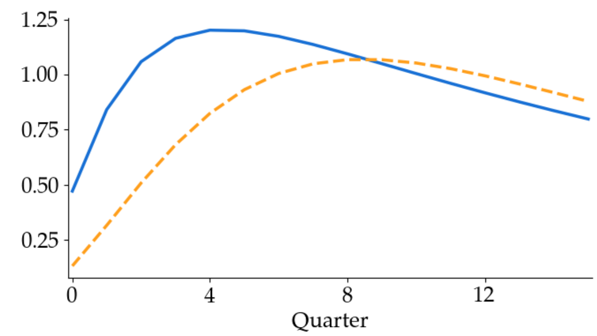
(a) Output response under FIRE



(b) Output response under learning



(c) Interest rate response under FIRE



(d) Interest rate response under learning

Figure 16: 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.

6.3 Deferred financing and the delayed impacts of fiscal stimulus

A key difference in the policy implications of heterogeneous-agent versus representative-agent macroeconomic models are the equilibrium responses of prices and quantities to deficit-financed fiscal transfers (Auclert et al. 2024). Representative-agent models with full-information, rational expectations have Ricardian equivalence, which implies that non-distortionary government spending and tax and transfer policies have identical effects on equilibrium outcomes regardless of the timing of financing. Conversely, deficit-financed fiscal policy can induce immediate and large equilibrium responses of quantities, such as consumption spending, in heterogeneous-agent models, where the magnitudes of these responses increase as financing is further delayed due to the compounding effects of future general equilibrium feedback into current demand (Angeletos et al. 2023).

When households form expectations with imperfect learning, 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 immediacy. Prolonging deficits additionally

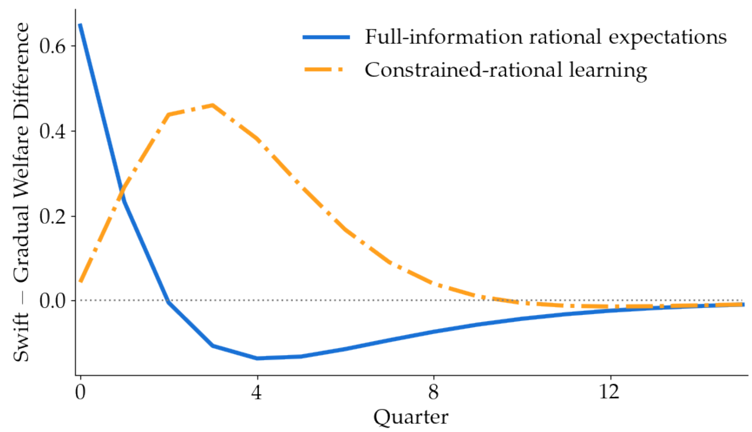


Figure 17: Monetary regime welfare differences under different models of expectation formation

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.

stretches out the cumulative response of output across a much longer horizon, running the risk that stimulus will last 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 linearized government budget constraint and tax rule are given by

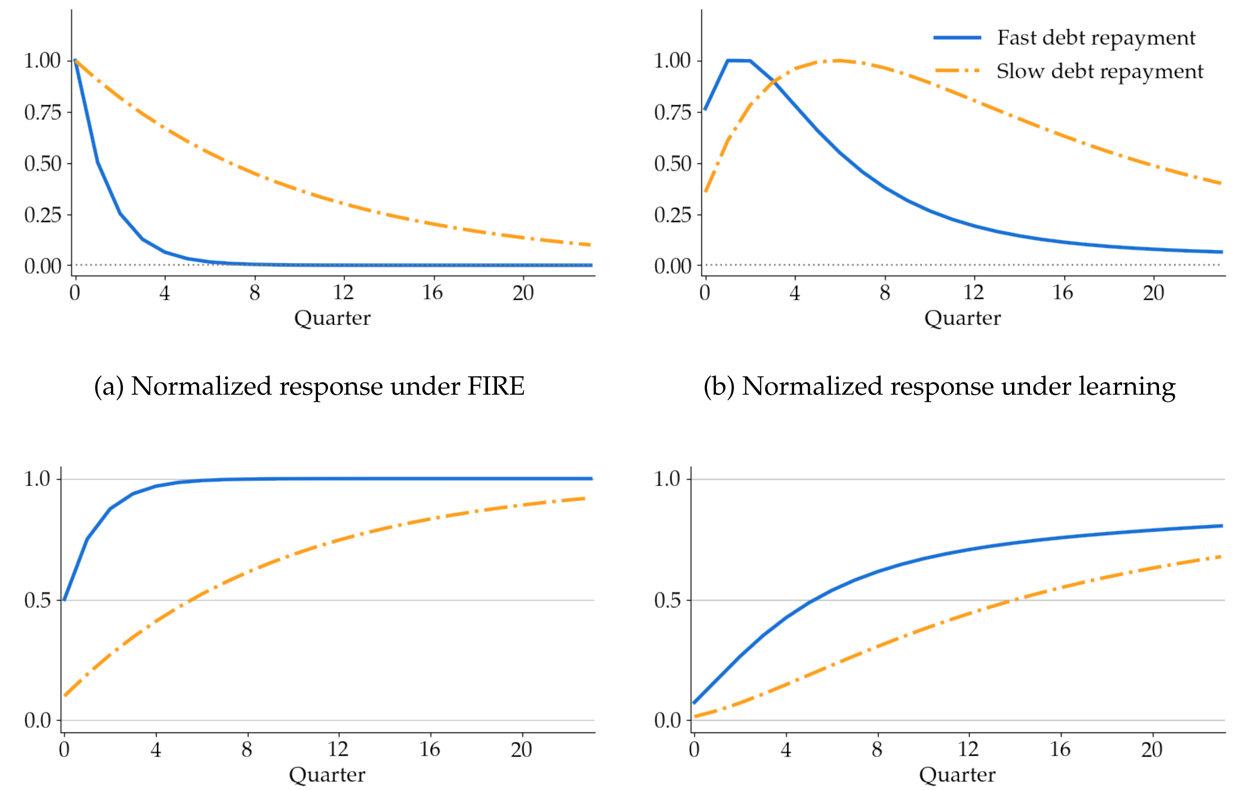
$$B_t = \frac{1}{\beta}(B_{t-1} - T_t)$$
$$T_t = \delta B_{t-1} - (1-\delta)\zeta_t$$

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.

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



(a) Normalized response under FIRE

(b) Normalized response under learning

(c) Discounted cumulative response under FIRE

(d) Discounted cumulative response under learning

Figure 18: 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.

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

Figure 18 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 of one year in the peak output response between regimes and a difference of ten quarters for the discounted cumulative response, more than double the gap under FIRE.

7 Conclusion

This paper argues 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 when supplemented with measured expectations data yield model-implied impulse responses of consumption that closely resemble observed consumption inertia.

Guided by new empirical evidence on the extrapolative tendencies of measured expectations, I embed an unobserved components model of expectation formation into a tractable heterogeneous-agent, general equilibrium environment. Learning is imperfect when the model of expectation formation is too simple to perfectly perceive the dynamic equilibrium effects of shocks. I show that this model can endogenously generate inertia in the dynamics of realized consumption and extrapolation in expectations due to a form of belief unanchoring. The propensity for beliefs to unanchor is tightly linked to the magnitude of general equilibrium amplification, which is large when marginal propensities to consume are large, as in heterogeneous-agent models, and when the policy regime is less responsive. I conclude by providing a new rationale that cautions against gradual approaches to monetary policy in the form of inertial policy rules and to fiscal policy in the form of deferred financing of fiscal deficits, both having the potential to inadvertently contribute to longer 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 Chen Lian, "Forward guidance without common knowledge," *American Economic Review*, 2018, *108* (9), 2477–2512.

_ 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 Christian Brownlees, "Impulse response estimation by smooth local projections," *Review of Economics and Statistics*, 2019, *101* (3), 522–530.

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

Bastianello, Francesca and Paul Fontanier, "Expectations and learning from prices," *Review of Economic Studies*, 2025, *92* (3), 1341–1374.

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.

Chen, Heng and Yicheng Liu, "Expectation and Confusion: Evidence and Theory," *Available at SSRN 5343753*, 2025.

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.

Hansen, Lars Peter and Kenneth J Singleton, "Generalized instrumental variables estimation of nonlinear rational expectations models," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1269–1286.

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.

Kučinskas, Simas and Florian S Peters, “Measuring under-and overreaction in expectation formation,” *Review of Economics and Statistics*, 2022, pp. 1–45.

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.

__ , “Survey measurement of probabilistic macroeconomic expectations: progress and promise,” *NBER Macroeconomics Annual*, 2018, *32* (1), 411–471.

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.

Piazzesi, Monika and Martin Schneider, "Momentum traders in the housing market: Survey evidence and a search model," *American Economic Review*, 2009, *99* (2), 406–411.

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.

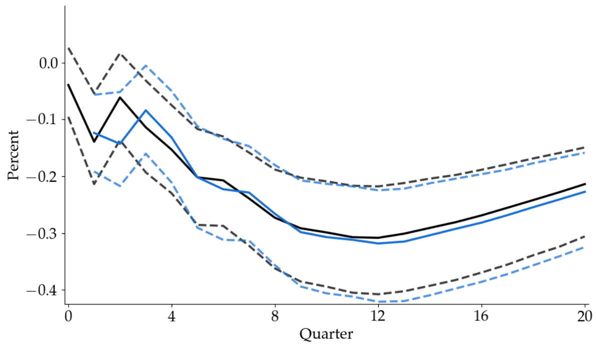
— , "Imperfect common knowledge and the effects of monetary policy," 2001.

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

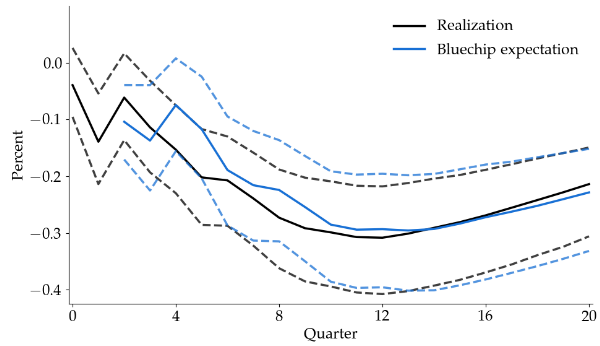
Yaari, Menahem E, "Uncertain lifetime, life insurance, and the theory of the consumer," *The Review of Economic Studies*, 1965, *32* (2), 137–150.

A Additional impulse response results

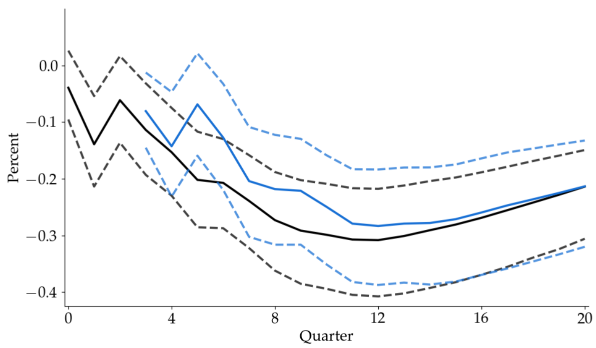
A.1 Impulse response comparisons with confidence bands



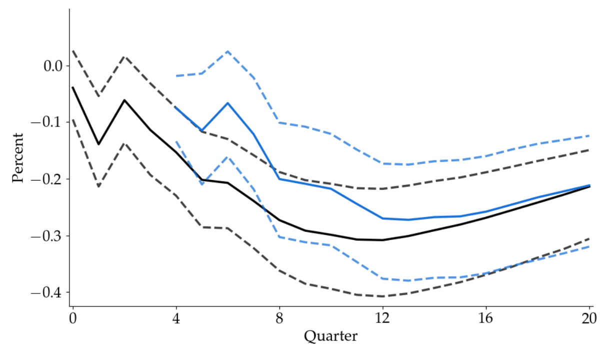
(a) One-quarter ahead Bluechip expectation



(b) Two-quarters ahead Bluechip expectation



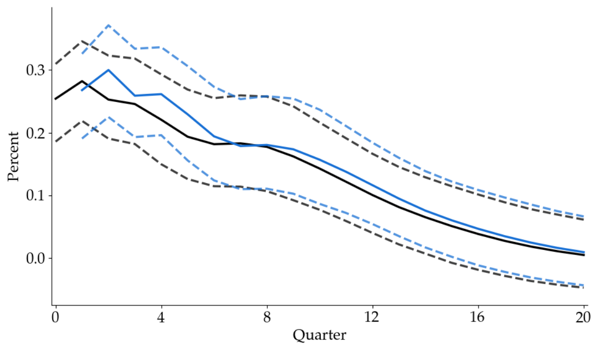
(c) Three-quarters ahead Bluechip expectation



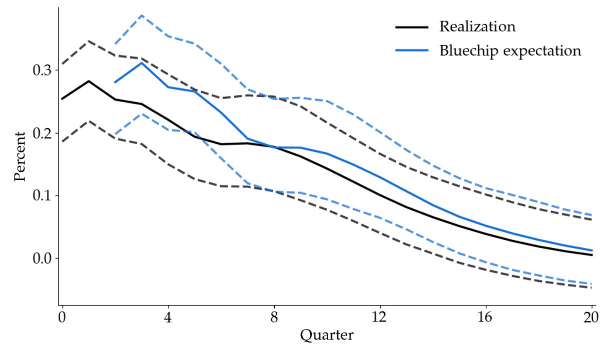
(d) Four-quarters ahead Bluechip expectation

Figure 19: Real disposable income impulse responses to a Känzig (2021) oil shock

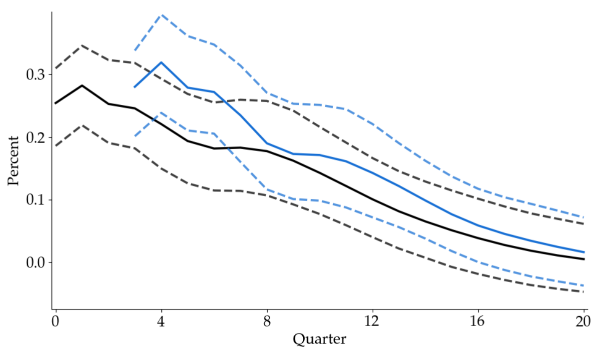
Note: each panel contains an impulse response function of realizations (black) and a fixed horizon-h (blue) forecast from Bluechip survey expectations data to a positive Känzig (2021) oil price news shock. The dashed lines are 68% confidence bands produced using moving block bootstrap by Jentsch and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



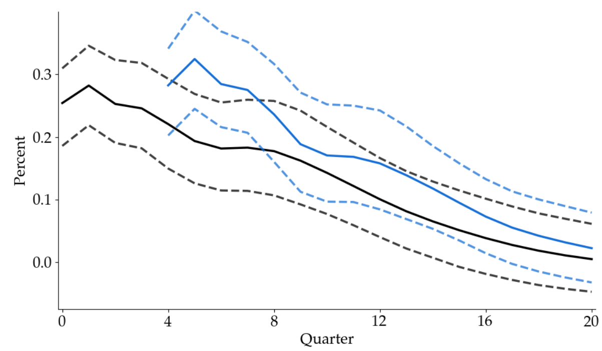
(a) One-quarter ahead Bluechip expectation



(b) Two-quarters ahead Bluechip expectation



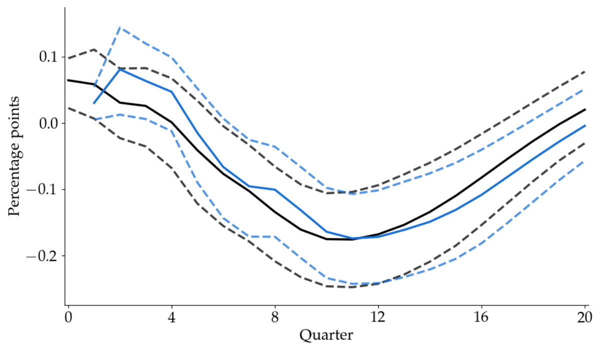
(c) Three-quarters ahead Bluechip expectation



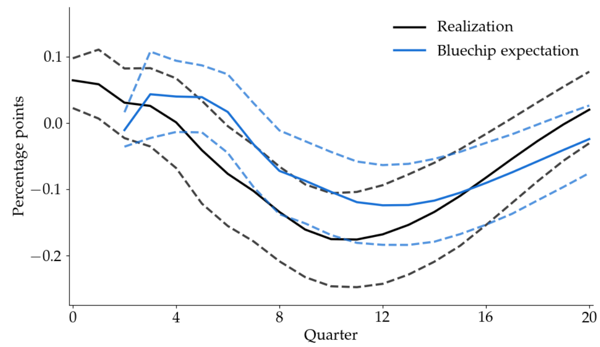
(d) Four-quarters ahead Bluechip expectation

Figure 20: CPI inflation impulse responses to a Känzig (2021) oil shock

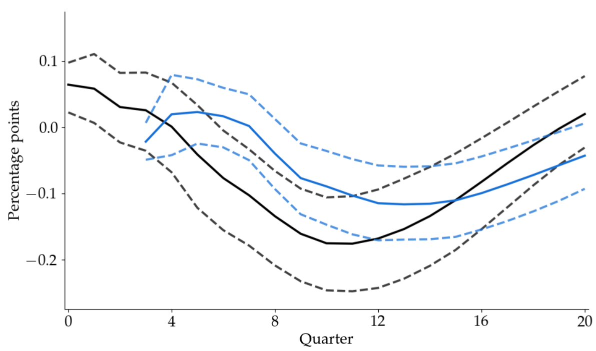
Note: each panel contains an impulse response function of realizations (black) and a fixed horizon-h (blue) forecast from Bluechip survey expectations data to a positive Känzig (2021) oil price news shock. The dashed lines are 68% confidence bands produced using moving block bootstrap by Jentsch and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



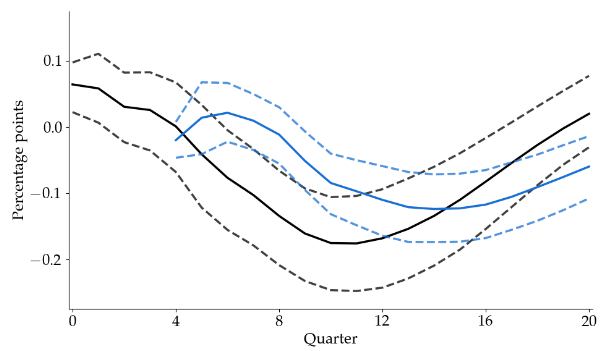
(a) One-quarter ahead Bluechip expectation



(b) Two-quarters ahead Bluechip expectation



(c) Three-quarters ahead Bluechip expectation



(d) Four-quarters ahead Bluechip expectation

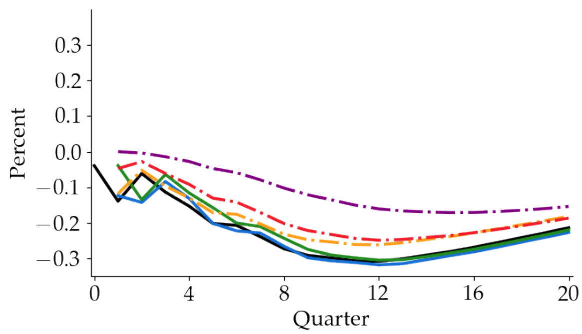
Figure 21: Nominal federal funds rate impulse responses to a Känzig (2021) oil shock

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).

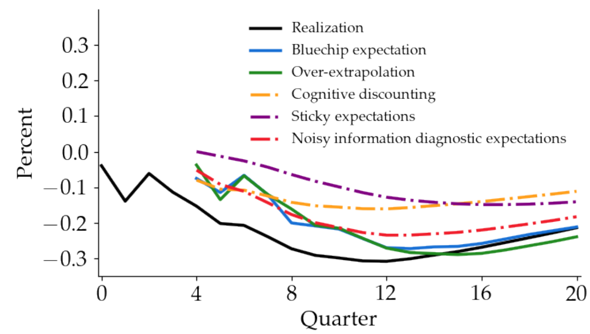
A.2 Other models of expectation formation

Figure 22 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.

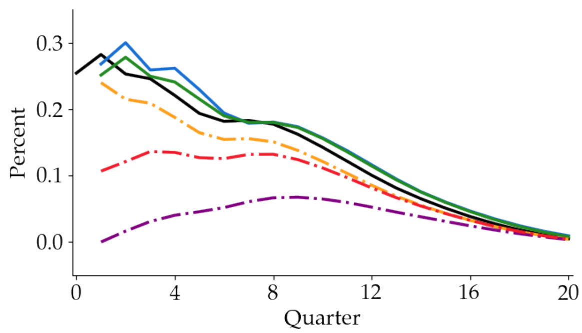
The over-extrapolation model is the model from Equation (**??**) and whose hair plot is displayed in Figure **??**. As shown earlier, the over-extrapolation model is able to rationalize the Bluechip expectations impulse responses across variables, time, and horizons. We can similarly construct impulse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



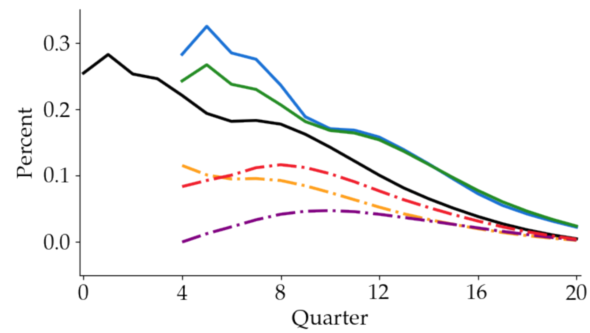
(a) Real disposable income, one-quarter ahead exp.



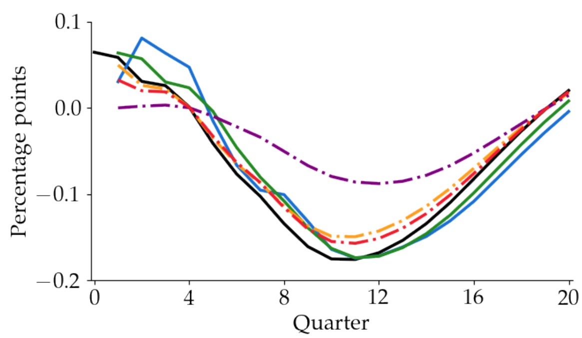
(b) Real disposable income, four-quarters ahead exp.



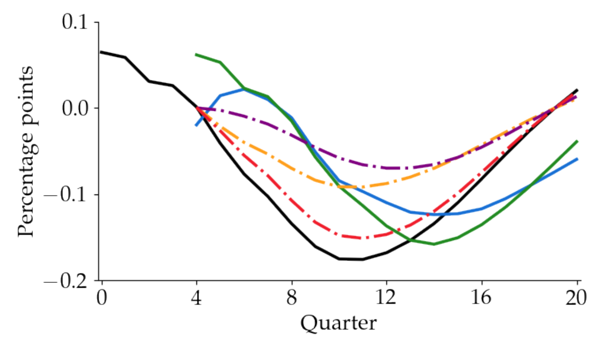
(c) CPI inflation, one-quarter ahead exp.



(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 22: 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 22 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 22 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 22. 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 3: 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 4.4, 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 (11) 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 3. 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 3 is the baseline extrapolation I choose in the main set of results reported in Table 2 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 5, 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 (**??**) and displayed in Figure **??**.

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 **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[22].

Positive, real eigenvalues of A
For the eigenvalues of **A** to be real (note that **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 **A** are real.

For the eigenvalues to be positive, we have

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

[22] This is potentially stronger than necessary for bounding the absolute value of the MA 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

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

a_ℓ denote the MA 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 MA representations 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 MA 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 MA 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 MA 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

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

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.