Micro Jumps, Macro Humps:

Monetary Policy and Business Cycles in an Estimated HANK Model

Adrien Auclert* Matthew Rognlie† Ludwig Straub‡

March 2020

Abstract

We estimate a Heterogeneous-Agent New Keynesian model with sticky household expectations that matches existing microeconomic evidence on marginal propensities to consume *and* macroeconomic evidence on the impulse response to a monetary policy shock. Our estimated model uncovers a central role for investment in the transmission mechanism of monetary policy, as high MPCs amplify the investment response in the data. This force also generates a procyclical response of consumption to investment shocks, leading our model to infer a central role for these shocks as a source of business cycles.

*Stanford University, CEPR and NBER. Email: aauclert@stanford.edu.

†Northwestern University and NBER. Email: matthew.rognlie@northwestern.edu.

‡Harvard University and NBER. Email: ludwigstraub@fas.harvard.edu.

The material in appendix A was presented at the SED 2018 in Mexico City under the title "Forward Guidance is More Powerful Than You Think", which this paper supersedes. For helpful comments, we thank Marios Angeletos, Florin Bilbiie, Luigi Bocola, Larry Christiano, Martin Eichenbaum, Emmanuel Farhi, Xavier Gabaix, Greg Kaplan, Rohan Kekre, Ralph Luetticke, Alisdair McKay, Kurt Mitman, Ben Moll, Ricardo Reis, Gianluca Violante, and Peter Zorn. This research is supported by the National Science Foundation grant SES-1851717.

1 Introduction

A central question in monetary economics is how to model the effects of monetary policy. How do we gauge the success of a model? What empirical moments should we use to judge?

The traditional approach in the literature is to build models that are consistent with *macro moments*, such as the impulse response to identified monetary policy shocks (Rotemberg and Woodford 1997, Christiano, Eichenbaum and Evans 2005) or the covariance structure of aggregate time series (Smets and Wouters 2007, Justiniano, Primiceri and Tambalotti 2010). One important fact in this literature is that the macroeconomic response to aggregate shocks tends to be hump-shaped. This is consistent with the conventional view in central banks that "monetary policy can have little immediate effect on either real activity or inflation" (Woodford 2003, p. 322). The left panel of Figure 1 displays the impulse response of output to an identified monetary policy shock—one of the targets of our estimated model in section 4—which clearly displays such a *macro hump*. To deliver these hump-shapes, the literature to date has used representative-agent models enhanced with a variety of adjustment frictions such as habit formation, or deviations from rational expectations such as inattention.[1] Estimated versions of these models are currently widely used by central banks for forecasting and policy analysis.

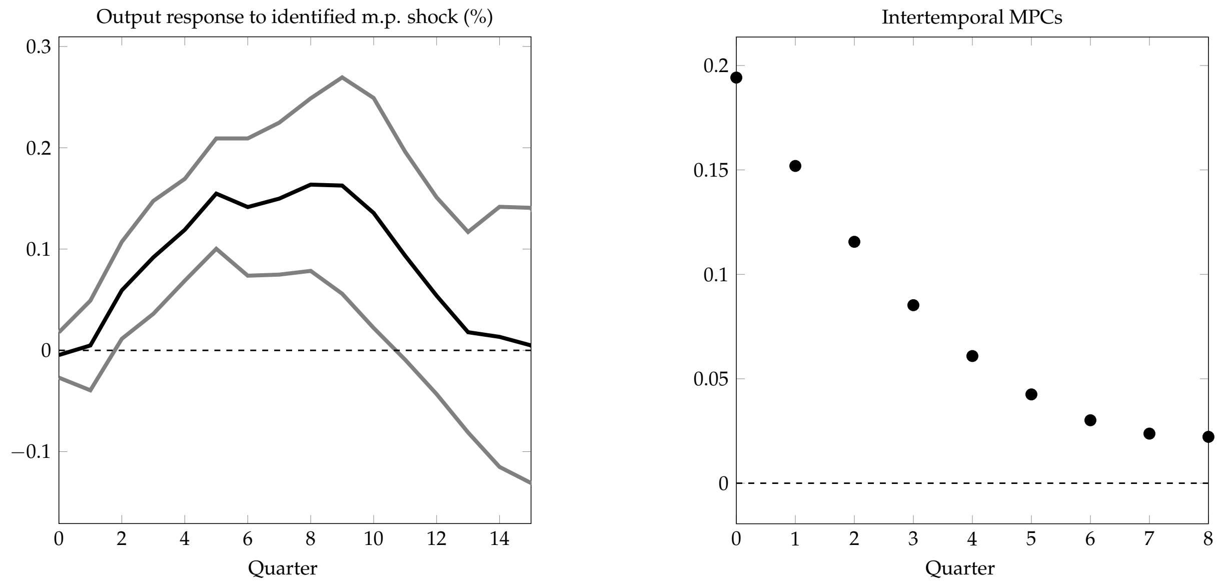
A recent literature has proposed instead to build models that are consistent with *micro moments*, such as the path of the consumption response to an identified transitory income shock. This path is generally characterized by an immediate *jump* on impact—an elevated Marginal Propensity to Consume, or MPC—followed by a less pronounced but still elevated level of spending in the following periods, as in the right panel of figure 1 (see Fagereng, Holm and Natvik 2018, Auclert, Rognlie and Straub 2018). To match these moments, the literature to date has used heterogeneous-agent models with incomplete markets, idiosyncratic risk and borrowing constraints.[2] These models have gathered a considerable amount of attention because they can speak to the interaction between monetary policy and distribution, and are often viewed as painting a realistic and intuitive picture of the monetary transmission mechanism.

At present, these two approaches are incompatible. As is well-known, representative-agent models feature MPCs that are much too low compared to the data, so models in the macro moments tradition fail to match micro jumps. Similarly, existing heterogeneous-agent models feature an aggregate impulse response to monetary policy that is peaked on impact (e.g. Kaplan, Moll and Violante 2018, figure 3), so models in the micro moments tradition fail to match macro humps, and are therefore ill-suited for estimation using macro data.

In this paper, we combine elements from both literatures to build and estimate a model of the monetary transmission mechanism that simultaneously matches these macro and micro moments. We overcome two major difficulties that had made this exercise difficult until now. First, we show that the standard approach to generating consumption hump-shapes—introducing habit formation in consumption—makes it very difficult to match MPCs. By contrast, we show that introduc-

[1] e.g. Christiano, Eichenbaum and Evans (2005), Smets and Wouters (2007), and Maćkowiak and Wiederholt (2015).
[2] e.g. McKay, Nakamura and Steinsson (2016), Kaplan, Moll and Violante (2018), and Bilbiie (2019).

Figure 1: Macro Humps, Micro Jumps.



Note. Left panel shows the impulse response of output to a Romer and Romer (2004) shock, estimated with a Jordà (2005) projection; see section 4.2 for details. Right panel shows the consumption response to a one-time unanticipated increase in average labor incomes; estimated by Fagereng, Holm and Natvik (2018) using Norwegian administrative data; interpolated to quarterly data using cubic interpolation on the cumulative spending response.

ing sluggishness in the adjustment of households' expectations of aggregate variables ("sticky expectations", as in Carroll, Crawley, Slacalek, Tokuoka and White 2018) allows the model to simultaneously match macro humps and micro jumps. Second, we build upon the rapid sequence-space simulation procedure developed in Auclert, Bardóczy, Rognlie and Straub (2019) by introducing a new and general methodology to handle departures from rational expectations. With these hurdles cleared, we are able to estimate a HANK model featuring sticky expectations, sticky prices and wages with indexation, long-term debt, and investment adjustment costs. We match a) a set of macro monetary policy impulse responses, as in Christiano, Eichenbaum and Evans (2005), and b) a set of macro aggregate time series, as in Smets and Wouters (2007). Our estimation procedure is reliable and fast, even on a laptop computer.

We use our estimated model to revisit the transmission mechanism of monetary policy and the sources of business cycle fluctuations. Our first main finding is that investment is a crucial driver of monetary policy transmission. If investment is constrained not to respond to a monetary shock, the cumulative output response over five years falls by over 80%. By contrast, in a standard representative-agent model with habits estimated using the same data, shutting off investment only causes the output response to decline by investment's accounting share of 40%. This result implies that factors affecting the responsiveness of investment to monetary policy are much more important for aggregate outcomes than previously thought.

This outsized role of investment follows naturally from our strategy of matching both MPCs and macro moments. Since we match the investment impulse response, investment in our model makes a significant contribution to aggregate output demand. As in the traditional Keynesian cross, this leads to a rise in household income, which—thanks to high MPCs—causes consumption to rise, making an additional contribution to output demand. Conceptually, this amplification mechanism only requires elastic investment and high MPCs, as we demonstrate in appendix A.[3]

The quantitative importance of this mechanism, however, hinges on our model's ability to also match the consumption impulse response. Our model achieves this with sticky expectations: it takes time for households to become informed about future macro variables. This dampens ex-ante intertemporal substitution and income effects, which otherwise would lead to a large, immediate consumption response. Once households actually receive the labor income from an output boom, however, they have high MPCs. This leads to a hump-shaped consumption response—driven, in the end, mostly by income effects that originate with the hump-shaped investment response.

We also use our estimated model to investigate two other aspects of monetary transmission. First, we examine the role of fiscal policy, which has been emphasized by Kaplan, Moll and Violante (2018) as an important source of indirect effects on consumption. We show that two features of our model make fiscal policy surprisingly unimportant for monetary transmission: long-term debt that matches the empirical duration in the US, and a more realistically delayed fiscal adjustment instead of a balanced-budget rule. Second, we look at the stock market response, finding that it has the right direction and timing—which has been challenging for HANK models to match, as stressed by Kaplan and Violante (2018)—and that the resulting consumption response is also consistent with the empirical literature.

Our second main finding in the paper is about the sources of business cycles. A fundamental puzzle in the business-cycle literature, dating back to Barro and King (1984), is to explain the procyclical comovement of consumption and investment. The literature has traditionally done so with shocks to TFP or markups: Smets and Wouters (2007), for instance, has a prominent role for both. Although investment shocks can play some role (see Justiniano, Primiceri and Tambalotti 2010), in representative-agent models it is difficult for these shocks to generate procyclical consumption.[4] In our model, on the other hand, the complementarity between high MPCs and investment naturally delivers procyclical consumption. As a result, when we estimate the shock processes underlying the US business cycle, we find a large role for investment shocks: they explain 65% of output variation at business cycle frequencies, vs. only 15% in the representative-agent version of our model. Crucially, they also explain the majority of consumption-investment comovement, vs. almost none in the representative-agent model.

All in all, our paper unifies three strands of the literature. First, we build on a large body

[3]Amplification is greater, however, when intertemporal MPCs are also high, since these lead to intertemporal demand feedbacks (Auclert, Rognlie and Straub 2018). To demonstrate this, we also estimate a two-agent model with habits, calibrated to have the same average MPC, on the same data. Investment is barely more important than in the representative-agent model: shutting off investment causes the output response to decline by 45%. This is low because the two-agent model, unlike our heterogeneous-agent model, fails to match the intertemporal MPCs in figure 1.

[4]One route is to put consumption-labor complementarity into utility (e.g. Furlanetto and Seneca 2014), but as Auclert and Rognlie (2017) point out, such complementarities can interact perversely with sticky prices.

of work that estimates representative-agent models with limited or full-information methods.[5] Second, we build on the active HANK literature pioneered by Gornemann, Kuester and Nakajima (2016), McKay, Nakamura and Steinsson (2016), Auclert (2019), Kaplan, Moll and Violante (2018) and Werning (2015),[6] by estimating a HANK model. Third, we build on the large literature that considers how various deviations from rational expectations can explain the effects of monetary policy (see Angeletos and Lian 2016 for a review).[7] The specific model we use is a type of sticky-information model, as in Gabaix and Laibson (2001) and Mankiw and Reis (2002, 2007), though our version with sticky expectations is closest to Carroll et al. (2018).

Our strategy of combining imperfect information with a strong general equilibrium consumption feedback is closely connected to Angeletos and Huo (2019), who show that these two forces are complements, as the Keynesian cross accentuates inertia at the macro level. It is also closely connected to Farhi and Werning (2019), who were the first to depart from rational expectations in a HANK model, obtaining a solution to the forward guidance puzzle.

An emerging literature also estimates HANK models using limited or full-information methods. Challe, Matheron, Ragot and Rubio-Ramirez (2017) was an early contribution, in a model with restricted heterogeneity. Hagedorn, Manovskii and Mitman (2019b) estimate a one-asset model to match the impulse response to identified technology shocks. They provide an analytical characterization of the role of redistribution and fiscal policy in shaping this impulse response in their model, relative to what a representative-agent model would deliver, and they obtain humps in consumption because their estimated responses of monetary and fiscal policy to technology shocks are themselves hump-shaped. Bayer, Born and Luetticke (2019a) perform a full-information Bayesian estimation of the two-asset HANK model in Bayer, Luetticke, Pham-Dao and Tjaden (2019b). They find that demand shocks play a somewhat larger role in driving business cycle fluctuations in their model relative to the Smets and Wouters (2007) model, and uncover a role for shocks to idiosyncratic income risk and the supply of liquid assets. Relative to their work, our focus is on simultaneously matching MPCs and hump-shapes in consumption, and on the role of investment in the transmission mechanism. We also use a complementary set of methods—sequence-space rather than state-space, which we show to be particularly well-suited to the deviation from rational expectations that we introduce.

Finally, the amplification mechanism at the core of our model—through which shocks to investment affect consumption via MPCs—is analytically distilled in subsequent and complementary work by Bilbiie, Känzig and Surico (2019), in the context of models with tractable heterogeneity. They emphasize the role of cyclical income inequality in further amplifying this channel.

[5]See, among many others, Del Negro, Schorfheide, Smets and Wouters (2007) and Altig, Christiano, Eichenbaum and Lindé (2011), in addition to the work already cited.

[6]See Bilbiie (2008) for an early contribution to this literature, in a model with limited heterogeneity.

[7]This includes models with cognitive discounting as in Gabaix (2016), lack of common knowledge of as in Angeletos and Lian (2018) and Angeletos and Huo (2019), rational inattention as in Sims (2003), Maćkowiak and Wiederholt (2009, 2015) and Zorn (2018), or level-k thinking as in García-Schmidt and Woodford (2019) and Farhi and Werning (2019).

Layout. The paper proceeds as follows. In section 2, we show why sticky expectations can simultaneously rationalize micro jumps and macro humps. In section 3, we present our general equilibrium HANK model. In section 4 we present details of our impulse-matching estimation procedure and discuss model fit. In section 5 we examine the transmission mechanism of monetary policy in our estimated model. In section 6 we enrich the model with a full set of shocks and estimate shock processes to revisit the sources of business cycles. We conclude in section 7.

2 Jumps and humps in consumption

We begin by describing the challenge of simultaneously matching micro jumps and macro humps in consumption. Micro "jumps" are intertemporal MPCs: the micro responses of consumption to transitory income shocks. In the data, these responses peak on impact. Macro "humps" are the aggregate consumption responses to persistent shocks, such as monetary policy shocks. In the data, these responses have delayed peaks.

We first set up a standard heterogeneous-agent model that matches jumps, but cannot generate humps. We then explain why habit formation can match humps, but cannot generate jumps. Finally, we introduce an informational rigidity that can successfully generate both humps and jumps, and will therefore serve as a basis for our general equilibrium model in section 3.

2.1 Heterogeneous households and iMPCs

We model the behavior of a mass one of heterogeneous households in discrete time. The model is in partial equilibrium, without aggregate risk: households face deterministic sequences of interest rates $\{r_t\}$ and average labor incomes $\{y_t\}$. Each household is indexed by its income state s and its liquid asset position ℓ. s follows a Markov process with transition matrix Π, and determines household productivity $e(s)$. Household behavior is characterized by the following dynamic programming problem:

$$V_t(\ell, s) = \max_{c,\ell'} u(c) + \beta \mathbb{E}\left[V_{t+1}(\ell', s') \,|s\right] \tag{1}$$

$$c + \ell' \leq (1 + r_t)\ell + y_t e(s) \tag{2}$$

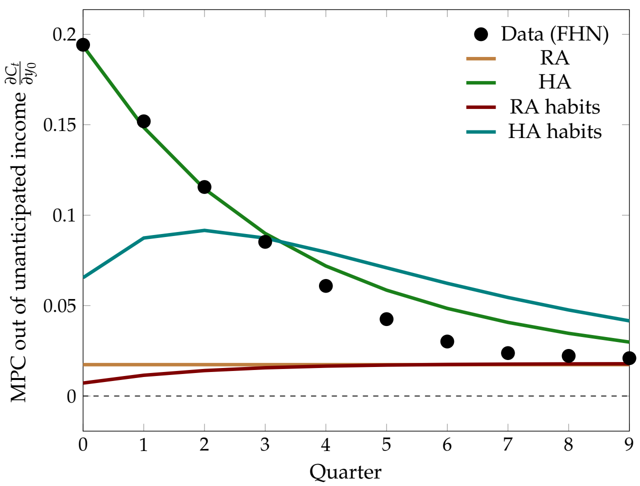
$$\ell' \geq 0 \tag{3}$$

Aggregate consumption C_t is the integral of household consumption choices $c_t(\ell, s)$ over the time-varying cross-sectional distribution of states $\mathcal{D}_t(\ell, s)$. Given an initial distribution $\mathcal{D}_0$, C_t can be expressed as a function of the sequence $\{y_s, r_s\}_{s\geq 0}$:

$$C_t = \mathcal{C}_t\left(\{y_s, r_s\}_{s\geq 0}\right), \quad t \geq 0$$

The consumption function $\mathcal{C}$ summarizes the aggregate behavior of households, holding con-

Figure 2: Intertemporal MPCs $\partial C_t / \partial y_0$ in models and in the data



Note. This figure displays intertemporal MPCs $\partial C_t / \partial y_0$ in various models and in the data. The black dots are constructed by fitting a cubic spline through the estimated cumulative annual iMPCs from Fagereng, Holm and Natvik (2018). The green line displays iMPCs in a baseline heterogeneous agent (HA) model calibrated to match a first-year annual iMPC of 0.55, as per footnote 10. The orange line displays iMPCs in a representative-agent (RA) model with $\beta(1+r) = 1$ and $r = 5\%$. The red line shows iMPCs in an RA model with habit formation in consumption calibrated to match a first-year annual iMPC of 0.55, as described in appendix B.1. The blue line shows iMPCs in a HA model with habit formation in consumption, as described in appendix B.2.

stant primitive parameters such as preferences and the process for idiosyncratic income risk.[8] In particular, the derivatives of $\mathcal{C}$ around the steady state $y_s = y^*, r_s = r^*$ characterize the first-order response of households to changes in aggregate income and real interest rates. These derivatives, or sequence-space Jacobians, are sufficient statistics for the aggregate behavior of households in general equilibrium (Auclert et al. 2019).

As we argued in Auclert, Rognlie and Straub (2018), the Jacobians $\partial C_t / \partial y_s$, which we call *intertemporal marginal propensities to consume* or iMPCs, play a particularly important role. First, they determine fiscal multipliers, and the general equilibrium response to demand shocks more broadly. Second, the impulse response of average consumption to a one-time unanticipated transitory increase in labor income, $\partial C_t / \partial y_0$, can be directly compared with the data and is helpful to discriminate across models. For example, the black dots in Figure 2 represent the quarterly iMPCs implied by the estimates in Fagereng, Holm and Natvik (2018), who combine Norwegian administrative data on income and wealth with data on lottery winnings to estimate $\partial C_t / \partial y_0$.[9] The data iMPCs clearly feature a "micro jump": the average response to an unanticipated income shock is highest on impact and then declines slowly towards zero. The high initial level $\partial C_0 / \partial y_0$ of about 0.2 is consistent with findings from the large empirical literature on MPCs.

The green line in Figure 2 shows the intertemporal MPCs $\partial C_t / \partial y_0$ for a version of the hetero-

[8]See for example Auclert, Rognlie and Straub (2018), Kaplan, Moll and Violante (2018), or Farhi and Werning (2019).

[9]We provide a detailed mapping from data to the model in Auclert, Rognlie and Straub (2018). To convert the estimates from annual to quarterly, we use cubic spline interpolation of the estimated cumulative annual iMPCs.

geneous-agent (HA) model described above, calibrated to match an average annual MPC of 0.55—the point estimate from Fagereng, Holm and Natvik (2018).[10] This strategy achieves a good fit overall, both to the initial level of iMPCs and their subsequent path. Hence, the model is able to replicate the "micro jumps" in the data. This is in direct contrast with a standard representative-agent (RA) model—a version of (1) with no income risk and no borrowing constraint—which predicts a low and flat iMPC schedule, as shown in the orange line of Figure 2.

2.2 Habit formation and iMPCs

One problem with the heterogeneous-agent model just described is that, if we simply embed it into general equilibrium with no further modification, it cannot generate hump-shaped consumption responses to persistent macroeconomic shocks, or "macro humps". This is a property of virtually all HANK models in the existing literature. For example, in all of McKay and Reis (2016), McKay, Nakamura and Steinsson (2016), Gornemann, Kuester and Nakajima (2016), Kaplan, Moll and Violante (2018), Auclert, Rognlie and Straub (2018), and Hagedorn, Manovskii and Mitman (2019a), the consumption response of the economy to monetary and fiscal shocks is peaked on impact, rather than hump-shaped as it typically is in the data. The reason is well-known from the representative-agent literature: standard models of preferences with rational expectations do not have any force delaying the consumption response to aggregate shocks. Yet, such a delay is typically seen as a prerequisite for model estimation. Hence, to make models suitable for estimation, this literature has either modified preferences, or deviated from the assumption of rational expectations.

A very popular approach, pursued by Fuhrer (2000), Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007) among many others, modifies *preferences* to feature habit formation in consumption. This approach successfully slows down the adjustment to macroeconomic shocks, but here we argue it is poorly suited to obtaining a good micro fit.[11,12]

Consider first the assumption of external habits, as in Smets and Wouters (2007). This modifies the utility function $u(c)$ in (1) to read $u(c - \gamma C_{-})$, where C_{-} represents average consumption in the previous period, and a good macro fit requires γ to be a large number like 0.6.[13] In a typical heterogeneous-agent model, this strategy is infeasible: if the model is to be consistent with the steady-state distribution of consumption, it has to feature many agents with consumption far below 0.6 times average consumption, whose marginal utility of consumption would be infinite.[14]

[10]In this calibration, we use the same income process $e(s)$, utility function u and steady-state real interest rate r^* as that of our quantitative model (see section 4). We then find the discount factor β that generates an aggregate MPC in the first year of 0.55, that is, $\sum_{s=0}^{3} \left(\frac{1}{1+r^*}\right)^s \frac{\partial C_s}{\partial y_0} = 0.55$. This procedure yields $\beta = 0.8422$ at an annual rate.

[11]We focus on additive habits, which are used by the majority of papers in the macro estimation literature. Multiplicative, external habits may provide an alternative to our solution of sticky expectations.

[12]In a related argument, Havranek, Rusnak and Sokolova (2017) show that micro and macro approaches to estimating habits tend to produce inconsistent estimates.

[13]For example, in the estimated Smets-Wouters model, $\gamma = 0.71$.

[14]Another way to see this is that the steady state of the model would feature a counterfactually high degree of precautionary savings, and therefore a counterfactual distribution of consumption, as all agents save enough to ensure

Consider next the alternative assumption of internal habits, as in Christiano, Eichenbaum and Evans (2005), which replaces the utility function $u(c)$ in (1) by $u(c - \gamma c_-)$, where c_- is the agent's own consumption in the previous period, and can therefore scale differently for rich and poor agents. The problem with this specification is that it substantially lowers marginal propensities to consume. In appendix B.1, we prove this formally for the case of the RA model: there, the impact MPC is reduced by a fraction $1 - \beta\gamma$ relative to the standard RA MPC. The reason is visible in the red line of figure 2: in response to a one-time income transfer, agents limit their initial increase in consumption to avoid raising their consumption habit too rapidly. The resulting pattern of iMPCs is increasing over time, the opposite of the jump in the data.

The intuition that habits reduce impact MPCs carries over to a HA model with internal additive habits. To make this point, we set up such a model with $\gamma = 0.6$ and otherwise calibrate it to the same parameters as our no-habit HA model (see appendix B.2 for details). As the blue line of figure 2 illustrates, that model features a micro hump: iMPCs themselves are hump-shaped, in contrast to the data.

These results lead us to pursue a different strategy to model hump-shaped responses to macroeconomic shocks, one that relies on a deviation from *rational expectations*. This route has the additional advantage that it does not involve any change to the ability of the HA model to match iMPCs.

2.3 Sticky expectations and iMPCs

We propose to extend the standard HA model with a version of the "sticky information" friction pioneered by Gabaix and Laibson (2001) and Mankiw and Reis (2002, 2007). The exact version we use is Carroll et al. (2018)'s formulation of *sticky expectations*. In order to formally express it, we work with an aggregate-risk version of the household problem (1) in which both r_t and y_t follow stochastic processes that are orthogonal to idiosyncratic risk.

In the sticky-expectations model, households update their information sets about aggregate shocks infrequently, with an iid probability of $1 - \theta$ each period. At time t, a household that last updated $k \geq 0$ periods ago bases its forecast of the future on the information available at time $t - k$. The parameter θ indicates the stickiness of expectations in the model, and rational expectations correspond to the special case where $\theta = 0$.

We assume that only expectations about future aggregates are sticky, not expectations about future idiosyncratic productivity shocks $e(s)$. This reflects the idea, formalized in the rational inattention literature, that idiosyncratic shocks have a much greater variability and therefore receive more attention by households (see e.g. Maćkowiak and Wiederholt 2009). A complementary interpretation, from Angeletos and Huo (2019), is that incomplete information about idiosyncratic shocks is less relevant because, unlike with aggregate shocks, there is no general equilibrium feedback that amplifies the inertia.

minimal consumption of γC_- in the face of large variability in incomes.

Formally, households solve, subject to the unchanged constraints (2)–(3):

$$V_t(\ell, s, k) = \max_{c, \ell'} u(c) + \beta \mathbb{E}_{t-k}\left[\theta V_{t+1}(\ell', s', k+1) + (1-\theta) V_{t+1}(\ell', s', 0) \,|s\right] \tag{4}$$

The number of periods k since the last information update enters as a state variable, which matters because it shapes the expectation $\mathbb{E}_{t-k}$ on the right. Since households are aware that, with probability θ, they will keep a stale information set and move to $k' = k+1$, while with probability $1-\theta$ they they will update their information set next period and move to state $k' = 0$, the expectation is taken over a convex combination of $V_{t+1}(\ell', s', k+1)$ and $V_{t+1}(\ell', s', 0)$. Note that as in Carroll et al. (2018), households always observe current r_t and y_t, ensuring that they know their current cash on hand $(1+r_t)\ell + y_t e(s)$ and do not violate their borrowing constraints.

There are two interpretations of the information friction embedded in (4). First, non-adjusting households might not use the information available in current aggregates like r_t and y_t to update their expectations about future aggregates. Alternatively, these households might update their expectations about future aggregates, but fail to incorporate this information into their consumption-savings decision, instead anchoring their decision rule to past expectations. Empirical estimates of sluggish expectation adjustment miss the latter possibility, and thus provide only a lower bound for the θ in our model.

The sticky-expectation formulation of informational rigidities has two main advantages for our purposes. First, intertemporal MPCs are unchanged[15] by sticky expectations: a one-time unanticipated income shock does not change future incomes or interest rates, and therefore does not interact with frictions in the adjustment of these expectations over time. Second, and more importantly, slow adjustment of expectations allows us to model hump-shaped impulse responses to macroeconomic shocks. We turn to this exercise next.[16]

3 An inattentive HANK model

We now embed our model of a population of heterogeneous households with sticky expectations—which we will call "inattentive" for brevity—into general equilibrium.

To maximize comparability with the existing representative-agent literature, we structure the rest of our general equilibrium model following Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007). In addition to the heterogeneous-agent inattentive household sector, we depart from these papers by abstracting away from endogenous capacity utilization, fixed costs in production, and a "cost channel" for monetary policy, since we did not find any of these

[15] That is, feeding a realized path of average incomes $\{y+\epsilon, y, y, y, \dots\}$ with constant r and small $\epsilon > 0$ into the economy at its deterministic steady state always generates a response of average consumption equal to that shown in Figure 2, irrespective of θ.

[16] The formulation in (4) is also consistent with several pieces of evidence. First, at the macro level, evidence from the cross section of monetary impulse responses (eg Cloyne, Ferreira and Surico 2020, Wong 2018) suggest that impulse responses are hump-shaped within groups, not just across groups, which our formulation delivers. Second, at the micro level, our formulation is consistent with the sluggish adjustment of household expectations documented in Coibion and Gorodnichenko (2012, 2015) (see section 4.4).

three features essential to match the impulse response. But we maintain the key features of these models that generate macroeconomic sluggishness in inflation (namely, price and wage indexation) and investment (namely, investment adjustment costs).[17]

3.1 Inattentive households

We build on the model presented in section 2, extending it by letting households hold a second, illiquid asset, and allowing for heterogeneity in illiquid asset holdings.

The economy is populated by a unit mass of households that face both idiosyncratic and aggregate uncertainty. Households transition stochastically between idiosyncratic productivity states s according to a Markov process with fixed transition matrix Π. The mass of households in state s is always equal to π_s, the probability of s in the stationary distribution of Π. We normalize idiosyncratic productivity levels $e(s)$ to have mean one, so $\sum_s \pi_s e(s) = 1$.

In addition to stochastic heterogeneity, we also allow for permanent heterogeneity. Households are each assigned a permanent type, g, such that there is a mass μ_g of households of type g. The permanent type influences three household attributes: their discount factor β_g; their group-average skill level $\bar{e}_g$; and their steady-state level of aggregate illiquid asset holdings $\bar{a}_g$. This permanent heterogeneity allows us to fit the large heterogeneity in illiquid asset holdings in the data while maintaining a simplified model of households' decisions to save in the illiquid asset. We normalize $\sum_g \mu_g = 1$.

In each period, a household enjoys the consumption of a generic consumption good c and gets disutility from working n hours according to the separable felicity function $u(c) - v(n)$. Because of sticky wages, n is determined by union labor demand. As described in section 3.4, unions allocate labor equally across all households, so that for every household, $n_t = N_t$ in period t, where N_t is aggregate labor demand. Individual after-tax labor income in period t is therefore given by $z_t \equiv Z_t \bar{e}_g e(s)$, where $Z_t \equiv (1 - \tau_t) w_t N_t$ is aggregate after-tax labor income.

Each household has access to two assets. As in section 2, they can trade in a liquid asset, ℓ, which has a return r_t^ℓ between $t-1$ and t that is predetermined in period $t-1$, and they are subject to a zero-borrowing constraint. In equilibrium, this asset will earn a relatively low return. They also hold an illiquid asset, a, which carries a higher equilibrium return r_t^a. In the non-stochastic steady state of the economy, agents behave in such a way that all agents in group g hold the same amount $\bar{a}_g$ in illiquid assets. Both liquid and illiquid assets are issued by a financial intermediary, which is introduced in section 3.2 below.

Formally, a household in group g and state s, with liquid assets ℓ, illiquid assets a, and who observed the aggregate state k periods ago when it had illiquid assets a_{-k}, solves the dynamic

[17] An alternative route to generate sluggishness in inflation and investment is to assume inattentiveness of unions and firms. We followed this route in a previous version of this paper. This alternative specification did not affect the main results of the paper.

programming problem

$$
\begin{aligned}
V_{g,t}\left(\ell, a, a_{-k}, s, k\right) \quad &= \quad \max_{c,\ell'} u(c) - v(N_t) + \beta_g \mathbb{E}_{t-k}\left[\theta V_{g,t+1}\left(\ell', a', a_{-k}, s', k+1\right)\right. \\
&\qquad\qquad \left. + \left(1-\theta\right) V_{g,t+1}\left(\ell', a', a', s', 0\right) |s\right] \\
&\quad c + \ell' = \left(1 + r_t^{\ell}\right)\ell + Z_t \bar{e}_g e(s) + d_{g,t}\left(a_{-k}, k\right) \\
&\quad a' = \left(1 + r_t^a\right) a - d_{g,t}\left(a_{-k}, k\right) \\
&\quad \ell' \geq 0
\end{aligned}
\tag{5}
$$

where θ denotes the stickiness of household expectations—the probability that they do not update their information about aggregates or the value of their illiquid account—and $d_{g,t}$ is a distribution from the illiquid account, described below. We assume θ to be the same across all groups.

While the return on the liquid asset is set by the financial intermediary in period $t-1$, the return on the illiquid asset r_t^a is stochastic and moves around with the value of the stock market. We assume that households update their information on the value of their illiquid account a infrequently, at the same time they update their expectations.[18] They then distribute an amount $d_{g,t}$ from their illiquid to their liquid account that depends on $a^e = a_t^e\left(a_{-k}, k\right) \equiv \mathbb{E}_{t-k}\left[\left(1+r_t^a\right) a | a_{-k}\right]$, the value of the portfolio they expect to have given their last information update. Specifically, we assume that households follow the rule

$$
d_{g,t}\left(a_{-k}, k\right) = \frac{r^{a,ss}}{1 + r^{a,ss}} a^e + \chi\left(a^e - \left(1 + r^{a,ss}\right)\bar{a}_g\right) \tag{6}
$$

where $r^{a,ss}$ is the steady-state level of illiquid returns and χ is a small, positive constant. The first term in (6) states that households increase their distributions only when they observe a higher value of the stock market, and by a fraction corresponding to the steady-state annuity value of this increase. As we show in section 5.5, this ensures that the marginal propensity to consume out of capital gains is small in the aggregate, and that the consumption response is delayed, both consistent with empirical evidence. The second term in (6) ensures that households save when they expect their portfolio value to be low, and dissave when they expect it to be high, so that in the long run their liquid asset position is equal to $\bar{a}_g$. We set the parameter χ to be positive but very close to 0.[19]

This set of assumptions delivers a two-asset model that is complex enough to deliver micro jumps (with MPCs as high as in the data) and macro humps (with consumption following a hump-shaped path after shocks), yet tractable enough to be estimated using the methodology developed in section 4.3.

[18] There is empirical evidence that households infrequently choose to observe the value of their financial assets (for example Alvarez, Guiso and Lippi 2012). Moreover, we show in section 5.5 that, at the aggregate level, our model predicts the pattern for the consumption response to capital gains documented in Chodorow-Reich, Nenov and Simsek (2019). Even though households are inattentive, the stock market is always priced by competitive, fully informed and rational financial traders—see section 3.2.

[19] Different small values of χ yield numerically identical results. A similar device is used in incomplete-market small open economy models to induce stationarity (see e.g. Schmitt-Grohé and Uribe 2003).

3.2 Financial intermediary

The financial intermediary in our model has two activities: a banking activity, performing maturity transformation by collecting liquid short-term deposits L_t and investing in long-term government debt B_t subject to an intermediation cost of ξL_t paid in the next period, and a mutual fund activity, collecting illiquid funds A_t and investing them in government bonds and shares in firms.

The consolidated representative financial intermediary faces the following sequence of flow-of-funds constraints. At the beginning of the period, the value of outstanding illiquid and liquid liabilities must be equal to the liquidation value of the portfolio of government bonds and shares in firms, net of the intermediation cost of liquid deposits, so that:

$$(1 + r_t^a)\, A_{t-1} + \left(1 + r_t^\ell\right) L_{t-1} = (1 + \delta q_t)\, B_{t-1} + \int \left(p_{jt} + D_{jt}\right) v_{jt-1} dj - \xi L_{t-1} \tag{7}$$

where v_{jt} denotes the shares of firm j with price p_{jt}. At the end of the period, the value of newly-purchased bonds and shares must be equal to the value of newly issued liquid and illiquid liabilities, so that:

$$\int p_{jt} v_{jt} dj + q_t B_t = A_t + L_t \tag{8}$$

The financial intermediary's problem is to choose v_{jt}, B_t and L_t so as to maximize the expected return on illiquid liabilities, $\mathbb{E}_t\left[r_{t+1}^a\right]$.[20] This leads to the following asset pricing equations:

$$\mathbb{E}_t\left[1 + r_{t+1}^a\right] = \frac{\mathbb{E}_t\left[1 + \delta q_{t+1}\right]}{q_t} = \frac{\mathbb{E}_t\left[p_{jt+1} + D_{jt+1}\right]}{p_{jt}} = 1 + r_{t+1}^\ell + \xi \equiv 1 + r_t \tag{9}$$

where we have defined r_t as the ex-ante real interest rate. Hence, in equilibrium, competitive intermediaries fully pass through the cost of deposit issuance to a lower deposit interest rate, $r_{t+1}^\ell = r_t - \xi$, and they equalize the expected returns on all other assets to the ex-ante real interest rate.

We also allow the financial intermediary to invest in nominal reserves issued by the central bank, that are in zero net supply and pay an interest rate i_t between t and $t+1$. In appendix C.2, we show that the first order condition with respect to reserve holdings implies the Fisher equation:

$$1 + r_t = (1 + i_t)\, \mathbb{E}_t\left[\frac{P_{t+1}}{P_t}\right] \tag{10}$$

Our assumption of perfect attention for the mutual fund is consistent with immediate reaction of financial markets to news, as documented, for example, by Bernanke and Kuttner (2005), Gürkaynak, Sack and Swanson (2005), and Nakamura and Steinsson (2013) for monetary policy.

[20] Here we assume that the financial intermediary uses a risk-neutral stochastic discount factor. However, since we are solving to first order in aggregates, we would obtain an identical solution under any alternative choice of stochastic discount factor M_t such that the intermediary maximizes $\mathbb{E}_t\left[M_{t+1} r_{t+1}^a\right]$.

3.3 Firms

Our specification of firms mostly follows Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007), except that we abstract from endogenous capacity utilization, fixed costs in production, and a cost channel of monetary policy.

Final good. A final good firm produces the homogeneous output good Y_t out of intermediate goods. Its production function is implicitly defined by the equation

$$\int_0^1 G_p\left(\frac{Y_{jt}}{Y_t}\right) dj = 1 \tag{11}$$

where G_p is the Klenow and Willis (2016) version of the Kimball (1995) aggregator, which satisfies $G_p(1) = 1$ and $G_p'(x) = \exp\left\{\frac{1-x^{\frac{v_p}{\epsilon_p}}}{v_p}\right\}$. v_p is the demand superelasticity, with $v_p = 0$ corresponding to standard CES demand.

Intermediate goods. All intermediate goods producers operate a common Cobb-Douglas production function under monopolistic competition and constant productivity Θ,

$$Y_{jt} = \Theta K_{jt}^{\alpha} N_{jt}^{1-\alpha}$$

They hire capital and labor from a common market. Hence all firms have the same capital-labor ratio $\frac{K_{jt}}{N_{jt}} = \frac{K_t}{N_t}$ and the same real marginal cost s_t, such that the real wage $w_t = \frac{W_t}{P_t}$ (the ratio of the aggregate nominal wage W_t and the price index for final output, P_t) and the rental rate on capital r_t^K are equal, respectively, to

$$\begin{aligned} w_t &= s_t(1-\alpha)\Theta K_t^{\alpha} N_t^{-\alpha} \\ r_t^K &= s_t \alpha \Theta K_t^{\alpha-1} N_t^{1-\alpha} \end{aligned}$$

We assume that prices are sticky à la Calvo and are fully indexed to inflation. With probability $1-\zeta_p$, firms are free to reset their price P_{jt}. From the financial intermediary, they face a menu of stock prices $p_{jt}(P_{jt})$ conditional on resetting their price to P_{jt}, and they choose P_{jt} so as to maximize the sum of stock price $p_{jt}(P_{jt})$ and dividend

$$D_{jt} = \left(\frac{P_{jt}}{P_t} - s_t\right) Y_{jt}$$

The fraction ξ_p of firms that do not reset their price index it to the previous period's inflation, so that their price follows

$$P_{jt} = \Pi_{t-1} P_{j,t-1} \tag{12}$$

where $\Pi_{t-1} \equiv \frac{P_{t-1}}{P_t}$. Appendix C.3 shows that optimal price-setting of firms, given the discount factor from (9), generates an indexed Phillips curve, which to first order is:

$$\pi_t - \pi_{t-1} = \frac{\left(1-\zeta_p\right)\left(1-\frac{\zeta_p}{1+r}\right)}{\zeta_p}\frac{\epsilon_p}{\nu_p+\epsilon_p-1}\mathbb{E}_t\left[\sum_k\left(\frac{1}{1+r}\right)^k\left(s_{t+k}-\frac{\epsilon_p-1}{\epsilon_p}\right)\right] \tag{13}$$

where $\pi_t \equiv \log\left(\Pi_t\right)$. As is well-known, a higher Kimball superelasticity $\nu_p > 0$ results in larger real rigidities, and a lower slope for the Phillips curve in (13). In the special case where $\xi_p = 0$ and prices are fully flexible, all firms set the same price at a constant markup $\frac{\epsilon_p}{\epsilon_p-1}$ over nominal marginal costs, and real marginal costs are equal to the constant $\frac{\epsilon_p-1}{\epsilon_p}$.

Capital firms. A capital firm owns the capital stock and rents it at rate r_t^K. It faces adjustment costs and one-period time-to-build in investment. When it has capital stock K_t and pre-planned investment I_t, this firm pays a dividend of

$$D_t^K = r_t^K K_t - I_t\left(1+S\left(\frac{I_t}{I_{t-1}}\right)\right)$$

where the investment adjustment cost function satisfies $S(1) = S'(1) = 0, S''(1) = \phi$, and it enters the next period with a capital stock of

$$K_{t+1} = \left(1-\delta\right)K_t + I_t$$

At date t, it chooses investment for next period I_{t+1} to maximize the sum of its dividend D_t^K and its stock price $p_t^K\left(K_{t+1}, I_{t+1}\right)$, taking as given the menu $p_t^K\left(\cdot,\cdot\right)$ of stock prices conditional on capital stocks and investment levels. Appendix C.4 shows that, if we define $Q_t \equiv \mathbb{E}_t\left[\frac{\partial p_{t+1}^K}{\partial K_{t+2}}\right]$, investment dynamics are characterized by the following set of equations:

$$1+S\left(\frac{I_{t+1}}{I_t}\right)+\frac{I_{t+1}}{I_t}S'\left(\frac{I_{t+1}}{I_t}\right) = Q_t + \mathbb{E}_t\left[\frac{1}{1+r_{t+1}}\left(\frac{I_{t+2}}{I_{t+1}}\right)^2 S'\left(\frac{I_{t+2}}{I_{t+1}}\right)\right] \tag{14}$$

and

$$Q_t = \mathbb{E}_t\left[\frac{1}{1+r_{t+1}}\left(r_{t+2}^K + \left(1-\delta\right)Q_{t+1}\right)\right] \tag{15}$$

Aggregate firm value. There is a mass 1 of outstanding shares in both intermediate goods and capital firms. The arbitrage conditions (9) ensure that the shares in intermediate goods firms are priced at $p_{jt} = \frac{1}{1+r_t}\mathbb{E}_t\left[p_{jt+1}+D_{j,t+1}\right]$ and the shares in capital firms are priced at $p_t^K = \frac{1}{1+r_t}\mathbb{E}_t\left[D_{t+1}^K + p_{t+1}^K\right]$. Aggregate dividends are equal to

$$D_t = \int D_{jt}dj + D_t^K = Y_t - w_t L_t - I_t\left(1+S\left(\frac{I_t}{I_{t-1}}\right)\right) \tag{16}$$

and the value of the aggregate stock market, $p_t = \int p_{jt} dj + p_t^K$, satisfies the equation

$$p_t = \frac{1}{1+r_t} \mathbb{E}_t \left[D_{t+1} + p_{t+1} \right] \tag{17}$$

3.4 Unions

We follow standard practice in the New Keynesian sticky-wage literature and assume that household labor hours n_{it} are determined by union labor demand (Erceg, Henderson and Levin 2000, Schmitt-Grohé and Uribe 2005, Auclert, Rognlie and Straub 2018). Specifically, we assume that in each period t, each household i provides n_{ijt} hours of work to each of a continuum of unions indexed by $j \in [0,1]$, with $n_{it} = \int n_{ijt} dj$.

Each union j aggregates efficient units of work into a union-specific task $N_{jt} = \int \bar{e}_i e(s_{it}) n_{ijt} di$. A competitive labor packer then packages these tasks into aggregate employment services using the Kimball technology implicitly defined by

$$\int_j G_w \left(\frac{N_{jt}}{N_t} \right) dj = 1$$

where $G_w'(x) = \exp \left\{ \frac{1 - x^{\frac{\upsilon_w}{\epsilon_w}}}{\upsilon_w} \right\}$, and sells these services at price W_t. Unions set wages à la Calvo, resetting wages with probability $1 - \zeta_w$ per period, and indexing to past price inflation when they cannot reset wages, as in Smets and Wouters (2007). In every period, unions call upon their members to supply hours according to a uniform rule, $n_{ijt} = N_{jt}$. Whenever it can reoptimize, a union sets wages to maximize the average utility of its members given this rule. In appendix C.5, we show that union maximization leads to a Phillips curve for wage inflation, which to first order is:

$$\pi_{w,t} - \pi_{t-1} = \frac{(1-\beta\zeta_w)(1-\zeta_w)}{\zeta_w} \frac{\epsilon_w}{\epsilon_w + \upsilon_w - 1} \mathbb{E}_t \left[\sum_k \beta^k \left(s_{w,t+k} - \frac{\epsilon_w - 1}{\epsilon_w} \right) \right] \tag{18}$$

where $\pi_{w,t} \equiv \log \left(\frac{W_t}{W_{t-1}} \right)$, and $s_{w,t} \equiv \frac{\int v'(n_{it}) di}{(1-\tau_t) w_t \int e_{it} u'(c_{it}) di}$ is the ratio of the average marginal disutility of labor to an average income-weighted marginal utility of consumption, divided by the net-of-tax real wage. In the special case where $\xi_w = 0$ and wages are fully flexible, all unions set the inverse wage markup to the constant $\frac{\epsilon_w - 1}{\epsilon_w}$.

3.5 Government policy

The government issues long-term bonds B_t, collects labor income taxes $\tau_t \frac{W_t}{P_t} N_t \mathbb{E} \left[\bar{e}_g e(s) \right] = \tau_t \frac{W_t}{P_t} N_t$, and spends on goods and services G_t. Its budget constraint is

$$q_t B_t + \tau_t \frac{W_t}{P_t} N_t = G_t + (1 + \delta q_t) B_{t-1} \tag{19}$$

Given the lack of Ricardian equivalence in our model, the fiscal rule of the government matters. Since our focus for the next section is on monetary policy rather than fiscal policy, we assume that the government sets its spending at a constant

$$G_t = G$$

and follows a rule for the tax rate

$$\tau_t - \tau^{ss} = \psi q^{ss} \frac{(B_{t-1} - B^{ss})}{Y^{ss}} \tag{20}$$

such that the tax rate rises when debt is above its long-run level, thereby eventually bringing debt back down. The parameter ψ governs the speed of this adjustment. We consider several alternative fiscal rules in section 5.3.

Finally, we assume that monetary policy follows a conventional inertial Taylor rule for the nominal interest rate:

$$1 + i_t = (1 + r^{ss})^{1-\rho^m} (1 + i_{t-1})^{\rho^m} \left(\frac{P_t}{P_{t-1}} \right)^{(1-\rho^m)\phi} (1 + \epsilon_t^m)$$

where r^{ss} is the steady-state real interest rate, $\rho^m \in [0, 1)$ is the persistence of the policy rate, and ϵ_t^m is a monetary policy shock. Given these elements, the definition of equilibrium is standard:

Definition. Given a stochastic process for the monetary policy shock ϵ_t^m, an initial nominal wage W_{-1} and price level P_{-1}, initial government debt B_{-1}, an initial capital level K_{-1}, and an initial distribution of agents $\mathcal{D}_{g,0}(\ell, a, a_{-k}, s, k)$ in each fixed group g, a *competitive equilibrium* is a stochastic sequence of prices $\{P_t, W_t, \pi_t, \pi_{w,t}, w_t, r_t^K, s_t, r_t, r_t^\ell, r_t^a, p_t, i_t, q_t, Q_t\}$, aggregates $\{Y_t, N_t, I_t, K_t, C_t, L_t, A_t, G_t, \tau_t, D_t\}$, individual policy rules $\{c_{g,t}(\ell, a, a_{-k}, s, k), \ell_{g,t+1}(\ell, a, a_{-k}, s, k), a_{g,t+1}(\ell, a, a_{-k}, s, k)\}$, and joint distributions of agents $\mathcal{D}_{g,t}(\ell, a, a_{-k}, s, k)$ such that households optimize, the financial intermediary optimizes, all firms optimize, unions optimize, monetary and fiscal policy follow their rules, and asset markets clear:

$$\begin{aligned} \sum_{g,s,k} \mu_g \pi_s (1-\theta)\theta^k \int \ell \mathcal{D}_{g,t}(d\ell, da, da_{-k}, s, k) &= L_t \\ \sum_{g,s,k} \mu_g \pi_s (1-\theta)\theta^k \int a \mathcal{D}_{g,t}(d\ell, da, da_{-k}, s, k) &= A_t \end{aligned}$$

Appendix C.6 shows that, when these conditions are satisfied, the goods market also clears:

$$C_t + G_t + I_t + I_t S\left(\frac{I_t}{I_{t-1}}\right) + \xi L_{t-1} = Y_t$$

which is a statement that aggregate demand for final goods—the sum of consumption, government spending, investment including adjustment costs, and liquidity costs—must be equal to total production of these goods.

Representative-agent model. It will be informative to contrast the predictions of our model with those of a standard representative-agent model similar to Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007). For this comparison, the model we use is identical to that described above, except that we replace the inattentive household sector in section 3.1 with a perfectly attentive representative agent with internal additive habits. We restrict this agent not to invest in the liquid asset, so $\ell = 0$ at all times, and hence in equilibrium $L_t = 0$.[21] Hence, the representative agent solves

$$\begin{aligned} V_t(a, c_-) &= \max_{c,a'} u(c - \gamma c_-) + \beta \mathbb{E}_t \left[V_{t+1}(a', c) \right] \\ & \quad c + a' = (1 + r_t^a) a + Z_t \end{aligned}$$

with no constraint on a, leading to an Euler equation described in appendix B.1. Given this, the definition of equilibrium is straightforward. In what follows, we will refer to this model as the "RA-habit" model, or RA model for short.

4 Estimating HANK

With the model set up, we are now ready to describe our estimation procedure.

4.1 Two-step estimation procedure

We follow a two-step procedure to estimate our model, very similar to the one followed by Christiano, Eichenbaum and Evans (2005). In the first step, we *calibrate* some parameters, including all those that are relevant for the steady state of our model. In the second step, we *estimate* the remaining parameters by matching impulse responses from an identified monetary policy shock.

The general philosophy guiding our calibration is that we are interested in the average transmission of monetary policy over the last fifty years in the United States, from 1969 to the present-day 2019. Whenever possible, our steady state captures the average macroeconomic environment over this period. However, since there has not been much signal from monetary shocks in recent years (see the discussion in Ramey 2016), our impulse responses to monetary policy shocks are obtained using the original sample from Romer and Romer (2004), 1969 through 1996.

First step: Calibration. The left panel of table 2 summarizes the calibrated parameters of our model. Our households have constant CES utility over consumption $u(c) = \frac{c^{1-\nu^{-1}}-1}{1-\nu^{-1}}$ with an EIS of $\nu = 1$, and a power disutility from labor $v(n) = v_0 n^{1+\varsigma^{-1}}/(1+\varsigma^{-1})$ with Frisch elasticity of $\varsigma = 0.5$. We set government spending to $\frac{G}{Y} = 16\%$ of output, the average of government consumption expenditures over GDP in the period from 1969 to 2019, and use v_0 to normalize the level of output in the no-inflation steady state to $Y = 1$. We assume that the annual steady-state

[21] Since the liquid asset delivers inferior average returns, under perfect foresight about future aggregates the agent would endogenously choose not to hold it, and instead stay at the constraint $\ell = 0$.

Table 1: Calibrating permanent household heterogeneity

Household group g	1	2	3	4	5	6
Population share (μ_g)	Bottom 50%	Next 20%	Next 10%	Next 10%	Next 5%	Top 5%
Illiquid asset share	2.7%	7.0%	7.0%	13.0%	12.2%	58.0%
Labor income share	26.7%	18.3%	10.8%	14.4%	11.0%	18.8%
Discount factors (p.a.)	0.905	0.919	0.933	0.946	0.950	0.975

real interest rate is $r = 5\%$. This corresponds to the average combined real return on capital and government bonds over the same period. We also assume an intermediation spread of $\xi = -6.5\%$, consistent with an average effective real return on deposits of -1.5%.[22]

We assume $G = 6$ permanent household groups to capture the very uneven distribution of illiquid asset holdings in the United States. Recall that, in the steady state of our model, all households within a group have the same amount of illiquid assets $\overline{a}_g$, so there is no steady-state within-group inequality in illiquid assets. Nevertheless, we aim to capture between-group inequality in illiquid assets by appropriately selecting groups. We define these groups as cuts of the illiquid asset distribution in the 2013 Survey of Consumer Finances (Bricker et al. 2014). We define illiquid assets to include retirement accounts, certificates of deposit, the cash value of life insurance, savings bonds, and total non-financial assets net of mortgages on all residences. Households in the bottom 50% of the illiquid asset distribution hold virtually no assets and comprise our first group $g = 1$. Group $g = 2$ contains the next 20% of households, groups $g = 3, 4$ each contain the next 10% of households, and groups $g = 5, 6$ contain the next 5% and the top 5% of households. For each of these groups, we compute the share of total illiquid assets held by the group, as well as the share of total labor income earned by the group, and report these shares in Table 1. The table reflects the well-known fact that the distribution of illiquid assets is heavily skewed at the top. For example, the bottom 50% of the illiquid asset distribution own only 2.7% of total illiquid assets, though they earn 27% of total labor income. Conversely, the top 5% of the illiquid asset distribution own 58% of illiquid assets, but earn 19% of labor income.

There is little evidence that MPCs vary systematically along the illiquid asset distribution. Instead, the evidence presented in the literature (for example, Johnson, Parker and Souleles 2006, Misra and Surico 2014, Kaplan and Violante 2014, and Fagereng, Holm and Natvik 2018) is consistent with the hypothesis that the average MPC is constant across the distribution of illiquid asset holdings. Furthermore, as discussed in section 2, targeting an average annual MPC of 0.55 delivers a good fit for the path of iMPCs, our central "micro jump" target. These observations lead us to calibrate our model to achieve an average annual MPC of 0.55 for each of our six permanent groups, which we accomplish with group-specific discount factors β_g, along with an aggregate ratio of liquidity to GDP of $\frac{L}{Y} = 0.23$. The discount factors are reported in the last row of table 1.

Our process for gross income is a quarterly discretization of the process estimated in Kaplan, Moll and Violante (2018), with one modification: we rescale the variance of innovations by one

[22] This corresponds to the average of the MZM own rate between 1974 and 2019, minus average realized inflation.

Table 2: Calibrated and estimated parameters.

Panel A: Calibrated parameters

Parameter		Value
ν	EIS	1
ς	Frisch	0.5
β_g	Discount factors (p.a.)	Table 1
r	Real interest rate (p.a.)	0.050
α	Capital share	0.24
δ_K	Depreciation of capital (p.a.)	0.053
μ_p	Steady-state retail price markup	1.06
K/Y	Capital to GDP (p.a.)	2.23
L/Y	Liquid assets to GDP (p.a.)	0.23
ξ	Intermediation spread (p.a.)	0.065
G/Y	Spending-to-GDP	0.16
qB/Y	Government bonds to GDP (p.a.)	0.42
$\frac{1+r}{1+r-\delta}$	Maturity of government debt (a.)	5
ψ	Response of tax rate to debt (p.a.)	0.1
ϕ_π	Taylor rule coefficient	1.5

Panel B: Estimated parameters

Parameter		Value	std. dev.
θ	Household inattention	0.935	(0.01)
ϕ	Investment adj. cost parameter	9.639	(2.428)
ζ_p	Calvo price stickiness	0.926	(0.012)
ζ_w	Calvo wage stickiness	0.899	(0.016)
ρ^m	Taylor rule inertia	0.890	(0.01)
σ^m	Std. dev. of monetary shock	0.057	(0.005)

minus the ratio of within-group variance to total variance in order to maintain consistency with the aggregate variance of log earnings in the data. To account for progressive taxation, we then further scale down the variance of innovations by $(1-0.18)^2$, where 0.18 is the degree of tax progressivity in Heathcote, Storesletten and Violante (2017).

Our calibration for the supply side of the model is as follows, using averages over the 1969–2019 period for all targets. We assume an annualized rate of capital depreciation of $\delta_K = 5.3\%$, equal to the average ratio of depreciation to private fixed assets. We calibrate α to achieve a capital-to-output ratio of 223%, yielding $\alpha = 0.24$. The value of government debt to GDP is set to 46%, and the decay rate δ of government bond coupons is set to match the average duration of US government debt of 5 years.[23] We calibrate the fiscal rule parameter to $\psi = 0.1$ following empirical estimates from the fiscal rule literature (see appendix D.1).[24] We calibrate aggregate household wealth to GDP to 382%, its average in the data, then use this to back out the steady-state markup μ_p. Since the ratio of the capitalized value of markups to GDP is $\left(1-\frac{1}{\mu_p}\right)\frac{1}{r} = 3.82 - 2.23 - 0.46$, this gives $\mu_p = 1.06$. Finally, we follow the standard procedure in the literature (e.g. Smets and Wouters 2007) of assuming a wage markup of $\frac{\epsilon_w}{\epsilon_w - 1} = \frac{\epsilon_p}{\epsilon_p - 1}$, and a Kimball superelasticity for prices and wages of $v_p = v_w = 10$.

Second step: Estimation. We estimate the remaining parameters by matching the impulse responses to identified monetary policy shocks, as in Christiano, Eichenbaum and Evans (2005). We

[23] The modified duration of bonds in the model is $\frac{1+r}{1+r-\delta}$ so that, at an annual rate, $\delta = (1+r)\left(1-\frac{1}{5}\right)$.

[24] As discussed in appendix D.1, our quantitative results are not sensitive to ψ lying within a wide but empirically reasonable range.

assume that in our model, the identified shocks correspond to iid shocks to the Taylor rule ϵ_t^m with standard deviation σ^m, which have endogenously persistent effects on i_t through inertia in the Taylor rule.

The set of parameters to be estimated then includes the degree of household inattention θ, the curvature of the adjustment cost function ϕ, the Calvo parameter for the stickiness of prices ζ_p and wages ζ_w, inertia ρ^m in the Taylor rule, as well as the standard deviation of monetary shocks σ^m. Collecting these parameters in the vector $\Psi \equiv \left(\theta, \phi, \zeta_p, \zeta_w, \rho^m, \sigma^m\right)'$, let $\mathbf{J}\left(\Psi\right)$ denote the model-implied first-order impulse responses and $\widehat{\mathbf{J}}$ their empirical counterpart. Our estimator $\widehat{\Psi}$ solves

$$\min_{\Psi} \left(\mathbf{J}\left(\Psi\right) - \widehat{\mathbf{J}}\right)' \mathbf{\Sigma}^{-1} \left(\mathbf{J}\left(\Psi\right) - \widehat{\mathbf{J}}\right) \tag{21}$$

where $\mathbf{\Sigma}$ is a diagonal matrix containing the estimated variances of the empirical impulse responses. We also compute an estimator $\hat{\mathbf{V}}$ for the asymptotic covariance matrix of $\widehat{\Psi}$ as

$$\hat{\mathbf{V}} = \left(\frac{\partial \mathbf{J}}{\partial \Psi}\left(\widehat{\Psi}\right)' \Sigma^{-1} \frac{\partial \mathbf{J}}{\partial \Psi}\left(\widehat{\Psi}\right)\right)^{-1}$$

where $\frac{\partial \mathbf{J}}{\partial \Psi}\left(\widehat{\Psi}\right)$ is the Jacobian of $\mathbf{J}\left(\Psi\right)$ at $\widehat{\Psi}$. In our application, we include in $\mathbf{J}$ the responses of output Y_t, consumption C_t, investment I_t, hours N_t, the price level P_t, the nominal wage level W_t, and the nominal rate i_t, truncating impulse responses at the 16th quarter.

This procedure can be implemented for any set of empirical impulse responses to an identified monetary shock. We next present our baseline approach for obtaining these impulse responses.

4.2 The empirical response to a monetary policy shock

As our measure of monetary policy shocks, we use the shocks constructed by Romer and Romer (2004) on their original sample (1969m3–1996m12). This is one of the leading approaches to identify monetary shocks in the data, and it delivers impulse responses which generally align well, in timing and magnitude, both with the conventional view from central banks and with results from alternative methods (see e.g. Ramey 2016).

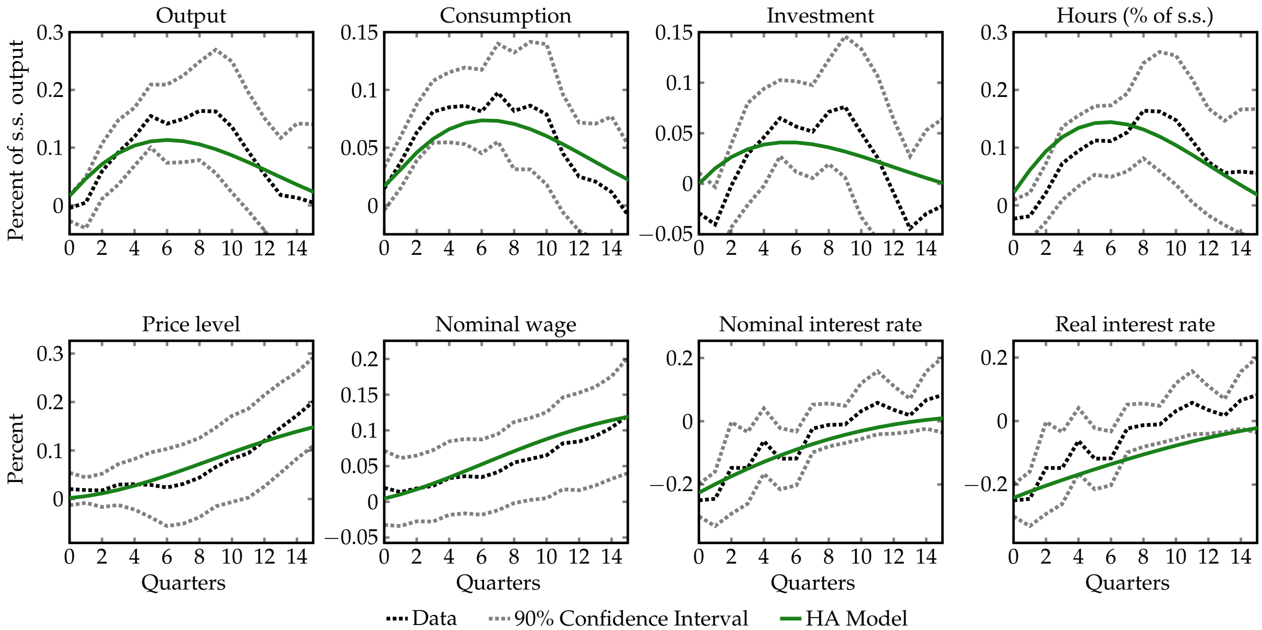
To obtain impulse responses, we use a Jordà (2005) projection. This standard procedure has the benefit of being able to recover the exact impulse responses in our model, provided that the Romer-Romer shocks represent iid innovations to the Taylor rule ϵ_t^m.[25]

We collect monthly data on eight standard macro time series: output, consumption, investment, hours, nominal prices, nominal wages, the nominal interest rate, and a measure of the real interest rate that uses ex-post inflation.[26] We then run a Jordà projection, which for a generic

[25] That is, given estimated parameters $\widehat{\Psi}$, if we simulate enough data from the model with aggregate shocks described in section 6, and run a Jordà projection on the innovations ϵ_t^m in the model-generated data, then we recover $\mathbf{J}\left(\widehat{\Psi}\right)$.

[26] Specifically, we use nominal GDP for Y_t, nominal PCE for C_t, and nominal investment for I_t, all deflated by the GDP deflator in order to preserve accounting identities. We also use the log of hours of all persons in nonfarm business sector for N_t, the log of the PCE deflator for P_t, the log of average hourly earnings of private production employees for W_t, the level of the Federal Funds rate for i_t, and the difference between that rate and one-month-ahead PCE inflation

Figure 3: Impulse response to a monetary policy shock vs. model fit



Note. This figure shows our estimated set of impulse responses to an identified Romer and Romer (2004) monetary policy shock (dashed black lines, with dashed gray confidence intervals). The solid green lines are the impulse responses implied by our estimated inattentive heterogeneous-agent model.

outcome such as Y_t reads

$$Y_{t+h} = J_h^Y \epsilon_t^m + \beta_h^Y X_t + \zeta_{t,h}^Y \tag{22}$$

separately for horizons $h = 1 \dots T$, up to $T = 48$ months, where ϵ_t^m is the Romer-Romer series and $\zeta_{t,h}^Y$ is a regression error term. To control for the potential endogeneity of ϵ_t^m in practice, we include in X_t the set of controls that Ramey (2016) uses in her specification for figure 2, panel B: lags of industrial production, unemployment, the consumer price index and a commodity price index. We compute the standard deviation of $\widehat{J_h^Y}$ using a Newey and West (1987) correction for the autocorrelation in $\zeta_{t,h}^Y$. For ease of interpretation, we normalize the impulse responses of output, consumption and investment to be in percentage points of output in the period before the shock. We finally aggregate the impulse responses $\left\{\widehat{J_h^Y}\right\}$ and their standard deviations to the quarterly level, and normalize them so that the impact fall in the nominal interest rate is equal to 25 bps. This delivers the empirical impulse response matrix $\widehat{\mathbf{J}}$ that we use in our estimation.

The dark dashed lines in figure 3 display our impulse responses, with lighter dashed lines indicating confidence intervals. Since actual inflation does not respond much on impact, the real interest rate, which is not targeted, falls by about as much as the nominal interest rate. As anticipated, output, consumption, investment and hours follow the "macro hump" pattern that is also documented in numerous alternative studies, with peak magnitudes that are also typical of those

for ex-post r_t. For series that are only available at the quarterly frequency, we interpolate to monthly frequency before running the regression.

found in other work, although here they occur after a somewhat longer delay. Prices and wages take time to respond, do not rise at all initially, and the magnitude of their eventual increase is small.[27]

4.3 Computational methodology

Our estimation procedure to match empirical impulse responses is very close to the one popularized by Christiano, Eichenbaum and Evans (2005): we simulate the model to first order in aggregate shocks, and find parameters that minimize the distance between the model's impulse responses and their empirical counterparts in the sense of equation (21). This requires simulating the model many times, one for each guess of the parameter vector Ψ. There are two features of our model that make this computation challenging. First, we have heterogeneous agents rather than a representative agent. Second, we have inattention—specifically, sticky expectations—rather than full-information rational expectations.

Our methodology for simulating the model can be separated into two parts. The first part deals with the simulation of a version of our model without sticky expectations, and draws on the tools developed in Auclert et al. (2019). The second handles sticky expectations, and can be generalized to other deviations from rational expectations. It is an important methodological contribution of this paper.

Simulation without sticky expectations. To solve the full-information rational expectations model, we use the method of Auclert et al. (2019), which solves for impulse responses as sequences following first-order, perfect-foresight (or "MIT") shocks.[28] Applying this methodology, we break down the model into "blocks", which take certain aggregate sequences as inputs and produce other aggregate sequences as outputs. For instance, the Taylor rule block has sequences for inflation π and the monetary shock ϵ^m as inputs, and the nominal interest rate i as its output. We depict all blocks in our model economy, and also provide a short summary of our method, in appendix D.2.

The key computational objects in this method are the Jacobians $\mathcal{J}$ of each block—the derivatives of its outputs with respect to its inputs. For example, one of the Jacobians of our household block is $J^{C,Z}$, a matrix containing $\partial C_t / \partial Z_s$, the model's iMPCs discussed in section 2. In Auclert et al. (2019), we provide methods to efficiently obtain all model Jacobians and combine them to obtain impulse responses. A major advantage of this approach for estimation is that most of this work does not need to be repeated across parameter draws. In particular, aside from the inattention parameter θ, none of the parameters in Ψ affect the Jacobians of the household block.

[27]For comparison, on average across the US estimates from the suite of models in Coenen et al. (2012), the peak output response to a 25bp shock to monetary policy is around 0.125%, while the cumulative inflation response after 4 years is about 0.2%. But our impulse responses are more delayed than theirs, since for them, the peak output effect is at 4 quarters and inflation starts to rise earlier.

[28]This merges Reiter (2009)'s approach of solving for equilibrium as a first-order linear system in aggregates with the MIT shock approach in Boppart, Krusell and Mitman (2018).

Therefore, in the rational expectations case $\theta = 0$, we can calculate these Jacobians a single time and reuse them on every parameter draw. This achieves a very large speed gain.

Including sticky expectations. We now introduce a method to deal with sticky expectations at almost no extra computational cost.

To understand this approach, it is useful first to consider a different problem: how to handle different permanent types of households, such as our types $g \in \mathcal{G}$ in (5). This is straightforward to handle using our sequence-space approach: we first calculate the Jacobians for each group individually using the methods in Auclert et al. (2019), then take their population-weighted average to obtain the Jacobians for the HA household sector as a whole.[29]

Our method to deal with sticky expectations proceeds similarly. At any point in time t, we partition households into different groups: those that have learned about the date-0 shock at different times $\tau \leq t$, and those that yet have to learn about it. We can then aggregate across this form of heterogeneity in the same way as with permanent types g: the aggregate response of all households is the weighted sum across household groups τ, weighted by their frequency. Since learning at date τ has probability $(1-\theta)\theta^\tau$ and is orthogonal to idiosyncratic shocks, the aggregate Jacobian relating output sequence o to input sequence i across all households is

$$\mathcal{J}^{o,i} = (1-\theta)\sum_{\tau=0}^{\infty}\theta^\tau \mathcal{J}^{o,i,\tau} \tag{23}$$

where $\mathcal{J}^{o,i,\tau}$ denotes the Jacobian for the group of households that learns about the shock to input sequence i at date τ.

To compute $\mathcal{J}^{o,i}$, we use the following insight. If $\tau \leq s$, the impulse response of a household learning at date τ about a date-s change in input i is the same as the impulse response of a household learning at date 0 about a date-$(s-\tau)$ change in i, shifted by τ periods. Moreover, if $\tau > s$, then τ is irrelevant since all households are aware that the shock has passed. Appendix D.3 shows that, combining this insight with equation (23), we obtain the following recursion relating $\mathcal{J}^{o,i}_{t,s}$ to the *full-information* Jacobian $\mathcal{J}^{o,i,FI}_{t,s} \equiv \mathcal{J}^{o,i,0}_{t,s}$:

$$\mathcal{J}^{o,i}_{t,s} = \begin{cases} \theta\mathcal{J}^{o,i}_{t-1,s-1} + (1-\theta)\mathcal{J}^{o,i,FI}_{t,s} & t>0,\ s>0 \\ \mathcal{J}^{o,i,FI}_{t,s} & s=0 \\ (1-\theta)\mathcal{J}^{o,i,FI}_{t,s} & t=0,\ s>0 \end{cases} \tag{24}$$

Given $\mathcal{J}^{o,i,FI}_{t,s}$, we can use (24) to build $\mathcal{J}^{o,i}$ with a simple operation for each entry. But $\mathcal{J}^{o,i,FI}_{t,s}$ is exactly what the the methods from Auclert et al. (2019) allow us to calculate efficiently. In short, starting with the full-information Jacobian, we can calculate a modified Jacobian with sticky

[29] By contrast, this would be much more costly with a state-space solution method. In the Reiter method, for instance, the cost of the bottleneck Schur decomposition would grow with the state space by a factor of $|\mathcal{G}|^3$, whereas the cost of our method only scales with $|\mathcal{G}|$. For the same reason, our sticky-expectations friction would be intractable in the state space.

expectations at almost no additional cost. Appendix D.3 shows how to generalize this approach to other deviations from rational expectations.

Overall, then, our computational approach with sticky expectations is very similar to the approach without sticky expectations, with one additional step. Before beginning estimation, we compute Jacobians $\mathcal{J}^{o,i,FI}$ for the household block for all o, i, since this does not vary with any of the parameters we seek to estimate. Then, during estimation, for each draw of a new θ parameter we apply (24) to obtain $\mathcal{J}^{o,i}$.

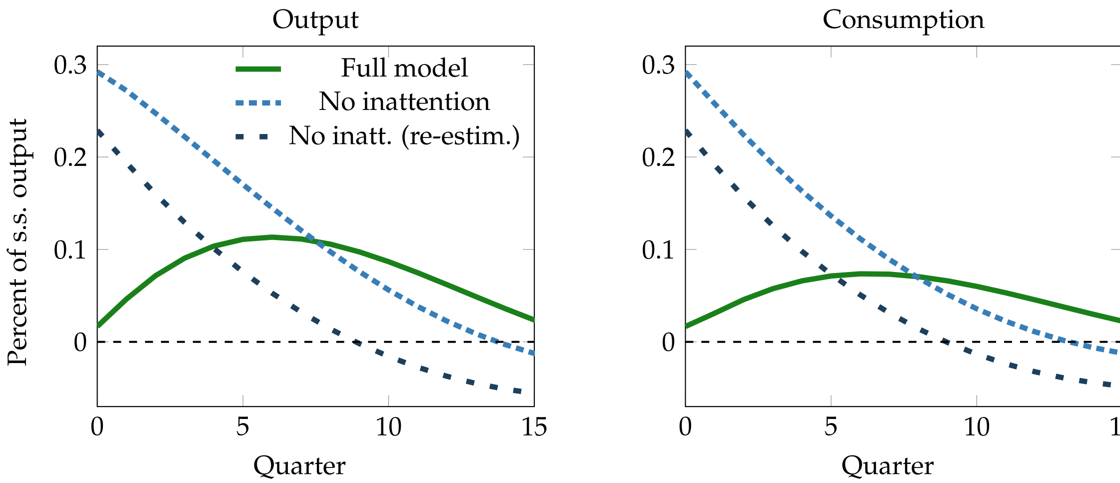
4.4 Model fit

The model-implied impulse responses at our estimated parameters are depicted in the solid green line of Figure 3. Overall, the model produces a good fit to the impulse responses. In order to compare this outcome to the fit of representative-agent models, in appendix D.4 we also estimate the RA-habit model, using the same procedure on the same set of impulse responses. Figure D.3 shows that this estimated RA model produces a similar set of impulse responses. Hence, the macro fit of our model is comparable to that achieved by typical medium-scale representative-agent models currently used for monetary policy analysis. Given the challenge that we posed in section 2, we view the achievement of such an equally good macro fit while simultaneously matching micro jumps as a clear success.

The right panel of Table 2 displays our parameter estimates. The Phillips curve parameters suggest a significant degree of price and wage stickiness. The estimated Calvo parameter of price stickiness is $\zeta_p = 0.93$, implying a 14-quarter average price duration, with the estimated Calvo parameter for wage stickiness just a little lower. This is more price and wage stickiness than in Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007), but it follows directly from our impulse responses: both prices and wages respond to the shock very slowly and with small magnitude. Indeed, our estimated RA model with habits implies very similar levels of ζ_p and ζ_w (see table D.1).

Our estimated inattention parameters show a fairly large degree of inattention for households, $\theta = 0.935$, implying an average duration of macro inattention of 15 quarters. This is required by our estimation procedure to match the delayed response of consumption, which peaks at nearly the two-year mark. (For a similar reason, the RA-habit model requires a habit parameter of $\gamma = 0.85$). This value may be compared to direct estimates from Coibion and Gorodnichenko (2012, 2015) and Coibion, Gorodnichenko and Kumar (2018) that speak to the quarterly rate of persistence of expectations for inflation and the output gap. For example, Coibion and Gorodnichenko (2012) estimate a quarterly degree of information rigidity for inflation around 0.8 for consumers and 0.86 for professional forecasters. These are below our estimate of θ—but note that, as discussed in section 2.3, these estimates provide a lower bound for the value of θ relevant for our model, since they miss the fraction of agents that adjust their expectations, but do not act on these adjusted expectations.

Figure 4: Impulse responses with and without inattention



Note. This figure shows the general equilibrium paths of output and consumption in our estimated HA model with different assumptions on inattention. The solid green uses our baseline estimates of household inattention. The dashed blue line is the impulse response when the inattention parameter is set to $\theta = 0$, holding all other parameters fixed at their estimated value in table 2. The dotted dark blue line reestimates the model parameters without inattention.

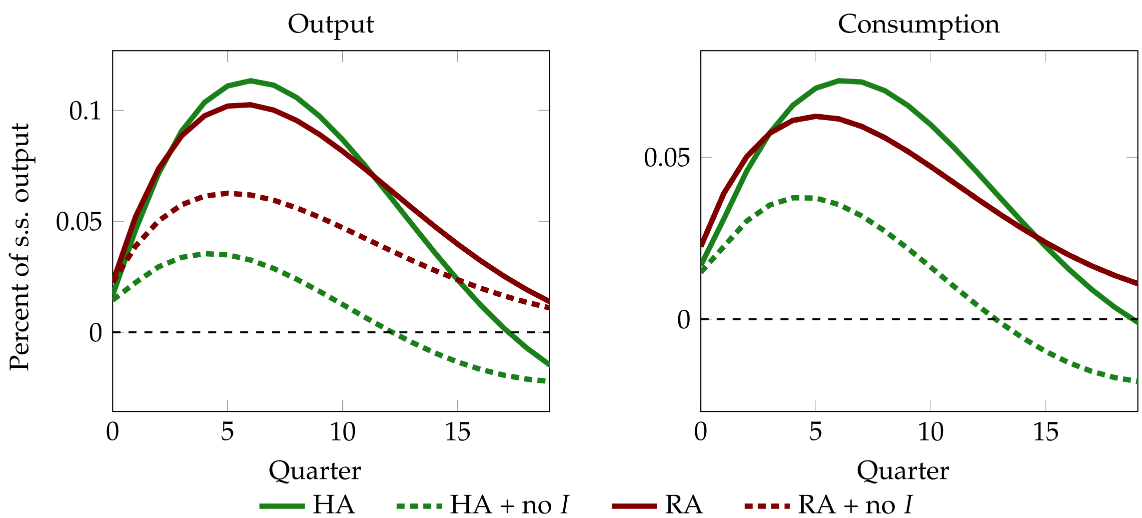
4.5 The importance of inattention

With our estimated model in hand, we can determine the extent to which household inattention is needed to match the impulse responses. In section 2, we argued that inattention was needed to generate hump-shaped consumption responses to persistent income shocks. But our investment adjustment cost specification implies that our model already features a general equilibrium source of humps in income, via investment. One might conjecture that this income hump-shape would translate directly into a consumption hump-shape, so that inattention is not needed after all.

This conjecture is incorrect, as we illustrate using two separate exercises. First, holding all parameters at their estimated values, we switch off inattention in the model, setting $\theta = 0$. The dashed light-blue line of figure 4 displays the resulting impulse responses of output and consumption, with the solid line displaying our baseline impulse responses for comparison. The consumption response has a counterfactual peak on impact. The reason is that, even though the response of investment is delayed, perfectly attentive households have enough liquidity to bring forward spending in anticipation of future changes in real interest rates and income.

Second, we reestimate the model to match the same empirical impulse responses, but this time constrain inattention to be switched off at $\theta = 0$. The resulting impulse response is displayed in the dotted dark blue line of figure 4. In this estimated model, there is barely a response of investment at all, and consumption again displays a counterfactual peak on impact. Here, the estimation procedure struggles to reconcile the large anticipatory consumption response in the model with the small impact responses of consumption, output, and hours in the data, and finds

Figure 5: Role of investment in the transmission mechanism



Note. This figure shows general equilibrium output and consumption in our estimated HA model (green) and an estimated representative-agent model with habits (red) in response to the same shocked real interest rate path, given different assumptions on the investment adjustment cost parameter ϕ. The solid lines are the baseline models, the dashed lines correspond to $\phi = \infty$.

the best fit when investment is almost infinitely frictional.

To sum up, in section 2 we introduced inattention as a natural route to generating hump-shapes in consumption in a model with high MPCs. In this section, we confirmed that inattention is both necessary and sufficient to generate humps in our model.[30]

5 Monetary policy transmission

Armed with our estimated model that matches both micro jumps and macro humps, we now examine the transmission mechanism of monetary policy.

5.1 Investment and the transmission mechanism

Our first main finding is that investment plays a central role in the transmission mechanism of monetary policy. To see this most cleanly, in figure 5 we shut off the investment response by setting the investment adjustment cost parameter to $\phi = \infty$ while assuming that monetary policy continues to implement the same shocked path for the real interest rate.[31] This reduces the peak

[30] The necessity of both inattention and high MPCs reflects, in part, a complementarity between the two forces, documented in Angeletos and Huo (2019) and Farhi and Werning (2019).

[31] Since monetary policy ultimately works through the real interest rate, this makes our counterfactual easier to interpret. An alternative is to hold constant the perturbation to the intercept of the Taylor rule, but since the Taylor rule endogenously changes the real interest rate in response to macro outcomes, the results—although similar—are harder to interpret in terms of monetary transmission.

output response by 69%, and the cumulative output response over 20 quarters by 84%.

These declines are far larger than investment's direct share of the output response, which at peak is slightly less than 40% in our baseline results. This is because shutting off investment dramatically weakens the effect of monetary policy on *consumption* as well: in figure 5, without investment the peak consumption response falls by 49%, and the cumulative consumption response over 20 quarters by 72%. This reflects a powerful investment-consumption feedback that is unique to models with high MPCs: as investment rises, the additional output demand leads to a rise in labor income—much of which is spent on consumption by high-MPC households, leading to even more output demand.

In short, investment is indirectly responsible for much of the consumption response to monetary policy. The exact shape of this indirect effect is influenced by the structure of our model. First, since inattentive households do not immediately realize that their labor incomes will rise—but spend the income once it arrives in their liquid accounts—the full effect of the consumption-investment feedback is backloaded, with sizable effects in figure 5 well after the $t = 6$ peak in output. Second, although elastic investment leads to a smaller capital price response to monetary policy, the relevance for consumption is limited because capital is held in the illiquid account. Neither of these features, however, is needed for the effect itself: in appendix A, we show that investment-consumption complementarity plays a qualitatively very similar role in a simplified HANK model without either inattention or illiquidity.[32]

On the other hand, this mechanism is not present in a representative-agent model. The red lines in figure 5 show the same impulse responses for our estimated RA model with habits. Shutting off investment does lead to a decline in the output impulse—by 38%, both at peak and cumulatively over 20 quarters—but this is solely due to the direct role of investment in output demand. Indeed, in this model there is no change in the consumption impulse at all, since it is pinned down by the path of real interest rates and intertemporal substitution via the Euler equation.

Indeed, a true heterogeneous-agent model turns out to be essential for the effects in figure 5. Even in a two-agent economy with habits—calibrated with a 20% share of hand-to-mouth households to match the same quarterly MPC, and estimated to match the same impulse responses—consumption is barely affected when investment is shut off (see appendix E.1). This is because a two-agent model, even if it has the same average MPC, cannot match the *intertemporal* MPCs shown in Figure 2. It therefore misses important intertemporal feedbacks from investment to consumption.

The centrality of investment to the monetary transmission mechanism has important practical implications. For instance, if investment is less elastic than usual—say, if there is an overhang of previously accumulated capital as in Rognlie, Shleifer and Simsek (2018)—any given change in interest rates will be much less effective, and policymakers may need to be more aggressive.

Given the importance of investment, one might wonder whether a richer model of investment,

[32]Our model would feature an additional source of amplification if we introduced unequal income incidence, making income inequality countercyclical (see Bilbiie, Känzig and Surico 2019).

beyond the standard investment adjustment cost specification, would give different results. For transmission through investment to consumption, however, the essential part of the model is the household side, with its high iMPCs. This is where we have concentrated most of the complexity of the model. Conditional on the investment response to monetary policy, which we match to the data, our amplification mechanism should be equally powerful, no matter what underlying forces are driving investment. In particular, it should persist with a more complex model of investment—for instance, a model with firm heterogeneity as in Ottonello and Winberry (2018), or a model where monetary policy affects investment via an interest coverage channel as in Greenwald (2019)—as long as that model, too, is estimated to match the same magnitude and hump-shape of the aggregate investment response.

5.2 Consumption in the transmission mechanism

So far, we have seen that investment is an important source of the consumption response to monetary policy. To further understand the determinants of consumption, we now decompose the aggregate consumption response into direct and indirect components, similar to the decomposition in Kaplan, Moll and Violante (2018). Since aggregate consumption can be written as a function $C_t = \mathcal{C}_t\left(\{r_s^\ell, r_s^a, Z_s\}\right)$ of liquid rates, illiquid rates, and aggregate after-tax labor incomes, we can decompose the aggregate consumption response into three terms, a direct and two indirect ones,

$$dC_t = \underbrace{\sum_s \frac{\partial \mathcal{C}_t}{\partial r_s^\ell} dr_s^\ell}_{\text{direct}} + \underbrace{\sum_s \frac{\partial \mathcal{C}_t}{\partial r_s^a} dr_s^a + \sum_s \frac{\partial \mathcal{C}_t}{\partial Z_s} dZ_s}_{\text{indirect}} \tag{25}$$

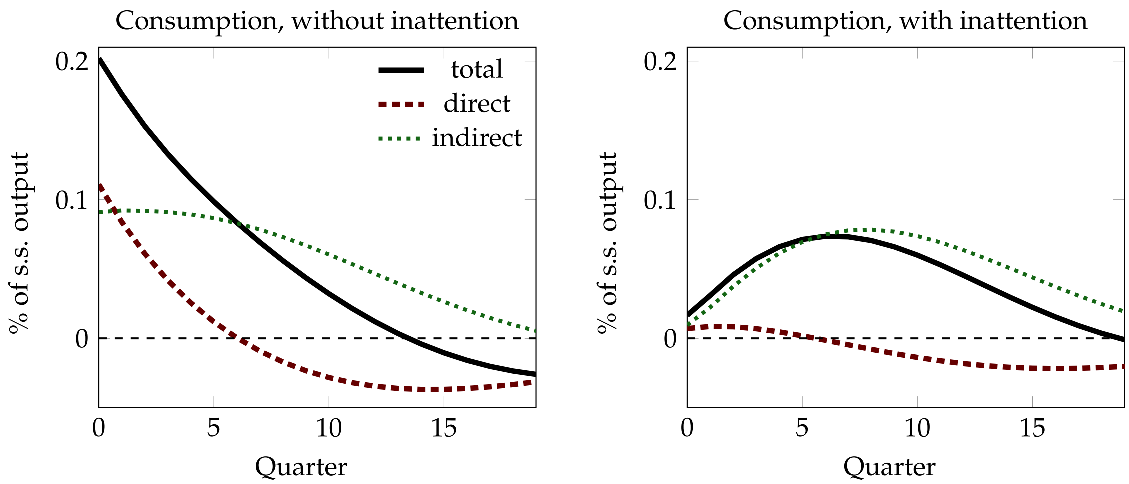
The direct term captures the direct effect of lower liquid real interest rates on consumption, through income and substitution effects. The indirect terms capture the effects of the return on the illiquid account and greater labor incomes.

Figure 6 shows the breakdown into direct and indirect effects, when the paths of $\{r_s^\ell, r_s^a, Z_s\}$ are those observed in response to the monetary shock in figure 3. The right panel does this for our inattentive HA model. Strikingly, almost the entire consumption response is accounted for by indirect effects. The direct effect is almost absent—a finding reminiscent of, but even more extreme than, Kaplan, Moll and Violante (2018).

To understand the near-absence of the direct effect, we switch off inattention in the left panel of figure 6, still feeding in the same equilibrium paths for $\{r_s^\ell, r_s^a, Z_s\}$. The direct effect is alive and well in this plot, accounting for a large share of the consumption response in the first few quarters. In fact, in large part it is the direct effect that causes the full-attention impulse response of consumption to be downward-sloping.

This shows that inattention not only generates a hump-shaped consumption response overall (see also figure 4), but does so by dampening the direct effect in figure 6, as well as by shifting the indirect consumption response into future periods. The reason for this is that the direct effect mainly reflects intertemporal substitution—an activity that relies on changes in future expected

Figure 6: Decomposition of consumption



Note. These figures decompose the consumption response of our estimated HA model, with and without inattention, into direct and indirect effects as in Kaplan, Moll and Violante (2018), see (25).

interest rates and is therefore heavily dampened in a model with inattention.

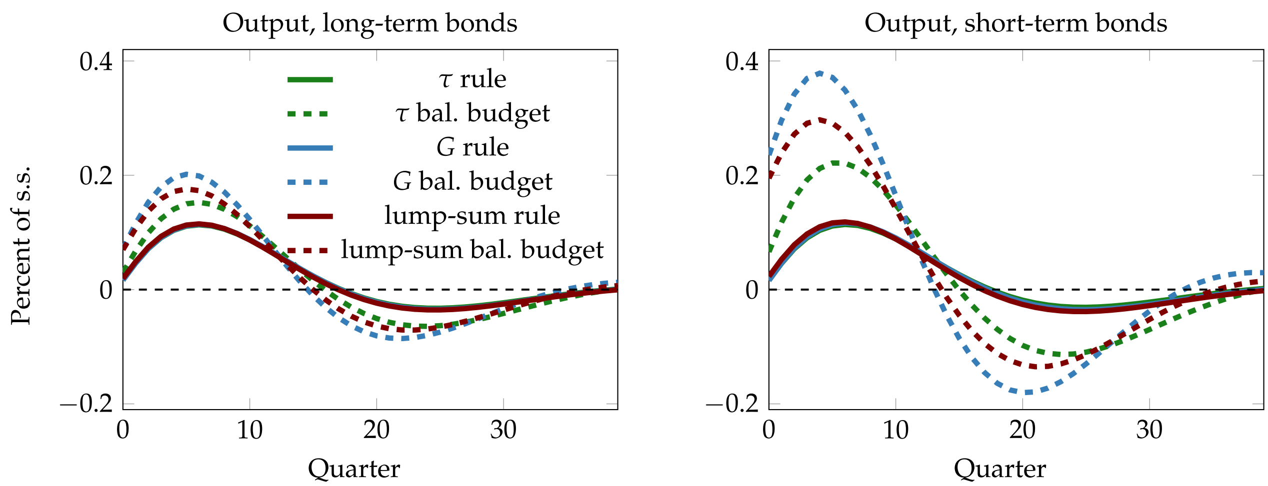
This dampening is necessary for the model to be consistent with the data. Otherwise, the combination of direct and indirect effects on impact will be far too large. To understand why this is inevitable in any model consistent with micro consumption data, in appendix E.2 we perform a simple experiment: we suppose that households' aggregate before-tax labor income is given by $(1-\alpha)Y_t^{data}$, where Y_t^{data} is the empirical impulse response of output to the monetary shock, and then feed in this labor income shock—*and no other shocks*—to the full-attention household sector. We find that the resulting consumption impulse is large, so that there is no room for intertemporal substitution to add to consumption in the first few quarters. We conclude that some friction must be dampening the overall consumption response on impact. In our model, this friction is inattention.

5.3 Fiscal policy in the transmission mechanism

We have uncovered investment as an important "indirect effect" of monetary policy on consumption. The prior HANK literature (e.g. Kaplan, Moll and Violante 2018) has also stressed indirect effects, but with greater emphasis on fiscal policy.

Is fiscal policy an important channel of monetary transmission in our model? Interestingly, no. This is due to two features that make the model a better quantitative fit for fiscal policy. First, rather than having only short-term debt, we allow for long-term debt and calibrate its duration to the average in the US time series, five years. This limits the effect of monetary policy shocks on the government budget, since only a small fraction of debt needs to be rolled over each quarter at

Figure 7: The role of fiscal policy for monetary transmission



Note. This figure shows the effects of the shocked real interest rate path under alternative fiscal rules. The colors correspond to different ways a given primary deficit is reduced: "τ rule" corresponds to a rule for labor tax rates; "G rule" means government spending adjusts; "lump-sum rule" means lump-sum transfers adjust.

a new interest rate.[33] Second, in our baseline we assume that the government follows a fiscal rule that gradually adjusts debt back to a steady-state target, rather than requiring a balanced budget each quarter.[34]

To gauge the quantitative importance of fiscal policy, in figure 7 we explore sensitivity to several different specifications. In the left panel, we assume alternative fiscal rules governing the adjustment of taxes or spending to changes in the government budget. The solid green, red, and blue lines maintain our baseline assumption that the adjustment takes place over the long term, but in addition to our baseline rule (20) where adjustment occurs through the tax rate τ, they alternatively vary government spending G or a lump-sum transfer T instead.[35] These show almost identical impulse responses. If, however, we instead assume that the adjustment happens on impact because the government must follow a balanced-budget rule, then government spending adjustment has the largest impact because it directly feeds into aggregate output demand, followed by lump-sum transfers and finally the tax rate. This is same ranking as in Kaplan, Moll and Violante (2018).

The right panel shows that the effects from the balanced-budget rule are much larger if we

[33]To our knowledge, this is the first quantitative HANK model to feature long-term debt. For prior contributions to the HANK literature with long-term debt, see Auclert (2019) and Nuño and Thomas (2019).

[34]A balanced-budget rule is, for instance, the baseline specification in both McKay, Nakamura and Steinsson (2016) and Kaplan, Moll and Violante (2018). Our empirically calibrated fiscal rule is closer to one of the alternative specifications in Kaplan, Moll and Violante (2018), which assumes that fiscal shocks are absorbed by debt and very slowly repaid.

[35]To be consistent, we define $\Delta_t \equiv \psi q^{ss} \frac{B_{t-1} - B^{ss}}{Y^{ss}} w_t N_t$ as the total size of the fiscal adjustment induced by our rule (20) and set $G_t - G^{ss} = \Delta_t$ and $T_t = \Delta_t$ to arrive at the G and T rules respectively.

counterfactually set $\delta = 0$—i.e. we assume that all debt is short-term—since then the monetary shock immediately affects the cost of rolling over the entire debt. Under the baseline rule with gradual debt adjustment, however, it still makes minimal difference whether the government adjusts τ, G, or T, nor does it matter much whether the debt itself is short- or long-term.

In short, fiscal policy can play a major role if we assume both short-term debt *and* a balanced-budget rule—with a peak output effect that is twice as large as our baseline when the tax rate τ is used to balance the budget, and even larger effects with other instruments like G and T. But with more quantitatively realistic assumptions, the role of fiscal policy in monetary transmission diminishes dramatically, to the point where the instrument used for fiscal adjustment makes little difference.

5.4 What gets monetary transmission started?

The prior literature, as well as our discussion of consumption in section 5.2, has focused on *partial equilibrium* decompositions. These decompositions tell us the proximate influences on an outcome like consumption: what fraction of the aggregate consumption response to a monetary shock is driven directly by households' response to interest rates, or indirectly through households' response to labor income or dividends.

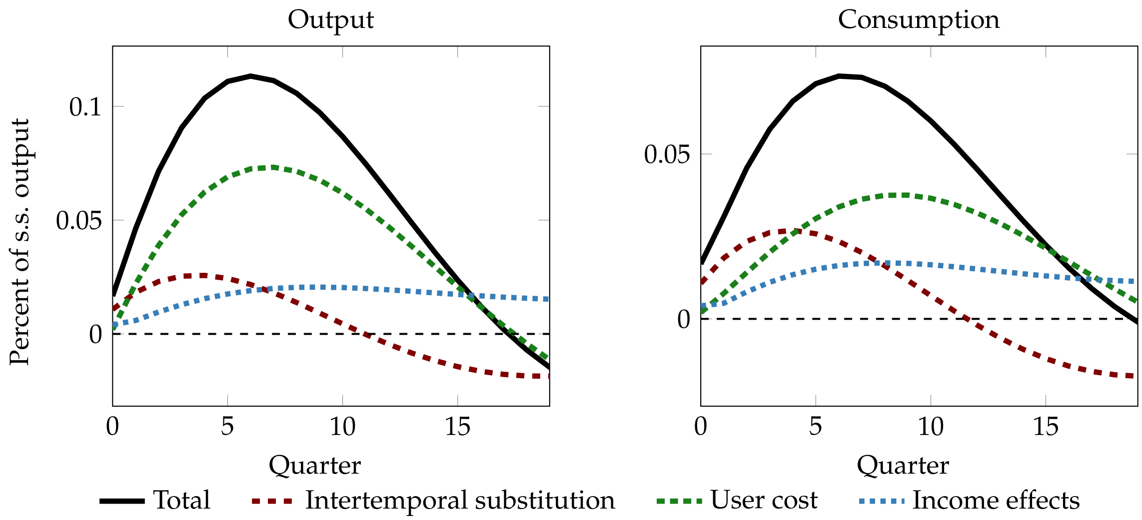
Although useful, these decompositions say nothing about how monetary policy caused variables like labor income to move in the first place. This is more difficult, since it is an inherently general equilibrium question. It is essential, however, if we want a complete picture of monetary transmission. For instance, even if households' direct response to interest rates explains only a small partial equilibrium share of consumption, it is possible that this response triggers general equilibrium forces that drive a much larger share of consumption through indirect effects. Alternatively, the indirect effects might come from some other source. With only a partial equilibrium decomposition, there is no way to know.

We therefore introduce a *general equilibrium* decomposition. The idea is that a shock to the real interest rate path $\{r_t\}$ influences equilibrium separately by entering into various equations. We can take some subset of these equations and ask what happens if $\{r_t\}$ is only perturbed in them, solving for general equilibrium. Splitting the set of equations where r_t enters into disjoint subsets, and repeating this process for each subset, by linearity the first-order outcomes will sum to the actual general equilibrium, giving us a decomposition that reveals the importance of each role for r_t in starting the monetary transmission mechanism.

Concretely, in this application we split the equations for the model in section 3 three ways. First, we assume that a real interest rate $r_{t+1}^{\ell,sub} = r_t^{sub} - \xi$ appears in the Euler equation (46), which governs intertemporal substitution. Second, we assume that a real interest rate r_t^{user} appears in (14) and (15), which is where real interest rates influence the investment decision by changing the user cost. Third, we assume that a real interest rate r_t^{inc} appears in all other equations, which govern asset pricing and income flows.[36]

[36]Since solving for general equilibrium only makes sense when the model is internally consistent, it is important that

Figure 8: General equilibrium decomposition: which effect gets monetary transmission started?



Note. This figure shows a general equilibrium decomposition of output and consumption into one of three effects of r.

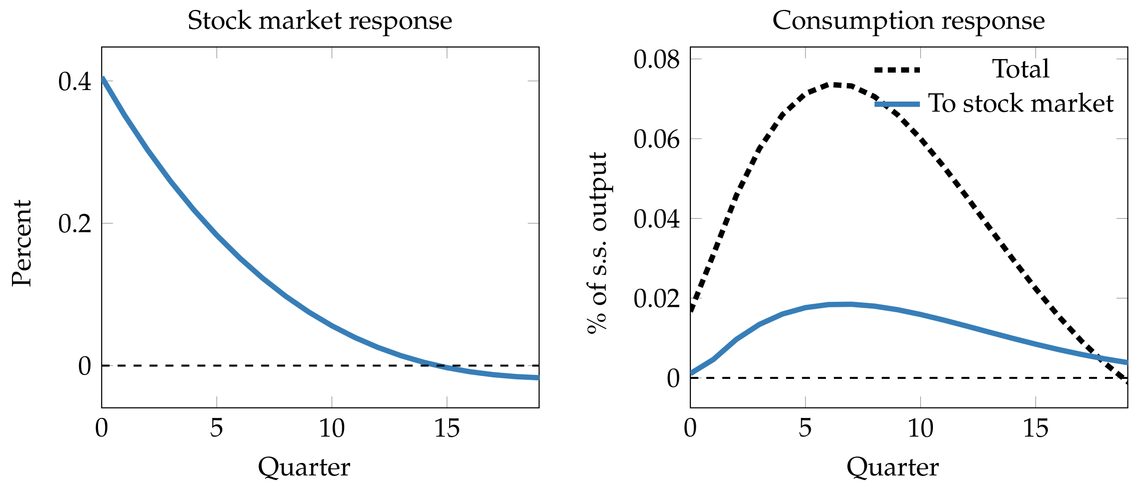
In figure 8, we apply the perturbed real interest rate from our estimated monetary shock to the paths of $\{r_t^{sub}\}$, $\{r_t^{user}\}$, and $\{r_t^{inc}\}$ separately, and calculate general equilibrium output and consumption for each, getting three paths that sum to the aggregate effect. By far the most important component of both output and consumption turns out to be the "user cost" effect from r_t^{user}. This effect answers the question: "if the only change is that capital firms perceive a different real interest rate path when making investment decisions, what happens in general equilibrium?" Remarkably, most monetary transmission appears to work through this one role of the real interest rate: about 65% of the output response, both at peak and cumulatively over 20 quarters. This clarifies the dominant role of investment from section 5.1: the importance of investment ultimately stems from the effect of interest rates on the user cost, rather than an endogenous response of investment to some other force.

On the other hand, the intertemporal substitution effect from r_t^{sub} is weak, becoming negative after 10 quarters. This effect answers the question: "if the only change is that households perceive a different real interest rate path when making intertemporal substitution decisions with their Euler equation, what happens in general equilibrium?" Notably, however, this *is* stronger than the direct effect in the right panel of figure 6, because it includes GE multiplier effects that amplify the original intertemporal substitution effect, while excluding the negative income effect from declining interest in the liquid account.

Finally, the income effect from r_t^{inc} is moderate but persistent. It reflects a negative contribution to consumption from declining liquid interest income, as well as the more backloaded positive contribution from the illiquid account rising in value, and the even more backloaded positive

the r_t corresponding to any flow of funds be perturbed simultaneously for both the sender and recipient. We achieve this by using r_t^{inc} for all such flows.

Figure 9: The stock market response to monetary policy and its effect on consumption



Note. The left figure shows the stock price response to the monetary policy shock in the model (in %), as well as the induced consumption response (in % of steady-state output).

contribution from lower taxes—as well as GE multiplier effects amplifying these forces.

5.5 Asset prices in the transmission mechanism

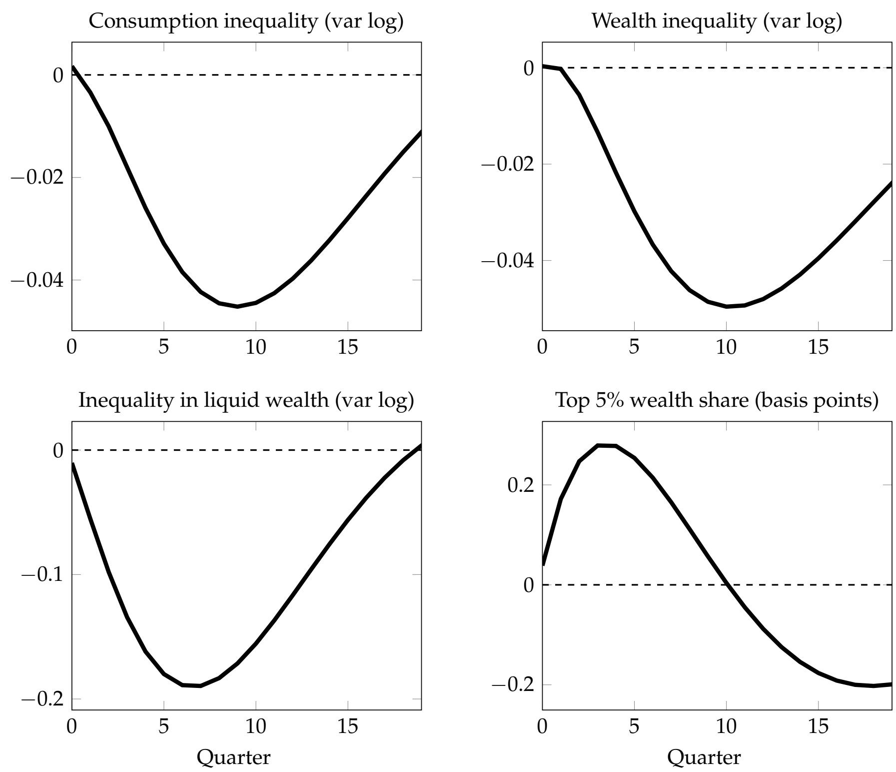
It is well-known that accommodative monetary policy raises stock prices (see Bernanke and Kuttner 2005 among many others). Figure 9 shows that our model is fully consistent with this observation. This is noteworthy because, as discussed in Kaplan and Violante (2018), many leading models in the literature do not even get the stock market response to go in the right direction.

In our model, the stock market rises both due to less discounting as well as an increase in future dividends. The left panel of figure 9 plots the overall response of stock prices to our accommodative monetary policy shock. The response is peaked on impact, consistent with the evidence in the literature. The magnitude, however, is somewhat smaller than what has been found in the data by Bernanke and Kuttner (2005) and the follow-up literature, reflecting the fact that our model does not feature effects from monetary policy on risk premia (see Kekre and Lenel 2019 for a HANK model that does).

The right panel of figure 9 plots the consumption response in the model, as well as the share of it that is explained by wealth effects from asset market returns.[37] The blue line shows that households respond to the stock market increase in our model, but only moderately and with a delay. The overall magnitude of this effect lies between the marginal propensity to consume out of stock market gains of 5 cents documented in di Maggio, Kermani and Majlesi (2018) and that of 2.8 cents documented in the cross-section of locations by Chodorow-Reich, Nenov and Simsek

[37] In the notation of section 5.2, this corresponds to $\sum_s \frac{\partial \mathcal{C}_t}{\partial r_s^a} dr_s^a$.

Figure 10: The distributional effects of monetary policy



(2019). Moreover, the delayed response is consistent with that documented in the latter paper.

5.6 Monetary policy and inequality

Figure 10 spells out the implications of our model for the effects of accommodative monetary policy shocks on measures of consumption and wealth inequality. Overall consumption and wealth inequality—as measured by the variance of the log—fall after accommodative monetary policy shocks, consistent with some empirical evidence (e.g. Coibion, Gorodnichenko, Kueng and Silvia 2017). This is largely driven by the positive response of labor incomes to monetary easing.

However, while overall wealth inequality falls, the wealth share of the top 5% household group actually rises slightly. This is due to the fact that this group holds a disproportionate fraction of the stock market, which increases in value.[38] Hence, the distributional effects of monetary policy implied by our model are subtle: rising labor incomes reduce inequality, but revaluation effects at the top accentuate inequality.

[38]Note that the share of equity within illiquid portfolios is the same for all household groups. Assuming that wealthier households hold larger equity shares would accentuate our results for the top 5%.

6 Investment and the business cycle

Thus far we have kept our focus on the transmission mechanism of monetary policy, uncovering a channel by which the response of investment stimulates consumption. Comovement of investment and consumption, however, is hardly limited to monetary policy shocks. In fact, a long literature following Barro and King (1984) has recognized that this comovement is at the heart of business cycle fluctuations.

This motivates us to investigate the investment-consumption comovement in our model in a business cycle context. To do so, we expand the model to allow for seven shocks and estimate it using Bayesian methods on historical business cycle data, as in Smets and Wouters (2007).[39]

6.1 Shocks

As we strive to stay as close to the previous literature as possible, almost all shocks are the same as in Smets and Wouters (2007). Also, almost all shocks are simply shocks to parameters that were assumed constant in the model description in section 3. In particular, we allow for 3 supply shocks (TFP shocks Θ_t, price markup shocks $\epsilon_{p,t}$, wage markup shocks $\epsilon_{w,t}$) and 4 demand shocks. Two of those demand shocks are exactly as in Smets and Wouters (2007) (monetary policy shocks ϵ_t^m, government spending shocks G_t). The two others in Smets and Wouters (2007) are risk premium shocks and investment-specific technology shocks. Risk premium shocks hit both investment and consumption in the model. Since we would like to focus on endogenous rather than exogenous comovement, we assume that there are two separate risk premium shocks, one for consumption and one for investment. We model the one for consumption as a discount factor shock. That is, the discount factor β_g of group g between periods t and $t+1$ is replaced by $\beta_g \exp\{\epsilon_t^C\}$ where ϵ_t^C is a mean-zero shock process. This specification implies that the only place ϵ_t^C enters is households' intertemporal optimality condition.

We proceed similarly for firms. The risk premium shock for investment is assumed to enter the optimality condition of mutual funds. In particular, we assume that mutual funds equalize the expected return on stocks with risk-premium adjusted return on bonds, replacing (9) with

$$\frac{\mathbb{E}_t\left[p_{jt+1} + D_{jt+1}\right]}{p_{jt}} = 1 + r_t + \epsilon_t^I$$

The shock ϵ_t^I is a mean-zero shock process and is our "investment shock": by revaluing capital, it directly affects the incentive to invest. We opt for this shock, rather than an investment-specific technology shock, because the literature shows that the latter has limited role in the business cycle when disciplined by data on investment prices (Justiniano, Primiceri and Tambalotti 2011).[40]

[39] A recent literature attempts to incorporate cross-sectional information to perform Bayesian estimation of heterogeneous-agent models (see Chang, Chen and Schorfheide 2018 and Plagborg-Møller and Liu 2019). We do not pursue this strategy here.

[40] An appealing implication of our investment shock specification is that it causes stock prices to be procyclical, as in Christiano, Motto and Rostagno (2014).

Table 3: Priors and posteriors

Supply shock		Prior distribution	Posterior Mode	Posterior std. dev	Demand shock		Prior distribution	Posterior Mode	Posterior std. dev
TFP Θ_t	s.d.	Invgamma$(0.1, 2)$	0.332	(0.016)	Mon. policy ϵ_t^m	s.d.	Invgamma$(0.1, 2)$	0.215	(0.010)
	AR	Beta$(0.5, 0.2)$	0.952	(0.016)		AR	Beta$(0.5, 0.2)$	0.171	(0.054)
w markup $\epsilon_{w,t}$	s.d.	Invgamma$(0.1, 2)$	0.362	(0.026)	G shock G_t	s.d.	Invgamma$(0.1, 2)$	0.319	(0.015)
	AR	Beta$(0.5, 0.2)$	0.916	(0.031)		AR	Beta$(0.5, 0.2)$	0.957	(0.015)
	MA	Beta$(0.5, 0.2)$	0.863	(0.048)	C shock ϵ_t^C	s.d.	Invgamma$(0.1, 2)$	3.369	(0.312)
p markup $\epsilon_{p,t}$	s.d.	Invgamma$(0.1, 2)$	0.248	(0.015)		AR	Beta$(0.5, 0.2)$	0.871	(0.018)
	AR	Beta$(0.5, 0.2)$	0.495	(0.155)	I shock ϵ_t^I	s.d.	Invgamma$(0.1, 2)$	4.452	(1.026)
	MA	Beta$(0.5, 0.2)$	0.791	(0.030)		AR	Beta$(0.5, 0.2)$	0.793	(0.029)

Note. For an ARMA(1,1) process of the form $x_{t+1} - \rho x_t = \epsilon_{t+1} - \theta \epsilon_t$, "AR" refers to ρ, "MA" refers to θ. To be comparable with Smets and Wouters (2007) we scale the markup shocks such that $\epsilon_{w,t}, \epsilon_{p,t}$ appear with a coefficient of 1 in the Phillips curves (18) and (13).

6.2 Estimation

We specify the full model, with its seven shocks, following Smets and Wouters (2007). In particular, we assume that all but the two markup shocks follow AR(1) processes around their deterministic steady-state values (which are zero for $\epsilon_t^m, \epsilon_t^C, \epsilon_t^I$). The two markup shocks are assumed to follow ARMA(1,1) processes. We depart from Smets and Wouters (2007) by not assuming a correlation between TFP and government spending shocks. Instead, all shocks are orthogonal.

We estimate our model to match the evolution of seven data series from 1966 Q1 (the beginning of the sample period in Smets and Wouters 2007) until 2018 Q4. The seven data series are real output Y_t, real consumption C_t, real investment I_t, real wages w_t, hours N_t (all in logs and linearly detrended), as well as inflation π_t and nominal interest rates i_t (both demeaned).[41]

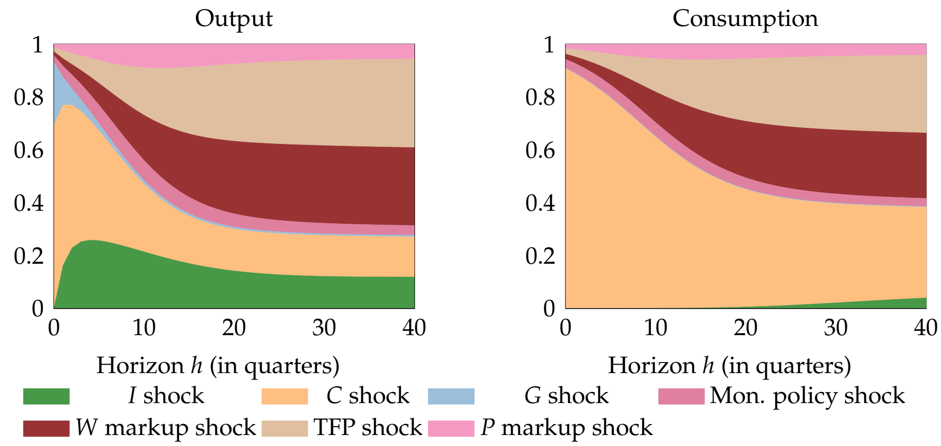
In order to maintain continuity within the paper and ensure that our results are not driven by unexplored features of impulse responses, we adopt the following strategy for estimation. We keep all parameters as in section 4—both the steady state and the parameters estimated from our monetary policy shocks. Instead, we estimate shock process parameters: all shocks' AR and MA parameters, as well as their standard deviations. For these parameters, we use the same priors as in Smets and Wouters (2007), shown in table 3.

It goes without saying that our inattentive heterogenous-agent model is a complicated one to estimate. We start with our insights from section 4.3 on how to efficiently simulate heterogeneous-agent models with information frictions, which allow us to quickly compute impulse responses to each shock. We then apply the approach from Auclert et al. (2019) to directly calculate the log-likelihood from these impulses.[42] We use this methodology to compute the posterior mode,

[41]Consistent with the data used in section 4.2, Y_t, C_t, I_t are nominal series normalized by the GDP deflator to preserve accounting identities; the nominal interest rate is the quarterly average fed funds rate; inflation is the annualized quarter-on-quarter change in the PCE deflator; employment is hours of all persons in the nonfarm business sector; the real wage is average hourly earnings of private production employees deflated by the GDP deflator.

[42]A forerunner of this approach in the sticky-information literature is Mankiw and Reis (2007), who also solve their

Figure 11: Forecast error variance decomposition for the RA model



Note. This figure decomposes $\text{Var}_{t-1}(Y_{t+h})$ and $\text{Var}_{t-1}(C_{t+h})$ into contributions of the seven shocks of the model.

as well as its standard errors from a Laplace approximation around the posterior mode. Starting from the prior means as initial guesses, the estimation takes around 120 seconds on a personal laptop. The sixteen estimated parameters are also shown in table 3.

For comparison, we also perform the exact same estimation, with the same shocks, in the representative-agent model with habits introduced in section D.4. Priors and posteriors for that model can be found in appendix F.1.

6.3 The role of investment for business cycles

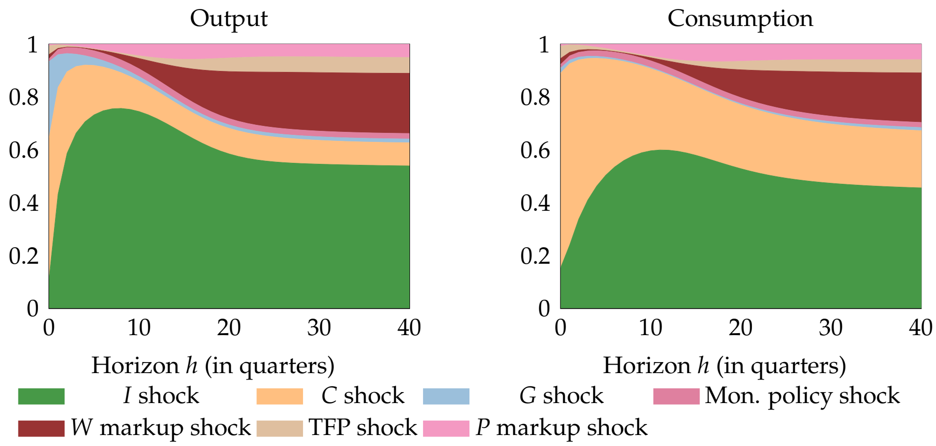
We use the estimated RA and HA models to explore the drivers of business cycles. To do so, we focus on forecast error variance decompositions of output and consumption. We construct these by additively decomposing $\text{Var}_{t-1}(Y_{t+h})$ and $\text{Var}_{t-1}(C_{t+h})$ at various horizons $h \geq 0$ into contributions from the seven orthogonal shocks.

The decompositions for the RA model are shown in figure 11. The most important drivers of output variation at business cycle horizons—6 to 32 quarters—in our RA model are supply shocks (TFP and markup shocks), explaining over half the forecast error variance. Supply shocks also matter for consumption, explaining a bit less than half the variation, with the rest being mostly explained by discount factor shocks ("C shocks"). Investment shocks matter somewhat for output (~15%), but do not explain consumption variation.

The decompositions for the HA model in figure 12 are strikingly different, even though the only change is in the model's household side. Investment shocks rise to prominence in the estimated HA model. Specifically, at business cycle horizons, they explain about 65% of the output variation and about 55% of the consumption variation. Discount factor shocks remain important

model in sequence space and calculate the log-likelihood directly from that representation.

Figure 12: Forecast error variance decomposition for the HA model



Note. This figure decomposes $\text{Var}_{t-1}(Y_{t+h})$ and $\text{Var}_{t-1}(C_{t+h})$ into contributions of the seven shocks of the model.

for consumption, but less so than in the RA model. Supply shocks become far less important.

Why does the estimated HA model favor investment shocks so much more than the RA model? As anticipated, the comovement between consumption and investment plays a crucial role. Figure 13 decomposes the forecast error *co*variance $\text{Cov}_t(Y_{t+h}, C_{t+h})$ between investment and consumption at various horizons $h \geq 0$.

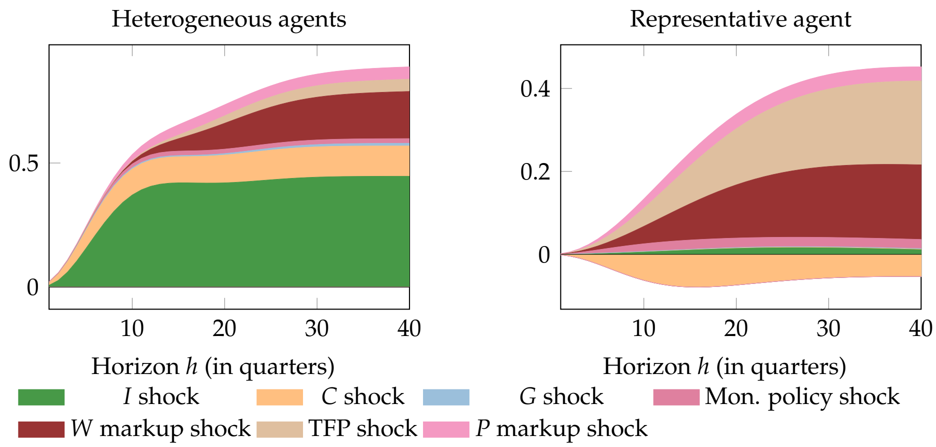
In our HA model, investment shocks drive most of the covariance—a striking consequence of the model's investment-consumption complementarity when MPCs are high. By contrast, investment shocks explain almost none of the covariance of consumption and investment in the RA model, in line with the Barro and King (1984) puzzle. Instead, supply shocks are needed.

The reason for this contrast is clear from the impulse responses to an investment shock in figure 14. In the HA model, consumption and investment move together and peak at similar times, whereas in the RA model there is a tiny and (due to habit formation) highly backloaded consumption response.[43]

The idea that investment shocks can help explain the business cycle comovement between consumption and investment in HA models is a robust insight of our paper. It also applies to other versions of the RA model that already attribute a large share of output fluctuations to investment shocks (see e.g. Justiniano, Primiceri and Tambalotti 2010). Typically, even these models do not find much role for investment shocks in moving consumption, at least without long lags. Presumably, this would be different in a heterogeneous-agent version of these models, for the same reason as here.

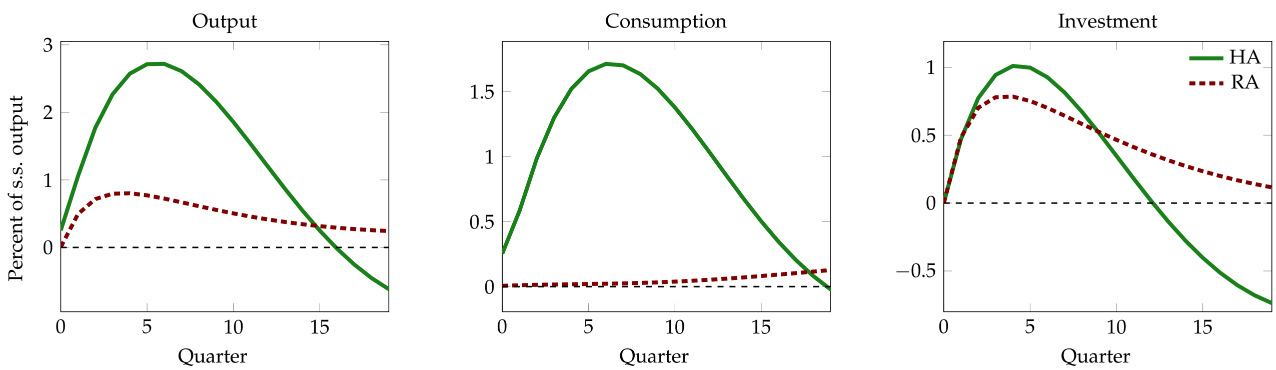
[43]Figures F.1—F.7 in appendix F.2 display all other impulse responses in the RA and HA models. Appendix F.3 shows historical decompositions of output and consumption.

Figure 13: Decomposition of forecast error covariance between consumption and investment



Note. This figure decomposes $\text{Cov}_{t-1}(Y_{t+h}, C_{t+h})$ into contributions of the seven shocks in the two models. Figure normalized by $sd(Y_t) \cdot sd(C_t)$.

Figure 14: Impulse response to a 1-standard deviation investment shock



Note. Impulse responses are to a one-time negative shock to risk premia ϵ_t^I, using the estimated standard deviation and persistence.

7 Conclusion

Our paper brings the current vintage of heterogeneous-agent New Keynesian models in line with the often sluggish behavior of macroeconomic aggregates. We do this by introducing sticky expectations and showing that the combination of heterogeneity and informational rigidity can be implemented very efficiently. Our model implies a very different view of the transmission mechanism of monetary policy and of the sources of business cycles—one in which investment plays a central role.

References

Altig, David, Lawrence J. Christiano, Martin Eichenbaum, and Jesper Lindé, "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics*, April 2011, *14* (2), 225–247.

Alvarez, Fernando, Luigi Guiso, and Francesco Lippi, "Durable Consumption and Asset Management with Transaction and Observation Costs," *American Economic Review*, May 2012, *102* (5), 2272–2300.

Alves, Felipe, Greg Kaplan, Benjamin Moll, and Gianluca Violante, "A Further Look at the Propagation of Monetary Policy Shocks in HANK," *Manuscript*, 2019.

Angeletos, George-Marios and Chen Lian, "Chapter 14 - Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2, Elsevier, January 2016, pp. 1065–1240.

_ and _ , "Forward Guidance without Common Knowledge," *American Economic Review*, September 2018, *108* (9), 2477–2512.

_ and Zhen Huo, "Myopia and Anchoring," *Manuscript*, September 2019.

Auclert, Adrien, "Monetary Policy and the Redistribution Channel," *American Economic Review*, June 2019, *109* (6), 2333–2367.

_ and Matthew Rognlie, "A Note on Multipliers in NK Models with GHH Preferences," *Manuscript*, August 2017.

_ and _ , "Inequality and Aggregate Demand," Working Paper 24280, National Bureau of Economic Research, February 2018.

_ , Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," Working Paper 26123, National Bureau of Economic Research, July 2019.

_ , Matthew Rognlie, and Ludwig Straub, "The Intertemporal Keynesian Cross," Working Paper 25020, National Bureau of Economic Research September 2018.

Barro, Robert J. and Robert G. King, "Time-Separable Preferences and Intertemporal-Substitution Models of Business Cycles," *Quarterly Journal of Economics*, November 1984, *99* (4), 817–839.

Bayer, Christian, Benjamin Born, and Ralph Luetticke, "Shocks and Frictions in US Business Cycles with Heterogeneous Agents," *Manuscript*, September 2019.

_ , Ralph Luetticke, Lien Pham-Dao, and Volker Tjaden, "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," *Econometrica*, 2019, *87* (1), 255–290.

Bernanke, Ben S. and Kenneth N. Kuttner, "What Explains the Stock Market's Reaction to Federal Reserve Policy?," *Journal of Finance*, June 2005, *60* (3), 1221–1257.

Bianchi, Francesco and Leonardo Melosi, "Escaping the Great Recession," *American Economic Review*, April 2017, *107* (4), 1030–1058.

Bilbiie, Florin O., "Limited Asset Markets Participation, Monetary Policy and (inverted) Aggregate Demand Logic," *Journal of Economic Theory*, May 2008, *140* (1), 162–196.

_ , "Monetary Policy and Heterogeneity: An Analytical Framework," *Manuscript*, August 2019.

_ **, Diego Känzig, and Paolo Surico**, "Capital, Income Inequality, and Consumption: the Missing Link," November 2019.

Boppart, Timo, Per Krusell, and Kurt Mitman, "Exploiting MIT Shocks in Heterogeneous-Agent Economies: The Impulse Response as a Numerical Derivative," *Journal of Economic Dynamics and Control*, April 2018, *89*, 68–92.

Bricker, Jesse, Lisa J. Dettling, Alice Henriques, Joanne W. Hsu, Kevin B. Moore, John Sabelhaus, and Richard A. Windle, "Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, September 2014, *100* (4).

Carroll, Christopher, Edmund Crawley, Jiri Slacalek, Kiichi Tokuoka, and Matthew N White, "Sticky Expectations and Consumption Dynamics," *Manuscript*, July 2018.

Challe, Edouard, Julien Matheron, Xavier Ragot, and Juan F. Rubio-Ramirez, "Precautionary Saving and Aggregate Demand," *Quantitative Economics*, July 2017, *8* (2), 435–478.

Chang, Minsu, Xiaohong Chen, and Frank Schorfheide, "Heterogeneity and Aggregate Fluctuations," *Manuscript*, October 2018.

Chodorow-Reich, Gabriel, Plamen T. Nenov, and Alp Simsek, "Stock Market Wealth and the Real Economy: A Local Labor Market Approach," Working Paper 25959, National Bureau of Economic Research, June 2019.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, February 2005, *113* (1), 1–45.

_ **, Roberto Motto, and Massimo Rostagno**, "Risk Shocks," *American Economic Review*, January 2014, *104* (1), 27–65.

Cloyne, James, Clodomiro Ferreira, and Paolo Surico, "Monetary Policy when Households have Debt: New Evidence on the Transmission Mechanism," *The Review of Economic Studies*, January 2020, *87* (1), 102–129.

Coenen, Günter, Christopher J Erceg, Charles Freedman, Davide Furceri, Michael Kumhof, René Lalonde, Douglas Laxton, Jesper Lindé, Annabelle Mourougane, Dirk Muir, Susanna Mursula, Carlos de Resende, John Roberts, Werner Roeger, Stephen Snudden, Mathias Trabandt, and Jan in't Veld, "Effects of Fiscal Stimulus in Structural Models," *American Economic Journal: Macroeconomics*, January 2012, *3* (4), 22–68.

Coibion, Olivier and Yuriy Gorodnichenko, "What Can Survey Forecasts Tell Us about Information Rigidities?," *Journal of Political Economy*, 2012, *120* (1), 116–159.

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

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

_ , _ , Lorenz Kueng, and John Silvia, "Innocent Bystanders? Monetary Policy and Inequality," *Journal of Monetary Economics*, June 2017, *88*, 70–89.

Davig, Troy and Eric M. Leeper, "Monetary–Fiscal Policy Interactions and Fiscal Stimulus," *European Economic Review*, February 2011, *55* (2), 211–227.

di Maggio, Marco, Amir Kermani, and Kaveh Majlesi, "Stock Market Returns and Consumption," Working Paper 24262, National Bureau of Economic Research, January 2018.

Erceg, Christopher J., Dale W. Henderson, and Andrew T. Levin, "Optimal Monetary Policy with Staggered Wage and Price Contracts," *Journal of Monetary Economics*, October 2000, *46* (2), 281–313.

Fagereng, Andreas, Martin B. Holm, and Gisle J. Natvik, "MPC Heterogeneity and Household Balance Sheets," *Manuscript*, October 2018.

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

Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, Keith Kuester, and Juan Rubio-Ramírez, "Fiscal Volatility Shocks and Economic Activity," *American Economic Review*, November 2015, *105* (11), 3352–84.

Fuhrer, Jeffrey C., "Habit Formation in Consumption and Its Implications for Monetary-Policy Models," *American Economic Review*, June 2000, *90* (3), 367–390.

Furlanetto, Francesco and Martin Seneca, "Investment Shocks and Consumption," *European Economic Review*, 2014, *66*, 111–126.

Gabaix, Xavier, "A Behavioral New Keynesian Model," *Mimeo*, May 2016.

— **and David Laibson**, “The 6D Bias and the Equity-Premium Puzzle,” *NBER Macroeconomics Annual*, January 2001, *16*, 257–312.

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

Gornemann, Nils, Keith Kuester, and Makoto Nakajima, “Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy,” *Manuscript*, April 2016.

Greenwald, Daniel L, “Firm Debt Covenants and the Macroeconomy: The Interest Coverage Channel,” *Manuscript*, July 2019.

Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson, “Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements,” *International Journal of Central Banking*, 2005.

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman, “The Fiscal Multiplier,” Working Paper 25571, National Bureau of Economic Research, February 2019.

— **,** — **, and** — , “Monetary Policy in Incomplete Markets Models: Theory and Evidence,” *Work in progress*, 2019.

Havranek, Tomas, Marek Rusnak, and Anna Sokolova, “Habit Formation in Consumption: A Meta-Analysis,” *European Economic Review*, June 2017, *95*, 142–167.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante, “Optimal Tax Progressivity: An Analytical Framework,” *The Quarterly Journal of Economics*, 2017, *132* (4), 1693–1754.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles, “Household Expenditure and the Income Tax Rebates of 2001,” *American Economic Review*, December 2006, *96* (5), 1589–1610.

Jordà, Òscar, “Estimation and Inference of Impulse Responses by Local Projections,” *The American Economic Review*, March 2005, *95* (1), 161–182.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti, “Investment Shocks and Business Cycles,” *Journal of Monetary Economics*, March 2010, *57* (2), 132–145.

— **,** — **, and** — , “Investment Shocks and the Relative Price of Investment,” *Review of Economic Dynamics*, 2011.

Kaplan, Greg and Giovanni L. Violante, “A Model of the Consumption Response to Fiscal Stimulus Payments,” *Econometrica*, July 2014, *82* (4), 1199–1239.

— **and** — , “Microeconomic Heterogeneity and Macroeconomic Shocks,” *Journal of Economic Perspectives*, August 2018, *32* (3), 167–194.

— **, Benjamin Moll, and Giovanni L. Violante**, "Monetary Policy According to HANK," *American Economic Review*, March 2018, *108* (3), 697–743.

Kekre, Rohan and Moritz Lenel, "Monetary Policy, Redistribution, and Risk Premia," 2019.

Kimball, Miles S., "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, November 1995, *27* (4), 1241–1277.

Klenow, Peter J. and Jonathan L. Willis, "Real Rigidities and Nominal Price Changes," *Economica*, July 2016, *83* (331), 443–472.

Leeper, Eric M., "Equilibria Under 'Active' and 'Passive' Monetary and Fiscal Policies," *Journal of Monetary Economics*, February 1991, *27* (1), 129–147.

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

— **and** — , "Business Cycle Dynamics under Rational Inattention," *The Review of Economic Studies*, October 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," *Quarterly Journal of Economics*, November 2002, *117* (4), 1295–1328.

— **and** — , "Sticky Information in General Equilibrium," *Journal of the European Economic Association*, May 2007, *5* (2-3), 603–613.

McKay, Alisdair and Ricardo Reis, "The Role of Automatic Stabilizers in the U.S. Business Cycle," *Econometrica*, January 2016, *84* (1), 141–194.

— **, Emi Nakamura, and Jón Steinsson**, "The Power of Forward Guidance Revisited," *American Economic Review*, October 2016, *106* (10), 3133–3158.

Misra, Kanishka and Paolo Surico, "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs," *American Economic Journal: Macroeconomics*, October 2014, *6* (4), 84–106.

Nakamura, Emi and Jón Steinsson, "High Frequency Identification of Monetary Non-Neutrality," *Manuscript, Columbia University*, December 2013.

Negro, Marco Del, Frank Schorfheide, Frank Smets, and Rafael Wouters, "On the Fit of New Keynesian Models," *Journal of Business & Economic Statistics*, April 2007, *25* (2), 123–143.

Newey, Whitney K. and Kenneth D. West, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, May 1987, *55* (3), 703–708.

Nuño, Galo and Carlos Thomas, "Optimal Monetary Policy with Heterogeneous Agents," *Manuscript, Banco de España*, December 2019.

Ottonello, Pablo and Thomas Winberry, "Financial Heterogeneity and the Investment Channel of Monetary Policy," Working Paper 24221, National Bureau of Economic Research, January 2018.

Plagborg-Møller, Mikkel and Laura Liu, "Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data," *Manuscript*, June 2019.

Ramey, Valerie A., "Chapter 2 - Macroeconomic Shocks and Their Propagation," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2, Elsevier, January 2016, pp. 71–162.

Reiter, Michael, "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, March 2009, *33* (3), 649–665.

Rognlie, Matthew, Andrei Shleifer, and Alp Simsek, "Investment Hangover and the Great Recession," *American Economic Journal: Macroeconomics*, April 2018, *10* (2), 113–53.

Romer, Christina D. and David H. Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, September 2004, *94* (4), 1055–1084.

Rotemberg, Julio and Michael Woodford, "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," *NBER*, January 1997, pp. 297–361.

Schmitt-Grohé, Stephanie and Martín Uribe, "Closing Small Open Economy Models," *Journal of International Economics*, October 2003, *61* (1), 163–185.

_ and _ , "Optimal Fiscal and Monetary Policy in a Medium-Scale Macroeconomic Model," *NBER Macroeconomics Annual*, January 2005, *20*, 383–425.

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

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, June 2007, *97* (3), 586–606.

Werning, Iván, "Incomplete Markets and Aggregate Demand," Working Paper 21448, National Bureau of Economic Research, August 2015.

Wong, Arlene, "Transmission of Monetary Policy to Consumption and Population Aging," *Manuscript*, January 2018.

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

Zorn, Peter, "Investment under Rational Inattention: Evidence from US Sectoral Data," February 2018.

Online Appendix for "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model"

A Investment in a canonical HANK model

In this section, we write down a canonical HANK model with investment.[44] We then show that, in such a model, investment acts a strong amplifier of monetary policy shocks. This confirms that our findings regarding the role of investment are due to the presence of heterogenous agents with high average MPCs, rather than some other feature of our model in the main text.

The model here is a sticky-wage, flexible-price HANK model with capital adjustment costs, as in Auclert and Rognlie (2018) and Auclert, Rognlie and Straub (2018). In Auclert and Rognlie (2017), we explain why this is a more natural starting point for a HANK model than the opposite assumption of flexible wages and sticky prices.[45] The model in the main text adds features that are necessary to obtain a good micro and macro fit: household inattention, sticky prices, government spending and debt, and indexation of both prices and wages; and it replaces capital with investment adjustment costs. But the core complementarity between investment and MPCs that we highlight here is robust to the addition of these additional features.

A.1 Model setup

Individual-level productivity states s follow a Markov process with transition matrix Π. Unions make all households work an equal number of hours N_t. There is no taxation, so take-home pay for a household in state s at time t is $w_t N_t e(s)$, where $e(s)$ is idiosyncratic productivity. Households can trade in one asset, a liquid deposit ℓ issued by financial intermediaries. The household problem is therefore

$$\begin{aligned} V_t(\ell, s) &= \max_{c,a'} u(c) + \beta \mathbb{E}_t \left[V_{t+1}(\ell', s') \right] \\ \text{s.t.} \quad & c + \ell' = \left(1 + r_t^{\ell}\right) \ell + w_t N_t e(s) \\ & \ell' \geq 0 \end{aligned}$$

Nominal wages are set by unions, subject to Calvo wage rigidity. The optimization problem of unions implies a standard Phillips curve for wages[46], which can be written to first order (see

[44] The model and main results in this appendix previously appeared in a June 2018 SED presentation, "Forward Guidance is More Powerful Than You Think," where we pointed out that the MPC-investment interaction, in the absence of other frictions like informational rigidities, could aggravate the forward guidance puzzle. Here we make the same point about amplification for standard AR(1) monetary policy shocks.

[45] This also avoids the (counterfactual) very high countercyclicality of profits in a flexible-wage, sticky-price model.

[46] This is included only for completeness: given the exogenous monetary policy for the real interest rate r_t and the lack of any other nominal rigidity, the slope of this Phillips curve is irrelevant for real equilibrium outcomes.

Auclert, Rognlie and Straub 2018) as

$$\pi_t^w = \kappa^w \int N_t \left(v'\left(N_t\right) - \frac{\epsilon - 1}{\epsilon} u'\left(c_{it}\right) \right) di + \beta \mathbb{E}_t \left[\pi_{t+1}^w\right]$$

A financial intermediary issues liquid deposits to households and invests them in firm shares. At the beginning of the period, the value of its outstanding deposits must be equal to the liquidation value of firm shares, i.e.

$$\left(1 + r_t^{\ell}\right) L_{t-1} = \left(p_t + D_t\right) v_{t-1}$$

At the end of the period, the value of newly-purchased shares must be equal to the value of newly issued deposits and reserves, i.e.

$$p_t v_t = L_t$$

We also allow the financial intermediary to invest in nominal reserves that pay a promised return of i_t and are in zero net supply. The financial intermediary maximizes the expected return to depositors $\mathbb{E}_t\left[r_{t+1}\right]$. The optimal portfolio choice of the financial intermediary results in the pricing equations

$$\mathbb{E}_t\left[1 + r_{t+1}^{\ell}\right] = \frac{\mathbb{E}_t\left[p_{t+1} + D_{t+1}\right]}{p_t} = \left(1 + i_t\right) \mathbb{E}_t\left[\frac{P_{t+1}}{P_t}\right] \equiv 1 + r_t$$

where we have defined r_t as the ex-ante real interest rate.

A representative final goods firm produces with technology

$$Y_t = \Theta K_t^{\alpha} N_t^{1-\alpha}$$

Prices are flexible, and firms have no monopoly power, so the real wage w_t and the rental rate of capital are respectively equal to

$$\begin{aligned} w_t &= \Theta\left(1-\alpha\right) K_t^{\alpha} N_t^{-\alpha} \\ r_t^K &= \Theta \alpha K_t^{\alpha-1} N_t^{1-\alpha} \end{aligned}$$

A capital firm owns the capital stock K_t and rents it to the representative final good producer. It faces quadratic adjustment costs to capital. In period t, it enters the period with capital stock K_t, invests I_t to obtain a capital next period of $K_{t+1} = \left(1-\delta\right) K_t + I_t$, and pays the adjustment cost, resulting in a dividend of

$$D_t = r_t^K K_t - I_t - \frac{\Psi}{2}\left(\frac{K_{t+1} - K_t}{K_t}\right)^2 K_t$$

where Ψ indexes the size of adjustment costs. The firm has a unit share outstanding, $v_t = 1$.

The capital firm chooses investment to maximize the sum of its dividend and its end-of period share price, $D_t + p_t$. Defining $Q_t = \frac{\partial p_t}{\partial K_{t+1}}$ as the responsiveness of the share price to the capital chosen by the firm, simple algebra shows that, given the asset pricing equations above, this optimization problem involves the standard equations from Q theory:

$$\frac{I_t}{K_t} - \delta = \frac{1}{\Psi}\left(Q_t - 1\right) \tag{26}$$

and

$$Q_t = \frac{1}{1+r_t}\mathbb{E}_t\left[r^K_{t+1} - \frac{I_{t+1}}{K_{t+1}} - \frac{\Psi}{2}\left(\frac{I_{t+1}}{K_{t+1}} - \delta\right)^2 + \frac{K_{t+2}}{K_{t+1}}Q_{t+1}\right]$$

Finally, monetary policy sets the nominal interest rate i_t in order to achieve a target for the ex-ante real interest rate r_t.

In an equilibrium of this model, households, unions, financial intermediaries, final goods firms and capital firms optimize, and markets clear, so that:

$$\begin{aligned} C_t + I_t + \frac{\Psi}{2}\left(\frac{K_{t+1}-K_t}{K_t}\right)^2 K_t &= Y_t \\ L_t &= p_t \quad (= Q_t K_{t+1}) \end{aligned}$$

A.2 Investment as an amplifier of monetary policy

We now demonstrate that investment acts as a amplifier of monetary policy in this model. This force is unique to the presence of heterogeneous agents (HA), in the sense that the amplification we highlight is entirely absent with a representative agent (RA).

To show this, we study the effects of an AR(1) monetary policy shock $r_t = r + \epsilon_0 \rho^t$ under several different sets of assumptions. First, we compare the HA model described above to an RA model that is identical except that the household sector is replaced by a representative agent whose consumption is governed by the Euler equation:

$$u'\left(C_t\right) = \beta\left(1+r_t\right)u'\left(C_{t+1}\right) \tag{27}$$

Second, we compare a model with "no investment", where the capital adjustment cost in (26) is infinite ($\Psi = \infty$) and investment therefore cannot respond to a monetary shock, to a model where it is finite ($\Psi = \Psi_0$) and the investment response has empirically reasonable magnitude.

Calibration. As in the main text, we assume an elasticity of intertemporal substitution of $\sigma = 1$ and target a steady-state $r = 5\%$, for both the HA and RA models. We depart, however, by choosing zero depreciation $\delta = 0\%$. This makes our point especially stark, since then the HA and RA output responses are exactly identical under the "no investment" case where capital adjustment costs are infinite. (In subsection A.5, we show that this equivalence carries over numerically, though not analytically, to the $\delta = 5.3\%$ case from the main text.)

In the HA model, we use the same Markov process for $e(s)$ as in the main text, but rescale $\log e(s)$ so that the standard deviation of log income is the same as in the main text, despite the absence of permanent type heterogeneity here. Also as in the main text, we target steady-state liquidity holdings L such that the income-weighted average quarterly MPC matches the value in figure 2, of 0.194. Since steady-state liquidity equals capital in our simplified model, this implies a very low capital-output ratio of $\frac{K}{Y} = .304$ in annualized terms. We set $\alpha = (r+\delta)\frac{K}{Y}$ for both models, and calibrate the discount factor β^{HA} for the HA households to be consistent with L. The discount rate β^{RA} for the RA households is, by necessity for a steady state, $1/(1+r)$.

We calibrate the persistence of the monetary shock to be $\rho = 0.9$—close to the persistence of the section 4 monetary shock via inertia in the Taylor rule—so that it lasts an average of 10 quarters. We calibrate the size ϵ_0 of the shock such that the date-0 consumption response in the RA model is 1%. We calibrate the capital adjustment cost with investment $\Psi = \Psi_0$ so that the ratio of the investment to consumption response in the RA model is comparable to the ratio of the peak empirical impulses in section 4.2, at about 2/3.[47]

Main result. Figure A.1 presents our main result. The left part of the graph shows the response in the RA model, the right part shows the response in the HA model. The top row shows the calibration with no investment ($\Psi = \infty$), the bottom row the calibration with investment ($\Psi = \Psi_0$). In the RA model, turning on investment has no effect on the consumption response to a monetary policy shock. This follows from the Euler equation in (27): in equilibrium, the path of consumption is entirely dictated by monetary policy $\{r_t\}$.

Similarly, when investment is turned off, the addition of heterogeneous agents makes no difference to the impulse response to a monetary shock. This result is an instance of Werning (2015)'s neutrality result under log utility, and we prove it formally in section A.4. Intuitively, the general equilibrium effects of monetary policy shocks in this model affect asset prices and labor earnings in proportion to output in every period. (Log utility is needed for the former and Cobb-Douglas production for the latter.) Therefore, agents just scale their decisions, relative to the steady state, by $\frac{Y_t^{RA}}{Y_{ss}^{RA}}$, where Y_t^{RA} is the representative-agent allocation at date t, and the representative-agent allocation obtains in the aggregate.

By contrast, when investment is turned on, the effects on consumption are more than *double* the case without investment. The effects of investment are also a little larger than in the representative agent model, reflecting the fact that output and hence the marginal product of capital is higher at every point. This complementarity is the main result of this section: it is the simultaneous presence of heterogeneous agents *and* investment that generates an amplification that is absent if only one of the two is present. Since MPCs are much higher in the HA model, households consume out of labor income from the investment boom, creating a large multiplier on the investment response to monetary policy.

Interestingly, this mechanism leads to continued amplification even after capital firms start to draw down their investment: for instance, the consumption response at $t = 10$ in the HA-investment model is more than double the others, despite the investment response being slightly

[47] We handle equilibrium selection by assuming that, in all models we consider, the economy returns to its initial steady state in the long-run. This is sufficient to uniquely pin down equilibrium for these cases, and can be implemented (for instance) by monetary policy reverting to a Taylor rule at some far-out date t.

Figure A.1: Complementarity between investment and high MPCs

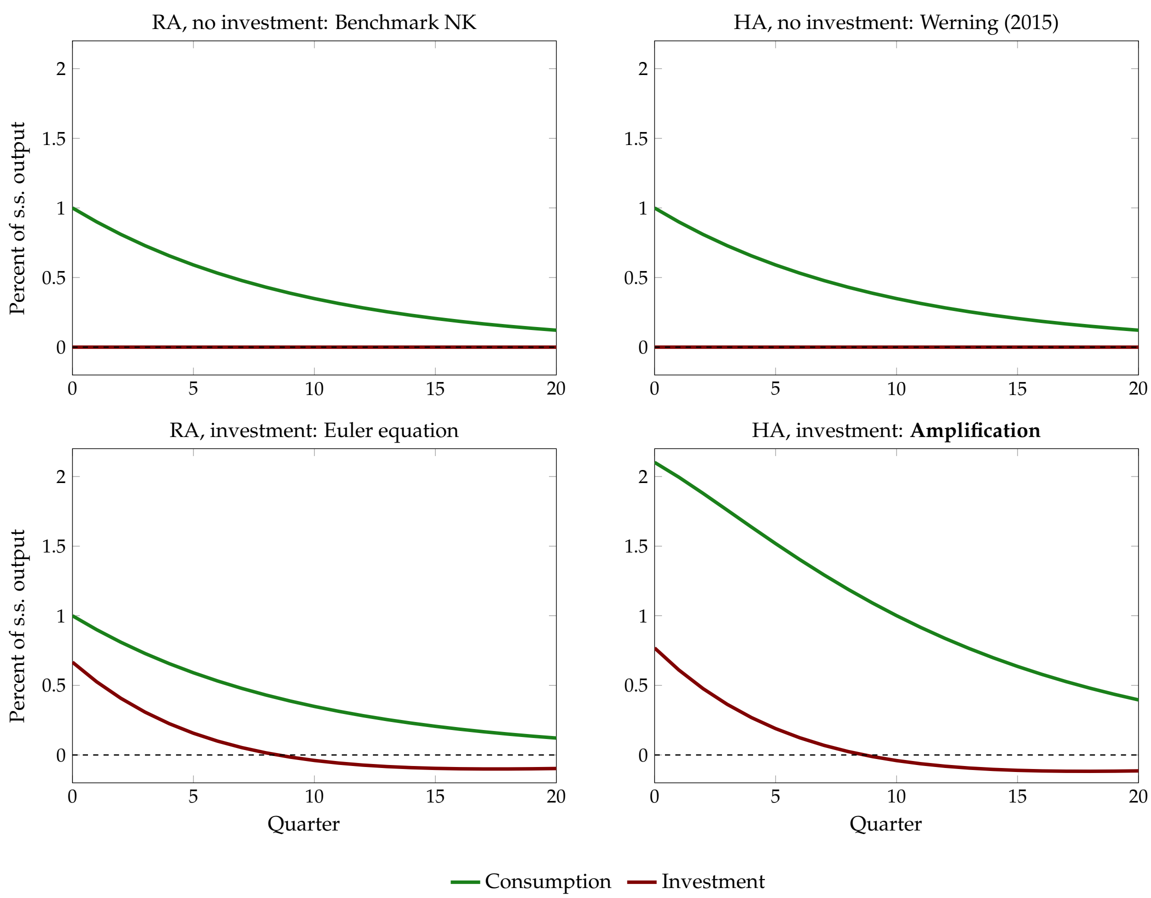
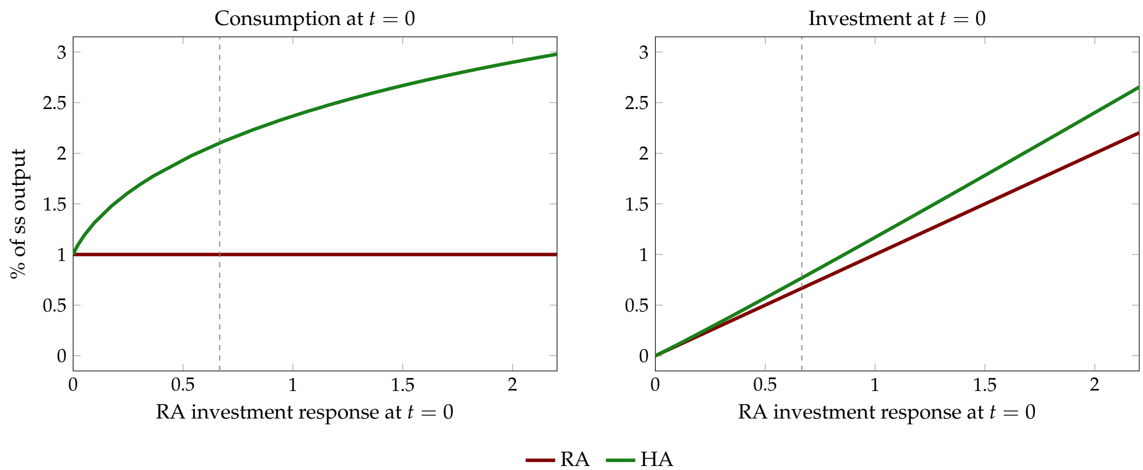


Figure A.2: Investment flexibility and amplification



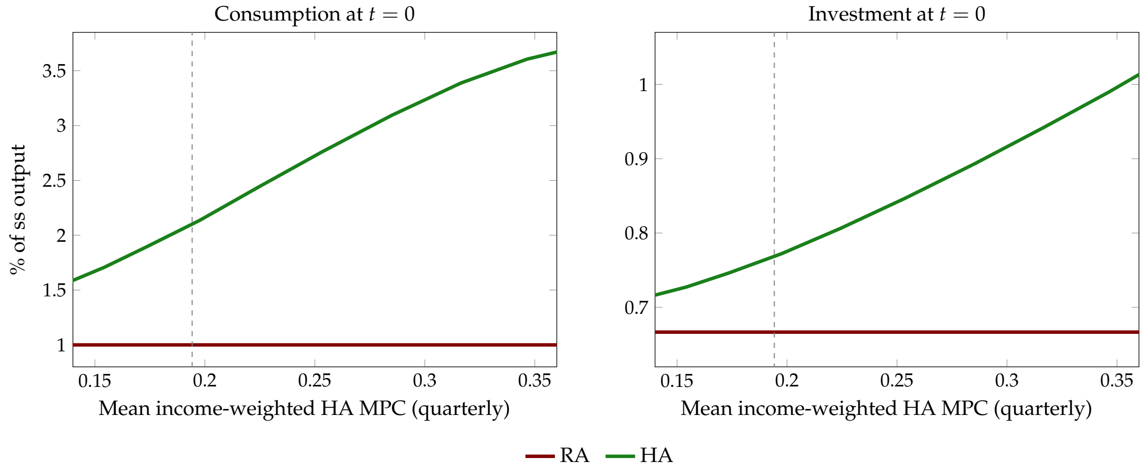
negative at that point. This is thanks to intertemporal demand spillovers, which result from the high iMPCs in the HA model.

Varying investment flexibility and MPCs. To illustrate this interaction further, in figure A.2 we vary the capital adjustment cost Ψ and study the effect on the date-0 responses of consumption C_0 and investment I_0. We quantify adjustment costs on the horizontal axis with the date-0 response of investment I_0 in the RA model (declining in Ψ); the gray dashed line at $2/3$ corresponds to our main calibration. Although the consumption response in the RA model is unaffected by investment flexibility, in the HA model the pattern is monotonic: more flexible investment implies a larger consumption response.

In figure A.3, we vary the level of labor income risk between 50% to 150% of its value in the HA model, while leaving all other aspects of the steady-state calibration, including total liquidity L, unchanged. This matters for our mechanism through the effect on iMPCs (see Auclert, Rognlie and Straub 2018), which here we summarize by plotting the average income-weighted MPC on the horizontal axis, with the gray dashed line at .194 corresponding to our main calibration. In the low-risk, low-MPC calibrations, amplification is small, since the household sector is closer to a representative agent; in high-risk, high-MPC calibrations, amplification becomes much larger. The high-MPC calibrations even have a slightly larger investment response, reflecting the elevated marginal product of capital from high consumption demand.

Together, figures A.2 and A.3 show the robustness of our mechanism: the interaction between investment and MPCs leads to amplification, which becomes stronger when we raise either investment flexibility or MPCs.

Figure A.3: MPCs and amplification



A.3 Understanding the mechanism: capital gains vs labor income

Thus far, we have emphasized transmission via labor income: when investment responds to monetary policy, the higher output demand leads to increased labor income, which raises consumption and output further via high MPCs.

There is another potentially important channel, however, through which the investment sector influences consumption: the return r_0^ℓ on assets between dates -1 and 0. In this model, $r_0^\ell = (p_0 + D_0)/p_{-1} - 1$, which includes the surprise revaluation effect on $p_0 = Q_0 K_0$ from the monetary shock. This effect shrinks when investment is made more flexible, since more capital investment leads to lower future rental rates on capital, offsetting the increase in valuation from lower real interest rates. A smaller return at date 0 makes households poorer, causing them to spend less.

To what extent does this limit amplification in the HA model? In figure A.4 we follow Auclert (2019) and Kaplan, Moll and Violante (2018) and decompose the first-order household consumption response into three sources, which together sum to the aggregate. First, there is the "direct" effect of changing ex-ante real interest rates $r_t = r_{t+1}^\ell$, which is unaffected by the general equilibrium investment response. Second, there is the "indirect" effect from higher labor income, which is the channel we have emphasized so far. Third, there is another "indirect" effect, from unexpected capital gains resulting in a higher date-0 return r_0^ℓ. We perform this decomposition both for the no-investment model $\Psi = \infty$ and the model with investment $\Psi = \Psi_0$.

The direct effect is the same in both models, but relatively muted. Instead, most of the increase in consumption is a response to rising labor income in general equilibrium. This response is much, much larger—by a factor of three—in the model with investment.

As expected, although capital gains contribute positively to consumption in both models, they play less of a role in the model with investment. Their influence is small enough in both cases, however, that this difference is barely visible in figure A.4.

Figure A.4: Decomposing the consumption response

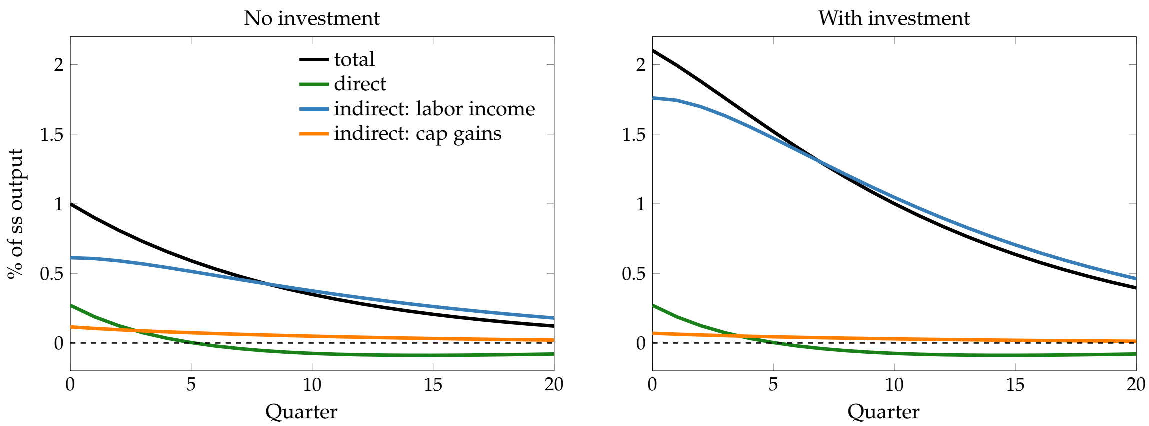
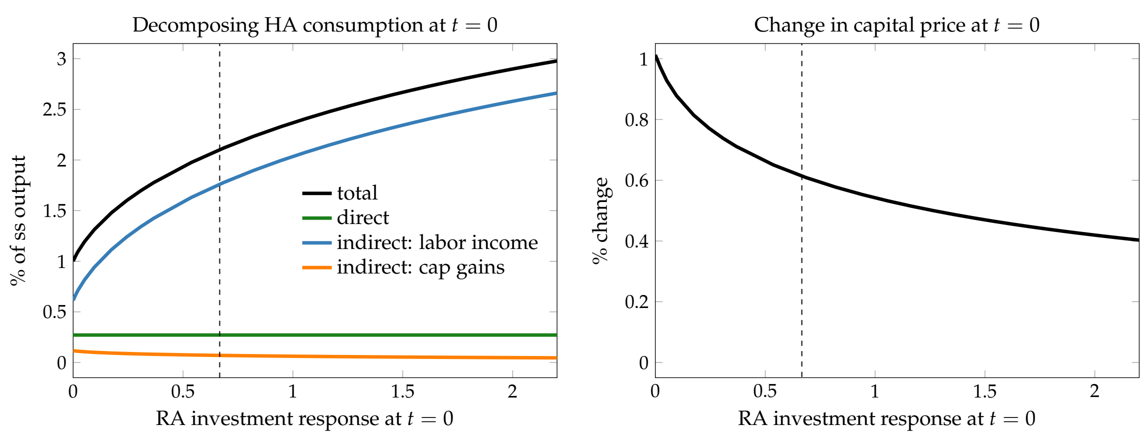


Figure A.5: Decomposition and capital gains by investment flexibility



Why are capital gains so unimportant in comparison to labor income? First, there is an important distributional difference: capital gains are earned by asset-holders, and MPCs are much lower for the asset-rich than for income-earners. For instance, in our calibration, the asset-weighted average quarterly MPC is 0.094, compared to the income-weighted quarterly MPC of 0.194 to which we calibrate. Second, although the decline in rental rates and corresponding rise in real wages causes redistribution from asset owners to income earners, the overall change in income is not zero-sum: indeed, this redistribution is swamped by the rise in aggregate income, most of which goes to labor.

Figure A.5 provides additional detail. The left panel decomposes the date-0 consumption response for the HA model, as a function of investment flexibility, that previously appeared in aggregate terms in figure A.2. We see that as investment flexibility rises, the indirect labor income effect steadily grows, while the indirect capital gains effect shrinks but with far smaller magnitude.

The right panel shows the impulse to date-0 return r_0^ℓ itself. The capital gains effect on consumption at $t = 0$ equals this, times the asset-weighted average MPC of 0.094. Although the change in r_0^ℓ is in relative terms quite dramatic—from a 1.0% to 0.6% increase as we go from the no-investment calibration to our main calibration—this becomes insignificant when multiplied by 0.094 to obtain the effect in the left panel.

Size and liquidity of the capital stock. Since all capital is liquid in this simple model, matching the average MPC from the main text implied a very low capital-output ratio, .304 in annual terms. The calibration in the main text instead has a value that is consistent with the macro data, 2.23.

Importantly, however, this larger capital stock in the main text is entirely held within the illiquid account. Although the larger stock suggests a larger role for capital gains, the illiquidity sharply limits this role: households receive the annuity value of their illiquid accounts into their liquid accounts as a flow, and this flow does not immediately change when the illiquid account gains value. As a result—as figure 5 makes clear—investment makes a strong positive contribution to output, despite its negative effect on the value of the illiquid account.

Our finding of a small contribution from capital gains is likely to be very robust to our calibration choice. If more capital is held in liquid accounts, then capital-holders will have low liquid MPCs. If, instead, more capital is held in illiquid accounts, capital prices' influence on consumption will be limited, since high liquid MPCs will then no longer be directly relevant. Furthermore, holding fixed the magnitude of the investment response, the effect of investment on total capital gains does not grow with the capital stock: if the stock is larger, the price moves proportionately less. Hence, even if a much larger capital stock was held in equally high-MPC liquid accounts, a realistic investment response would not exert any more downward pressure on consumption through the capital gains channel than in figure A.4.

Comparison to other papers. In ongoing and parallel work, Alves, Kaplan, Moll and Violante (2019) argue the opposite: that capital adjustment costs do not matter for aggregate consumption, since the capital gains and labor income effects offset. However, in their framework, another friction shapes capital accumulation: an illiquid account that cannot hold bonds, only capital and monopolists' equity. The properties of this friction may play an important role in their contrasting result.[48]

Bilbiie, Känzig and Surico (2019) argue that investment plays an important role in amplification primarily in conjunction with unequal cyclical incidence of labor income. By contrast, in both the main text and this section, we intentionally abstract away from unequal incidence to show that a combination of high MPCs and investment, on its own, is enough to deliver substantial amplification.

A.4 Neutrality proof with inelastic investment ($\Psi = \infty$)

Here we prove a neutrality result for the model with a fixed capital stock and no investment, explaining why the top left and right panels of figure underlying figure A.1 are identical. This is an instance of Werning (2015)'s finding for an EIS of 1.

First we need a lemma.

Lemma 1. *In perfect foresight equilibria of the RA model with $\sigma = 1$ and no investment ($\Psi = \infty$ and $\delta = 0$), $D_t = \alpha Y_t$ and $p_t = \frac{\alpha\beta^{RA}}{1-\beta^{RA}} Y_t$.*

Proof. Since $\Psi = \infty$, capital is always at its steady-state level K^{ss} and investment is always 0, so the dividend at t is $D_t = \alpha \frac{Y_t}{K^{ss}} K^{ss} = \alpha Y_t$, as desired. The asset price is

$$p_t = \frac{1}{1+r_t}\left(\alpha Y_{t+1} + p_{t+1}\right)$$

Iterating forward, we can write[49]

$$p_t = \sum_{s=1}^{\infty}\left(\prod_{u=0}^{s-1}\frac{1}{1+r_{t+u}}\right)\alpha Y_{t+s} \tag{28}$$

Note that iterating forward the Euler equation with $\sigma = 1$, we also have

$$C_t = (\beta^{RA})^{-s}\left(\prod_{u=0}^{s-1}\frac{1}{1+r_{t+u}}\right)C_{t+s}$$

which, using $Y_t = C_t$, we can substitute into (28) to get $p_t = \sum_{s=1}^{\infty}(\beta^{RA})^{-s}\alpha Y_t = \frac{\alpha\beta^{RA}}{1-\beta^{RA}} Y_t$, as desired. □

[48]They also perform a different experiment: they compare the model with *no* adjustment costs to the model *with* adjustment costs, whereas we compare the model *with* adjustment costs to the model with *infinite* adjustment costs (and more generally among different levels of the adjustment cost). We do not perform the first comparison, since the standard New Keynesian model explodes without adjustment costs in response to a real interest rate shock.

[49]This assumes that monetary policy $\{r_t\}$ does not permanently deviate so far from $r^{ss} > 0$ that the product does not converge to zero as $s \to \infty$.

Proposition 1. *Given any monetary policy $\{r_t\}$, a perfect foresight equilibrium allocation $\{Y_t, C_t, N_t, D_t, w_t, p_t\}$ in the representative-agent model with $\sigma = 1$ and no investment ($\Psi = \infty$ and $\delta = 0$) is also an equilibrium allocation in the heterogeneous-agent model with $\sigma = 1$ and no investment.*

In particular, monetary policy shocks will have the same effects on output.

Proof. Consider the HA model. Starting from the ergodic steady-state distribution at $t = 0$, and at first assuming real interest rates remain at r^{ss}, let $c_t^{ss}(\ell_{-1}, s^t)$ and $\ell_t^{ss}(\ell_{-1}, s^t)$ denote the date-t policies of agents as a function of liquidity at ℓ_{-1} and the history $s^t = (s_0, \ldots, s_t)$ of idiosyncratic shocks from date 0 to date t. (This *sequential form* of the problem will be more convenient for the proof.)

Optimal behavior is characterized for all t by

$$\left(c_t^{ss}\left(\ell_{-1}, s^t\right)\right)^{-1} \geq \beta\left(1+r^{ss}\right) \mathbb{E}_t\left[\left(c_{t+1}^{ss}\left(\ell_t^{ss}\left(\ell_{-1}, s^t\right), s^{t+1}\right)\right)^{-1} | s^t\right] \tag{29}$$

$$c_t^{ss}\left(\ell_{-1}, s^t\right)+\ell_t^{ss}\left(\ell_{-1}, s^t\right) = \left(1+r^{ss}\right) \ell_{t-1}^{ss}\left(\ell_{-1}, s^{t-1}\right)+w^{ss} N^{ss} e\left(s_t\right) \tag{30}$$

$$\ell_t^{ss}\left(\ell_{-1}, s^t\right) \geq 0 \tag{31}$$

where the Euler equation and the borrowing constraint hold with complementary slackness.

Now take an arbitrary monetary policy path $\{r_t\}$ and corresponding RA equilibrium sequences $\{Y_t, C_t, N_t, D_t, w_t, p_t\}$, and consider optimal household behavior subject to these equilibrium sequences . Denote policies by $c_t(\ell_{-1}, s^t)$ and $\ell_t(\ell_{-1}, s^t)$. Optimal behavior is characterized by

$$\left(c_t\left(\ell_{-1}, s^t\right)\right)^{-1} \geq \beta\left(1+r_{t+1}^{\ell}\right) \mathbb{E}_t\left[\left(c_{t+1}\left(\ell_{-1}, s^{t+1}\right)\right)^{-1} | s^t\right] \tag{32}$$

$$c_t\left(\ell_{-1}, s^t\right)+\ell_t\left(\ell_{-1}, s^t\right) = \left(1+r_t^{\ell}\right) \ell_{t-1}\left(\ell_{-1}, s^{t-1}\right)+w_t N_t e\left(s_t\right) \tag{33}$$

$$\ell_t\left(\ell_{-1}, s^t\right) \geq 0 \tag{34}$$

We now guess and verify that if $c_t^{ss}(\ell_{-1}, s^t)$ and $\ell_t^{ss}(\ell_{-1}, s^t)$ satisfy (29)-(31), then $c_t(\ell_{-1}, s^t) = \frac{Y_t}{Y^{ss}} c_t^{ss}(\ell_{-1}, s^t)$ and $\ell_t(\ell_{-1}, s^t) = \frac{Y_t}{Y^{ss}} \ell_t^{ss}(\ell_{-1}, s^t)$ satisfy (32)-(34). We will do so by explicitly showing the two sets of equations are equivalent.

This is trivially true for (34) and (31). For (32), plug in the candidate policy to obtain

$$\left(\frac{C_t}{C^{ss}}\right)^{-1}\left(c_t^{ss}\left(\ell_{-1}, s^t\right)\right)^{-1} \geq \beta\left(1+r_{t+1}^{\ell}\right)\left(\frac{C_{t+1}}{C^{ss}}\right)^{-1} \mathbb{E}_t\left[\left(c_{t+1}^{ss}\left(\ell_{-1}, s^{t+1}\right)\right)^{-\sigma} | s^t\right]$$

Here, dividing both sides by the RA Euler equation $C_t^{-1} = \beta(1+r_t)C_{t+1}^{-1}$, applying perfect foresight $r_t = r_{t+1}^{\ell}$, and dividing by C^{ss} gives us (29).

Similarly, for (33), plug in the candidate policy to obtain for the $t > 0$ case

$$\frac{Y_t}{Y^{ss}} c_t^{ss}\left(\ell_{-1}, s^t\right)+\frac{Y_t}{Y^{ss}} \ell_t^{ss}(\ell_{-1}, s^t) = \left(1+r_t^{\ell}\right) \frac{Y_t}{Y^{ss}} \ell_{t-1}^{ss}\left(\ell_{-1}, s^{t-1}\right)+w_t N_t e\left(s_t\right)$$

Here, noting that $w_t N_t = (1-\alpha)Y_t = \frac{Y_t}{Y^{ss}} w^{ss} N^{ss}$, we see that the equation is just (30) with $\frac{Y_t}{Y^{ss}}$ multiplying every term.

For $t = 0$, we have instead

$$\frac{Y_0}{Y^{ss}} c_0^{ss}\left(\ell_{-1}, s^t\right) + \frac{Y_0}{Y^{ss}} \ell_0^{ss}(\ell_{-1}, s) = \left(1 + r_0^{\ell}\right)\ell_{-1} + w_0 N_0 e\left(s_t\right)$$

which seems problematic since there is no longer a $\frac{Y_0}{Y^{ss}}$ multiplying the first term on the right. However, applying lemma 1,

$$1 + r_0^{\ell} = \frac{p_0 + D_0}{p^{ss}} = \frac{\frac{\alpha\beta^{RA}}{1-\beta^{RA}} Y_0 + \alpha Y_0}{\frac{\alpha\beta^{RA}}{1-\beta^{RA}} Y^{ss}} = (\beta^{RA})^{-1}\frac{Y_0}{Y^{ss}} = (1 + r^{ss})\frac{Y_0}{Y^{ss}}$$

so we again have equation (30) multiplied by $\frac{Y_0}{Y^{ss}}$. We conclude that $c_t(\ell_{-1}, s^t) = \frac{Y_t}{Y^{ss}} c_t^{ss}(\ell_{-1}, s^t)$ and $\ell_t(\ell_{-1}, s^t) = \frac{Y_t}{Y^{ss}} \ell_t^{ss}(\ell_{-1}, s^t)$ are an optimal plan for each household faced with RA equilibrium sequences $\{Y_t, C_t, N_t, D_t, w_t, p_t, r_t\}$.

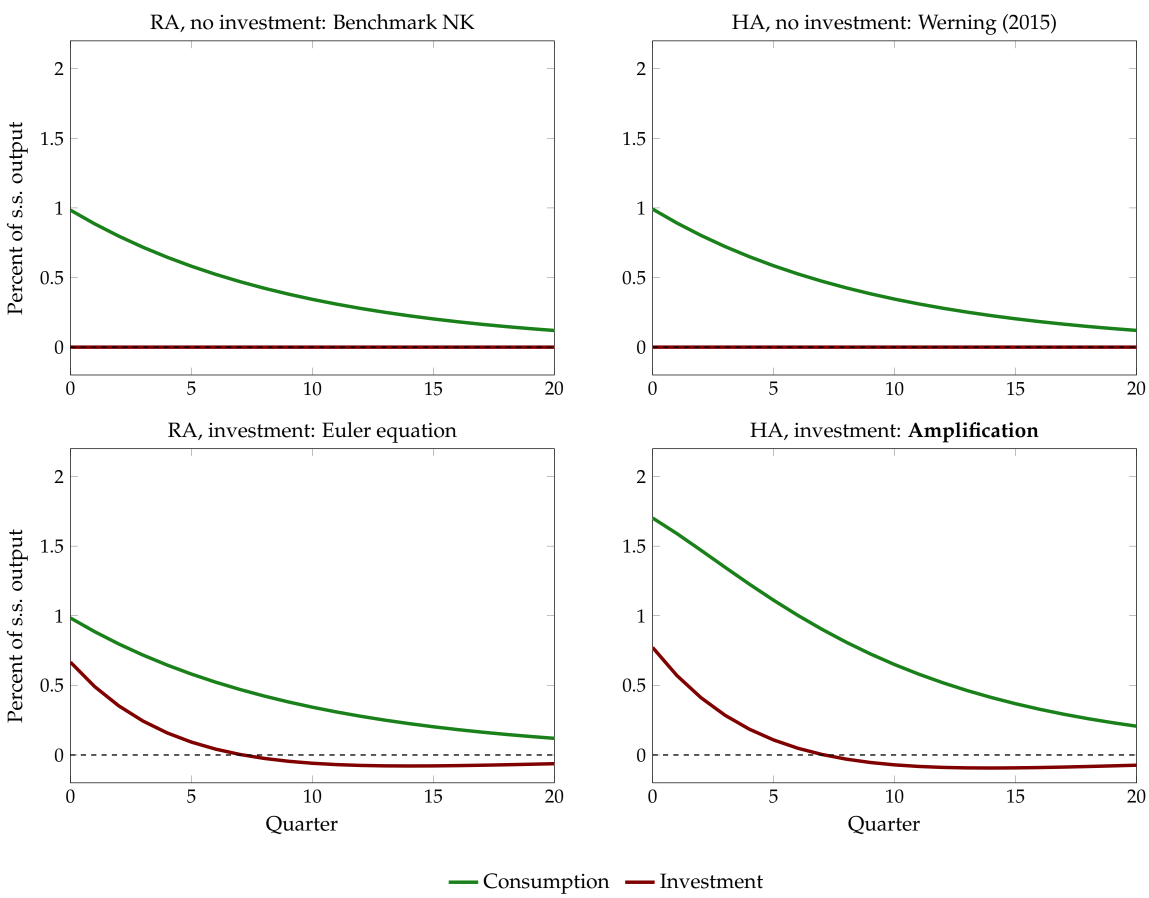
Since each household's consumption is scaled up by the same factor $\frac{Y_t}{Y^{ss}}$, aggregate consumption is also scaled up by that factor. Hence consumption $C_t = Y_t$ is the same as its RA equilibrium value, and goods market clearing holds. Asset market clearing follows from Walras' law, and all other equilibrium conditions are the same as in the RA model. We conclude that $\{Y_t, C_t, N_t, D_t, w_t, p_t, r_t\}$ is also an equilibrium for the HA model. □

A.5 Robustness to positive depreciation

In figure A.6, we recalculate figure A.1 in a model that is calibrated in exactly the same way, except that depreciation is set at its value $\delta = .054$ (annualized) from the main text, rather than at zero. The "no investment" case still features $\Psi = \infty$, in which case $K_t = K^{ss}$ and $I_t = \delta K^{ss}$ in all periods.

Although the analytical proof of proposition 1 no longer goes through when $\delta > 0$, we see numerically in the top panel of figure A.6 that the RA and HA models still deliver nearly identical results when $\Psi = \infty$. Similarly, in the bottom right panel, the interaction of the HA model with positive investment delivers substantial amplification. This is slightly smaller than in figure A.1 because a shock to r_t has less proportional effect on the user cost $r_t + \delta$, and therefore the incentive to invest, when δ is positive. (Although Ψ is recalibrated to match the same investment response at $t = 0$ in the RA model, the cumulative investment response adding subsequent periods is smaller.)

Figure A.6: Complementarity between investment and high MPCs: positive depreciation $\delta > 0$



B Appendix to section 2

B.1 RA model with additive internal habits

In partial equilibrium, given a process for income and interest rates $\{y_t, r_t\}$ and initial assets a_{-1}, a representative agent with external additive habit formation solves the following problem:

$$\begin{aligned} \max \quad & \mathbb{E}\left[\sum \beta^t u\left(c_t - \gamma c_{t-1}\right)\right] \\ & c_t + a_t = y_t + \left(1 + r_{t-1}\right) a_{t-1} \end{aligned}$$

where $0 \leq \gamma < 1$. The associated Euler equation is:

$$u'\left(c_t - \gamma c_{t-1}\right) - \mathbb{E}_t\left[\beta\gamma u'\left(c_{t+1} - \gamma c_t\right)\right] = \beta\left(1 + r_t\right)\mathbb{E}_t\left[u'\left(c_{t+1} - \gamma c_t\right) - \beta\gamma u'\left(c_{t+2} - \gamma c_{t+1}\right)\right] \tag{35}$$

Linearized solution. Linearizing (35) around a steady state with constant consumption c and $\beta^{-1} = (1 + r)$, we obtain

$$-\gamma dc_{t-1} + \left(1 + \gamma + \beta\gamma^2\right) dc_t - \left(1 + \beta\gamma + \beta\gamma^2\right)\mathbb{E}_t\left[dc_{t+1}\right] + \beta\gamma\mathbb{E}_t\left[dc_{t+2}\right] = -\frac{1}{\sigma}\left(1 - \beta\gamma\right)\left(1 - \gamma\right) c\frac{dr_t}{1 + r} \tag{36}$$

where $\frac{1}{\sigma} = -\frac{u''(c)c}{u'(c)}$ is the inverse curvature of u. We can rewrite (36) as

$$\mathbb{E}_t\left[P\left(L\right)\left(1 - L\right) dc_t\right] = -\kappa dr_t \tag{37}$$

where $\kappa \equiv \frac{1}{\sigma}\left(1 - \beta\gamma\right)\left(1 - \gamma\right)\frac{c}{1+r}$, and

$$P\left(X\right) \equiv \frac{b\left(X - \beta\gamma\right)\left(X - \frac{1}{\gamma}\right)}{X^2}$$

has two roots, one greater and one smaller than 1. The linear solution to the habits problem therefore jointly solves the Euler equation and the linearized version of the budget constraint,

$$\mathbb{E}_t\left[dc_{t+1} - dc_t\right] = \gamma\left(dc_t - dc_{t-1}\right) + \kappa\mathbb{E}_t\left[\sum_{k \geq 0}\left(\beta\gamma\right)^k dr_{t+k}\right] \tag{38}$$

$$dc_t = dy_t + \frac{1}{\beta} da_{t-1} - da_t + a dr_{t-1} \tag{39}$$

Intertemporal marginal propensities to consume. If $\mathbb{E}_t\left[dr_t\right] = 0$ for all t, then we can solve (38) to obtain

$$\mathbb{E}_0\left[dc_t\right] = \frac{1 - \gamma^{t+1}}{1 - \gamma} dc_0 \tag{40}$$

Integrating (39) and plugging in (40) , we then find

$$\frac{1}{1-\gamma}\left(\sum_{t=0}^{\infty}\beta^{t}\left(1-\gamma^{t+1}\right)\right)dc_{0}=dy_{0}$$

Hence, the expected path of consumption after an initial increase in income is:

$$\frac{\mathbb{E}_{0}\left[dc_{t}\right]}{dy_{0}}=\left(1-\beta\right)\left(1-\beta\gamma\right)\frac{1-\gamma^{t+1}}{1-\gamma} \tag{41}$$

which is plotted on the red line of figure 2, for an illustrative calibration with $\beta = 0.95$ and $\gamma = 0.6$. The initial MPC is depressed relative to $(1-\beta)$—that of the representative-agent model—by a factor $(1-\beta\gamma)$, reflecting the desire of the agent with additive habits to limit the initial increase in his habit stock.

B.2 HA model with additive internal habits

The heterogeneous-agent habit problem can be formulated as follows:

$$\begin{aligned}
V\left(\ell, c_{-}, s\right) \quad &= \quad \max_{c,\ell'} u\left(c-\gamma c_{-}\right)+\beta\mathbb{E}\left[V\left(\ell', c, s'\right)|s\right] \\
&\quad\quad c+\ell' = \left(1+r\right)\ell + ye\left(s\right) \\
&\quad\quad \ell' \geq 0
\end{aligned}$$

The first-order conditions for c and ℓ' are

$$\lambda = u'\left(c-\gamma c_{-}\right)+\beta\mathbb{E}\left[V_{c}\left(\ell', c, s'\right)|s\right] \tag{42}$$

$$\lambda+\mu = \beta\mathbb{E}\left[V_{\ell}\left(\ell', c, s'\right)|s\right] \tag{43}$$

where $\mu \geq 0$ is the multiplier on the borrowing constraint $\ell' \geq 0$. Moreover, the envelope conditions for ℓ and c_{-} imply

$$V_{\ell}\left(\ell, c_{-}, s\right) = \lambda\left(1+r\right) \tag{44}$$

$$V_{c_{-}}\left(\ell, c_{-}, s\right) = -\gamma u'\left(c-\gamma c_{-}\right) \tag{45}$$

We calibrate the model to $\gamma = 0.6$, $r = 0.05$, $u = \log$, and the same annual $\beta = 0.8422$ as that found in calibrating our no-habit HA model to match a first-year MPC of 0.55. We solve the model using standard methods, on a grid for (ℓ, c_{-}).

C Appendix to section 3

C.1 Euler equation for inattentive households

The optimal policy functions $c_{g,t}(\ell, a, a_{-k}, s, k)$ and $\ell'_{g,t}(\ell, a, a_{-k}, s, k)$ for the household problem (5), when the household is not constrained at the liquid asset lower bound $\ell'_{g,t}(\ell, a, a_{-k}, s, k) \geq 0$, satisfy the intertemporal Euler equation

$$u'(c_{g,t}(\ell, a, a_{-k}, s, k)) = \beta_g \mathbb{E}_{t-k}\left[(1+r^{\ell}_{t+1})\left(\theta u'(c_{g,t+1}(\ell'_{g,t}(\ell, a, a_{-k}, s, k), a', a_{-k}, s', k+1)\right.\right.$$
$$\left.\left. +(1-\theta)u'(c_{g,t+1}(\ell'_{g,t}(\ell, a, a_{-k}, s, k), a', a', s', 0))\right) |s\right] \quad (46)$$

which follows immediately from combining the first-order condition and envelope condition for (5). When the household is constrained, (46) is an inequality $\geq$.

C.2 Extended financial intermediary problem and monetary policy implementation

Here we extend the model of the financial intermediary in section 3.2 to allow it to allow nominal reserves M_t at the central bank, that pay a pre-determined interest rate of i_t. Since all assets are real, the flow-of-funds constraint (7) in date-t nominal units is modified to

$$(1+r^a_t)\,P_t A_{t-1} + \left(1+r^{\ell}_t\right) P_t L_{t-1} = (1+\delta q_t)\,P_t B_{t-1} + \int \left(p_{jt} + D_{jt}\right) P_t v_{jt-1} dj - \xi P_t L_{t-1} + (1+i_{t-1})\,M_{t-1} \quad (47)$$

and portfolio-investment constraint (8) now reads

$$P_t \int p_{jt} v_{jt} dj + P_t q_t B_t + M_t = P_t A_t + P_t L_t \quad (48)$$

The financial intermediary's problem is now to choose v_{jt}, B_t, L_t and M_t so as to maximize the expected return on illiquid liabilities, $\mathbb{E}_t\left[r^a_{t+1}\right]$, subject to (48) and (47). Since (47) implies that

$$\mathbb{E}_t\left[1+r^a_{t+1}\right] = \mathbb{E}_t\left[\frac{(1+\delta q_{t+1})\,B_t + \int \left(p_{jt+1} + D_{jt+1}\right) v_{jt} dj + (1+i_t)\,\frac{M_t}{P_{t+1}} - \left(1+r^{\ell}_{t+1}+\xi\right) L_t}{\int p_{jt} v_{jt} dj + q_t B_t + \frac{M_t}{P_t} - L_t}\right]$$

the first order conditions lead to equalization of all expected returns

$$\mathbb{E}_t\left[\frac{1+\delta q_{t+1}}{q_t}\right] = \mathbb{E}_t\left[\frac{p_{jt+1}+D_{jt+1}}{p_{jt}}\right] = (1+i_t)\,\mathbb{E}_t\left[\frac{P_t}{P_{t+1}}\right] = 1 + r^{\ell}_{t+1} + \xi$$

which are equations (9) and (10) in the main text, where we also define these to all be equal to the ex-ante real interest rate $1 + r_t$.

The central bank implements monetary policy by setting the nominal interest rate on reserves i_t, using open-market operations. We consider the limit where it does so using a net supply of reserves that is at all times equal to $M_t = 0$.

C.3 Intermediate goods firm price-setting

We first derive final demand. Individual consumers minimize $\int P_{jt}Y_{jt}dj$ subject to (11), which results in the first order condition

$$\frac{P_{jt}}{P_t} = G_p'\left(\frac{Y_{jt}}{Y_t}\right) = \frac{1-\left(\frac{Y_{jt}}{Y_t}\right)^{\frac{\upsilon_p}{\epsilon_p}}}{\upsilon_p}$$

hence, intermediate goods firms face the static demand curve

$$\frac{Y_{jt}}{Y_t} = \mathcal{Y}_p\left(\frac{P_{jt}}{P_t}\right)$$

where we have defined $\mathcal{Y}_p$ as

$$\mathcal{Y}_p(x) \equiv (1-\nu_p \log x)^{\frac{\epsilon_p}{\nu_p}}$$

Define the static profit function of an intermediate goods firm with current price p, when the price index is P, real marginal costs are s and aggregate demand is Y as

$$D(p;P,Y,s) \equiv \left(\frac{p}{P}-s\right)\mathcal{Y}_p\left(\frac{p}{P}\right)Y$$

and note that the derivative of D with respect to own price p is

$$\frac{\partial D}{\partial p} = \left(\frac{1}{P}+\left(\frac{s}{p}-\frac{1}{P}\right)\epsilon_p\left(\frac{p}{P}\right)\right)\mathcal{Y}_p\left(\frac{p}{P}\right)Y \tag{49}$$

where $\epsilon_p(x)$ is the elasticity of demand,

$$\epsilon_p(x) \equiv -\frac{\mathcal{Y}_p'(x)\,x}{\mathcal{Y}_p(x)} = \frac{\epsilon_p}{1-\nu_p \log x} \tag{50}$$

We next work out the optimal reset price for a firm. Upon receiving an opportunity to reset its price, a firm chooses P_t^* to maximize the sum of its dividend and its stock price,

$$D(P_t^*;P_t,Y_t,s_t) + p_t(P_t^*)$$

where by the no-arbitrage condition in (9) and the price indexation formula (12), we have

$$\begin{aligned} p_t(P_t^*) \quad = \quad & \frac{1}{1+r_t}\mathbb{E}_t\left[\zeta_p\left(D\left(P_t^*\frac{P_t}{P_{t-1}};P_{t+1},Y_{t+1},s_{t+1}\right)+p_{t+1}\left(P_t^*\frac{P_t}{P_{t-1}}\right)\right)\right. \\ & \left. +\,(1-\zeta_p)\max_{\widehat{p}}\theta\left(D(\widehat{p};P_{t+1},Y_{t+1},s_{t+1})+p_{t+1}(\widehat{p})\right)\right] \end{aligned}$$

Hence, defining $M_{t,t+k} \equiv \prod_{s=t}^{t+k-1} \frac{1}{1+r_s}$, P_t^* also solves

$$P_t^* = \arg\max_x \mathbb{E}_t \left[\sum_{k\geq 0} \zeta_p^k M_{t,t+k} D\left(x \frac{P_{t+k-1}}{P_{t-1}}; P_{t+k}, Y_{t+k}, s_{t+k} \right) \right]$$

Taking the first-order condition and using (49), we find that P_t^* solves

$$\begin{aligned}
&\mathbb{E}_t \left[\sum_{k\geq 0} \zeta_p^k \cdot M_{t,t+k} \cdot Y_{t+k} \cdot \mathcal{Y}_p \left(\frac{P_t^*}{P_{t-1}} \frac{P_{t+k-1}}{P_{t+k}} \right) \left(\frac{P_t^*}{P_{t-1}} \frac{P_{t+k-1}}{P_{t+k}} + \epsilon_p \left(\frac{P_t^*}{P_{t-1}} \frac{P_{t+k-1}}{P_{t+k}} \right) \left(s_{t+k} - \frac{P_t^*}{P_{t-1}} \frac{P_{t+k-1}}{P_{t+k}} \right) \right) \right] \\
&= \mathbb{E}_t \left[\sum_{k\geq 0} \zeta_p^k \cdot M_{t,t+k} \cdot Y_{t+k} \cdot f_p \left(\frac{P_t^*}{P_{t-1}} \frac{P_{t+k-1}}{P_{t+k}}, s_{t+k} \right) \right] \qquad (51) \\
&= 0
\end{aligned}$$

where we have defined the function $f_p(x,s)$ as

$$f_p(x,s) \equiv \mathcal{Y}_p(x) \left(x + \epsilon_p(x)(s-x) \right)$$

To derive a first-order approximation to the solution to (51), observe that

$$\begin{aligned}
\frac{\partial f_p}{\partial x} &= \mathcal{Y}_p'(x) \left(x + \epsilon_p(x)(s-x) \right) + \mathcal{Y}_p(x) \left(1 + \epsilon_p'(x)(s-x) - \epsilon_p(x) \right) \\
&= \mathcal{Y}_p(x) \left(-\frac{\epsilon_p(x)}{x} \left(x + \epsilon_p(x)(s-x) \right) + 1 + \epsilon_p(x) \frac{\epsilon_p'(x)\, x}{\epsilon_p(x)} \left(\frac{s-x}{x} \right) - \epsilon_p(x) \right) \\
&= \mathcal{Y}_p(x) \epsilon_p(x) \left(\frac{1}{\epsilon_p(x)} + \left[\frac{\nu_p}{\epsilon_p} - 1 \right] \epsilon_p(x) \left(\frac{s-x}{x} \right) - 2 \right)
\end{aligned}$$

where we have made use of the fact that $\frac{\epsilon_p'(x)x}{\epsilon_p(x)} = \frac{\nu_p}{\epsilon_p}\epsilon_p(x)$ from (50). Hence, around the steady state where $x^{ss} = 1$ and $s^{ss} = \frac{\epsilon_p - 1}{\epsilon_p}$, we have

$$\frac{\partial f_p}{\partial x}(x^{ss}, s^{ss}) = 1 \cdot \epsilon_p \cdot \left(\frac{1}{\epsilon_p} - \frac{\nu_p}{\epsilon_p} + 1 - 2 \right) = \epsilon_p \left(\frac{1 - \nu_p}{\epsilon_p} - 1 \right) = 1 - \nu_p - \epsilon_p$$

and similarly,

$$\frac{\partial f_p}{\partial s}(x^{ss}, s^{ss}) = \mathcal{Y}_p(x^{ss}) \epsilon(x^{ss}) = \epsilon_p$$

Totally differentiating (51) around the steady state where $M_{t,t+k}^{ss} Y_{t+k}^{ss} = \left(\frac{1}{1+r}\right)^k Y^{ss}$, we next find

$$\begin{aligned}
\mathbb{E}_t \Bigg[\sum_{k\geq 0} \zeta_p^k d\left(M_{t,t+k} \cdot Y_{t+k} \right) \cdot f_p(x^{ss}, s^{ss}) + \sum_{k\geq 0} \left(\frac{\zeta_p}{1+r} \right)^k \cdot Y^{ss} \frac{\partial f_p}{\partial x}(x^{ss}, s^{ss})\, d\left(\frac{P_t^*}{P_{t-1}} \frac{P_{t+k-1}}{P_{t+k}} \right) \\
+ \sum_{k\geq 0} \left(\frac{\zeta_p}{1+r} \right)^k Y^{ss} \frac{\partial f_p}{\partial s}(x^{ss}, s^{ss})\, ds_{t+k} \Bigg] = 0
\end{aligned}$$

and since $f_p\left(x^{ss}, s^{ss}\right) = 0$, this gives

$$\left(\epsilon_p + \nu_p - 1\right)\mathbb{E}_t\left[\sum_{k\geq 0}\left(\frac{\zeta_p}{1+r}\right)^k d\left(\frac{P_t^*}{P_{t-1}}\frac{P_{t+k-1}}{P_{t+k}}\right)\right] = \epsilon_p \mathbb{E}_t\left[\sum_{k\geq 0}\left(\frac{\zeta_p}{1+r}\right)^k ds_{t+k}\right] \tag{52}$$

Now write $p_t^* \equiv \log P_t^*$, $p_t \equiv \log P_t$, and $\pi_t \equiv \log\left(\frac{P_t}{P_{t-1}}\right)$. Since we are linearizing around a zero inflation steady state, we have

$$d\left(\frac{P_t^*}{P_{t-1}}\frac{P_{t+k-1}}{P_{t+k}}\right) = d\log\left(\frac{P_t^*}{P_{t-1}}\frac{P_{t+k-1}}{P_{t+k}}\right) = p_t^* - p_{t-1} + \pi_{t+k}$$

hence

$$\left(\epsilon_p + \nu_p - 1\right)\mathbb{E}_t\left[\sum_{k\geq 0}\left(\frac{\zeta_p}{1+r}\right)^k \left(p_t^* - p_{t-1} + \pi_{t+k}\right)\right] = \epsilon_p \mathbb{E}_t\left[\sum_{k\geq 0}\left(\frac{\zeta_p}{1+r}\right)^k ds_{t+k}\right]$$

which can be rewritten as

$$p_t^* - p_{t-1} \quad = \quad \left(1 - \frac{\zeta_p}{1+r}\right)\mathbb{E}_t\left[\sum_{k\geq 0}\left(\frac{\zeta_p}{1+r}\right)^k \cdot \left(\pi_{t+k} + \frac{\epsilon_p}{\epsilon_p + \nu_p - 1} ds_{t+k}\right)\right]$$

or recursively as

$$p_t^* - p_{t-1} = \left(1 - \frac{\zeta_p}{1+r}\right)\left(\pi_t + \frac{\epsilon_p}{\epsilon_p + \nu_p - 1} ds_t\right) + \frac{\zeta_p}{1+r}\mathbb{E}_t\left[p_{t+1}^* - p_t\right] \tag{53}$$

Moreover, the price index satisfies

$$\zeta_p\left(\frac{P_{t-1}\Pi_{t-1}}{P_t}\right)\cdot \mathcal{Y}\left(\frac{P_{t-1}\Pi_{t-1}}{P_t}\right) + \left(1-\zeta_p\right)\frac{P_t^*}{P_t}\cdot \mathcal{Y}\left(\frac{P_t^*}{P_t}\right) = 1 \tag{54}$$

Differentiating (54) around the zero inflation steady state, we obtain

$$\zeta_p\left(\pi_{t-1} - \pi_t\right) + \left(1-\zeta_p\right)\left(p_t^* - p_t\right) = 0$$

or

$$\pi_t = \zeta_p \pi_{t-1} + \left(1-\zeta_p\right)\left(p_t^* - p_{t-1}\right)$$

Combining with (53) and rearranging delivers

$$\pi_t - \zeta_p\pi_{t-1} = \left(1-\zeta_p\right)\left(1 - \frac{\zeta_p}{1+r}\right)\left(\frac{\epsilon_p}{\epsilon_p + \nu_p - 1} ds_t + \pi_t\right) + \frac{\zeta_p}{1+r}\mathbb{E}_t\left[\pi_{t+1} - \zeta_p \pi_t\right]$$

which can be rearranged as

$$\pi_t = \frac{1}{1+\frac{1}{1+r}}\pi_{t-1} + \frac{(1-\zeta_p)\left(1-\frac{\zeta_p}{1+r}\right)}{\zeta_p\left(1+\frac{1}{1+r}\right)}\frac{\epsilon_p}{\epsilon_p+\nu_p-1}ds_t + \frac{\frac{1}{1+r}}{1+\frac{1}{1+r}}\mathbb{E}_t\left[\pi_{t+1}\right]$$

or alternatively as

$$\pi_t - \pi_{t-1} = \frac{(1-\zeta_p)\left(1-\frac{\zeta_p}{1+r}\right)}{\zeta_p}\frac{\epsilon_p}{\epsilon_p+\nu_p-1}\mathbb{E}_t\left[\sum_{k\geq 0}\left(\frac{1}{1+r}\right)^k ds_t\right]$$

which is the expression used in the main text, equation (13).

C.4 Capital firms

The capital firm comes in with planned investment I_t and capital stock K_t, conducts this investment and pays dividend of

$$D_t^K\left(K_t, I_{t-1}, I_t\right) = r_t^K K_t - I_t\left(1+S\left(\frac{I_t}{I_{t-1}}\right)\right)$$

leaving it with $K_{t+1} = (1-\delta)K_t + I_t$ for next period. It also chooses I_{t+1} for next period in order to maximize its stock price of

$$p_t^K\left(K_{t+1}, I_t, I_{t+1}\right) = \frac{1}{1+r_t}\mathbb{E}_t\left[D_{t+1}^K\left(K_{t+1}, I_t, I_{t+1}\right) + p_{t+1}^K\left(K_{t+2}, I_{t+1}, I_{t+2}\right)\right]$$

The problem can therefore be written as

$$\max_{I_{t+1}}\left\{\mathbb{E}_t\left[D_{t+1}^K\left(K_{t+1}, I_t, I_{t+1}\right) + \max_{I_{t+2}} p_{t+1}^K\left(I_{t+1}+(1-\delta)K_{t+1}, I_{t+1}, I_{t+2}\right)\right]\right\}$$

The first order condition for I_{t+1} is

$$\mathbb{E}_t\left[1+S\left(\frac{I_{t+1}}{I_t}\right)+\frac{I_{t+1}}{I_t}S'\left(\frac{I_{t+1}}{I_t}\right)\right] = Q_t + Q_t^I \tag{55}$$

where we have defined Q_t and Q_t^I as, respectively,

$$\begin{aligned} Q_t &\equiv \mathbb{E}_t\left[\frac{\partial p_{t+1}^K}{\partial K_{t+2}}\right] \\ Q_t^I &\equiv \mathbb{E}_t\left[\frac{\partial p_{t+1}^K}{\partial I_{t+1}}\right] \end{aligned}$$

The envelope conditions for capital and previous investment are

$$\begin{aligned}\frac{\partial p_t^K}{\partial K_{t+1}} &= \frac{1}{1+r_t}\mathbb{E}_t\left[r_{t+1}^K + (1-\delta)\frac{\partial p_{t+1}^K}{\partial K_{t+2}}\right] \\ &= \frac{1}{1+r_t}\mathbb{E}_t\left[r_{t+1}^K + (1-\delta)\,Q_t\right]\end{aligned}$$

hence

$$Q_t = \mathbb{E}_t\left[\frac{\partial p_{t+1}^K}{\partial K_{t+2}}\right] = \mathbb{E}_t\left[\frac{1}{1+r_{t+1}}\left(r_{t+2}^K + (1-\delta)\,Q_{t+1}\right)\right]$$

which is equation (15) in the main text.

The envelope condition for previous investment is

$$\frac{\partial p_t^K}{\partial I_t} = \frac{1}{1+r_t}\mathbb{E}_t\left[\left(\frac{I_{t+1}}{I_t}\right)^2 S'\left(\frac{I_{t+1}}{I_t}\right)\right]$$

hence

$$Q_t^I = \mathbb{E}_t\left[\frac{1}{1+r_{t+1}}\left(\frac{I_{t+2}}{I_{t+1}}\right)^2 S'\left(\frac{I_{t+2}}{I_{t+1}}\right)\right]$$

plugging this expression back in (55) implies

$$1 + S\left(\frac{I_{t+1}}{I_t}\right) + \frac{I_{t+1}}{I_t}S'\left(\frac{I_{t+1}}{I_t}\right) = Q_t + \mathbb{E}_t\left[\frac{1}{1+r_{t+1}}\left(\frac{I_{t+2}}{I_{t+1}}\right)^2 S'\left(\frac{I_{t+2}}{I_{t+1}}\right)\right]$$

which is equation (14) in the main text. (15) and (14) jointly characterize investment dynamics.

C.5 Unions

Demand for labor services from union j is given by:

$$\frac{N_{jt}}{N_t} = \mathcal{Y}_w\left(\frac{W_{jt}}{W_t}\right) \tag{56}$$

where $\mathcal{Y}_w$ is Kimball demand,

$$\mathcal{Y}_w(x) \equiv (1 - \nu_w \log x)^{\frac{\epsilon_w}{\nu_w}}$$

The maximization problem of union j at time t is then

$$\mathbb{E}\left[\sum_{k\geq 0}\beta^k\zeta_w^k\left(\int\{u(c_{it+k}) - v(n_{it+k})\}\,d\mathcal{D}_{it+k}\right)\right]$$

taking as given (56), the wage indexation rule, and households budget constraints. Since each union is infinitesimal, it only takes into account its marginal effect on every household's consumption and labor supply. Note that the total real earnings of household i are

$$\begin{aligned} z_{it} &= (1-\tau_t) e_{it} \frac{\int_0^1 W_{jt} n_{ijt} dj}{P_t} \\ &= (1-\tau_t) e_{it} \frac{\int_0^1 W_{jt} \mathcal{Y}_w \left(\frac{W_{jt}}{W_t}\right) dj}{P_t} N_t \end{aligned}$$

The envelope theorem implies that we can evaluate indirect utility as if all income from the union wage change is consumed. Hence $\frac{\partial c_{it}}{\partial W_{jt}} = \frac{\partial z_{it}}{\partial W_{jt}}$, where

$$\begin{aligned} \frac{\partial z_{it}}{\partial W_{jt}} &= (1-\tau_t) e_{it} \frac{N_t}{P_t} \left(\mathcal{Y}_w \left(\frac{W_{jt}}{W_t}\right) + \frac{W_{jt}}{W_t} \mathcal{Y}_w' \left(\frac{W_{jt}}{W_t}\right) \right) \\ &= (1-\tau_t) e_{it} \frac{N_t}{P_t} \mathcal{Y}_w \left(\frac{W_{jt}}{W_t}\right) \left(1 - \epsilon_w \left(\frac{W_{jt}}{W_t}\right)\right) \end{aligned}$$

where $\epsilon_w(x) \equiv -\frac{x\mathcal{Y}_w'(x)}{\mathcal{Y}_w(x)}$. On the other hand, household i's total hours worked are

$$n_{it} \equiv \int_0^1 \mathcal{Y}_w \left(\frac{W_{jt}}{W_t}\right) N_t dj$$

so that

$$\begin{aligned} \frac{\partial n_{it}}{\partial W_{jt}} &= \mathcal{Y}_w' \left(\frac{W_{jt}}{W_t}\right) \frac{N_t}{W_t} \\ &= -\epsilon \left(\frac{W_{jt}}{W_t}\right) \mathcal{Y}_w \left(\frac{W_{jt}}{W_t}\right) \frac{N_t}{W_{jt}} \end{aligned}$$

hence, denoting by

$$U_s \equiv \int \{u(c_{is}) - v(n_{is})\} d\mathcal{D}_{is}$$

as average utility in period s, we have that the marginal change in aggregate utility induced by a change in a wage w is

$$\begin{aligned} \frac{\partial U_s}{\partial w} &= N_s \mathcal{Y}_w \left(\frac{w}{W_s}\right) \int \left\{ (1-\tau_s) u'(c_{is}) e_{is} \frac{1}{P_s} \left(1 - \epsilon_w \left(\frac{w}{W_s}\right)\right) + v'(n_{is}) \epsilon_w \left(\frac{w}{W_s}\right) \frac{1}{w} \right\} d\mathcal{D}_{is} \\ &= \left(\frac{MU_s}{P_s} + \left(\frac{MV_s}{w} - \frac{MU_s}{P_s} \right) \epsilon_w \left(\frac{w}{W_s}\right) \right) \mathcal{Y}_w \left(\frac{w}{W_s}\right) N_s \end{aligned}$$

where we have defined $MU_s \equiv (1-\tau_s) \int u'(c_{is}) e_{is} d\mathcal{D}_{is}$ and $MV_s \equiv \int v'(n_{is}) d\mathcal{D}_{is}$.

The union resets the wage knowing that, if it chooses wage w today, then its wage at any future time before it can reindex will be $w\frac{P_{t+k-1}}{P_{t-1}}$. The reset wage then solves

$$W_t^* = \arg\max \mathbb{E} \left[\sum_{k\geq 0} \beta^k \zeta_w^k U_{t+k} \left(w \frac{P_{t+k-1}}{P_{t-1}} \right) \right]$$

whose first order condition is

$$\mathbb{E}_t\left[\sum_{k\geq 0}\beta^k\zeta_w^k N_{t+k}\mathcal{Y}_w\left(\frac{W_t^*}{W_{t+k}}\frac{P_{t+k-1}}{P_{t-1}}\right)\right.$$
$$\left.\cdot\left(\left(1-\epsilon_w\left(\frac{W_t^*}{W_{t+k}}\frac{P_{t+k-1}}{P_{t-1}}\right)\right)\frac{W_t^*}{W_{t+k}}\frac{P_{t+k-1}}{P_{t-1}}\frac{MU_{t+k}W_{t+k}}{P_{t+k}}+\epsilon_w\left(\frac{W_t^*}{W_{t+k}}\frac{P_{t+k-1}}{P_{t-1}}\right)MV_{t+k}\right)\right]=0$$

We can rewrite this condition as

$$\mathbb{E}\left[\sum_{k\geq 0}\beta^k\zeta_w^k N_{t+k}\frac{MU_{t+k}W_{t+k}}{P_{t+k}}f_w\left(\frac{W_t^*}{W_{t+k}}\frac{P_{t+k-1}}{P_{t-1}},s_{w,t+k}\right)\right]=0$$

with f_w defined symmetrically to f_p in section C.3, and the inverse wage markup is defined as

$$s_{w,t}\equiv\frac{MV_t}{\frac{MU_tW_t}{P_t}}=\frac{\int v'\left(n_{it}\right)di}{\left(1-\tau_t\right)w_t\int e_{it}u'\left(c_{it}\right)di}=\frac{v'\left(N_t\right)}{\left(1-\tau_t\right)w_tu'\left(C_t^*\right)}$$

with C_t^* satisfying $u'\left(C_t^*\right)\equiv\int e_{it}u'\left(c_{it}\right)di$, as in Auclert, Rognlie and Straub (2018).

The derivation follows the same steps as in section C.3. Linearizing around the steady state, where $s_w=\frac{\epsilon_w-1}{\epsilon_w}$, we obtain

$$\left(\epsilon_w+\nu_w-1\right)\mathbb{E}_t\left[\sum_{k\geq 0}\left(\beta\zeta_w\right)^k\frac{d\left(\frac{W_t^*}{W_{t+k}}\frac{P_{t+k-1}}{P_{t-1}}\right)}{\frac{W_t^*}{W_{t+k}}\frac{P_{t+k-1}}{P_{t-1}}}\right]=\epsilon_w\mathbb{E}_t\left[\sum_{k\geq 0}\left(\beta\zeta_w\right)^k d\left(s_{w,t+k}\right)\right]$$

rearranging, and defining $\omega_t=\log W_t$, this is also

$$\left(\epsilon_w+\nu_w-1\right)\mathbb{E}_t\left[\sum_{k\geq 0}\left(\beta\zeta_w\right)^k\left(\omega_t^*-p_{t-1}-\left(\omega_{t+k}-p_{t+k-1}\right)\right)\right]=\epsilon_w\mathbb{E}_t\left[\sum_{k\geq 0}\left(\beta\zeta_w\right)^k d\left(s_{w,t+k}\right)\right]$$

or

$$\omega_t^*-p_{t-1}\quad=\quad\left(1-\beta\zeta_w\right)\left(\omega_t-p_{t-1}+\frac{\epsilon_w}{\epsilon_w+\nu_w-1}ds_t^w\right)+\beta\zeta_w\mathbb{E}_t\left[w_{t+1}^*-p_t\right]\tag{57}$$

Moreover, the wage index satisfies

$$\zeta_w\frac{W_{t-1}}{W_t}\Pi_{t-1}\mathcal{Y}_w\left(\frac{W_{t-1}}{W_t}\Pi_{t-1}\right)+\left(1-\zeta_w\right)\frac{W_t^*}{W_t}\mathcal{Y}_w\left(\frac{W_t^*}{W_t}\right)$$

which, in linear form, reads

$$\left(1-\zeta_w\right)\left(\omega_t^*-p_{t-1}\right)=\omega_t-p_{t-1}-\zeta_w\left(\omega_{t-1}-p_{t-2}\right)$$

Then, plugging in (57)

$$\begin{aligned}
(1-\zeta_w)\left(\omega_t^*-p_{t-1}\right) &= \omega_t-p_{t-1}-\zeta_w\left(\omega_{t-1}-p_{t-2}\right) \\
&= (1-\zeta_w)\left(1-\beta\zeta_w\right)\left(\omega_t-p_{t-1}+\frac{\epsilon_w}{\epsilon_w+\nu_w-1}ds_{w,t}\right)+\beta\zeta_w\left(1-\zeta_w\right)\mathbb{E}_t\left[\omega_{t+1}^*-p_t\right] \\
&= (1-\zeta_w)\left(1-\beta\zeta_w\right)\left(\omega_t-p_{t-1}+\frac{\epsilon_w}{\epsilon_w+\nu_w-1}ds_{w,t}\right) \\
&\quad +\beta\zeta_w\mathbb{E}_t\left[\omega_{t+1}-p_t-\zeta_w\left(\omega_t-p_{t-1}\right)\right]
\end{aligned}$$

and using $1-(1-\zeta_w)\left(1-\beta\zeta_w\right)+\beta\zeta_w^2=\zeta_w\left(1+\beta\right)$, we obtain

$$\omega_t-p_{t-1}=\frac{1}{1+\beta}\left(\omega_{t-1}-p_{t-2}\right)+\frac{\beta}{1+\beta}\mathbb{E}_t\left[\omega_{t+1}-p_t\right]+\frac{(1-\zeta_w)\left(1-\beta\zeta_w\right)}{1+\beta}\frac{\epsilon_w}{\epsilon_w+\nu_w-1}ds_{w,t}$$

In present value form, defining $\pi_{w,t}\equiv\omega_t-\omega_{t-1}=\log\left(\frac{W_t}{W_{t-1}}\right)$, this reads

$$\pi_{w,t}-\pi_{t-1}=\frac{\left(1-\beta\zeta_w\right)\left(1-\zeta_w\right)}{\zeta_w}\frac{\epsilon_w}{\epsilon_w+\nu_w-1}\mathbb{E}_t\left[\sum_k\beta^k\left(s_{w,t+k}-\frac{\epsilon_w-1}{\epsilon_w}\right)\right]$$

which is expression (18) in the main text.

C.6 Walras's law

Aggregate across all households,

$$\begin{aligned}
C_t+L_t &= \left(1+r_{t-1}-\xi\right)L_{t-1}+Z_t+d_t \\
A_t &= \left(1+r_t^a\right)A_{t-1}-d_t
\end{aligned}$$

where d_t are aggregate distributions from liquid to illiquid account. Consolidating, and using the definition of Z_t, we find

$$C_t+L_t+A_t=\left(1+r_{t-1}^{\ell}\right)L_{t-1}+\left(1+r_t^a\right)A_{t-1}+\left(1-\tau_t\right)w_tN_t$$

Using the government budget constraint (19), we next have

$$C_t+G_t+L_t+A_t+\left(1+\delta q_t\right)B_{t-1}=\left(1+r_{t-1}^{\ell}\right)L_{t-1}+\left(1+r_t^a\right)A_{t-1}+w_tN_t+q_tB_t$$

Finally, using the incoming flow of funds constraint for the financial intermediary (7),

$$C_t+G_t+L_t+A_t=\left(p_t+D_t\right)v_{t-1}-\xi L_{t-1}+w_tN_t+q_tB_t$$

then using the outgoing flow of funds constraint (8),

$$C_t + G_t + p_t v_t = (p_t + D_t) v_{t-1} - \xi L_{t-1} + w_t N_t$$

using market clearing condition for shares $v_t = 1$,

$$C_t + G_t + \xi L_{t-1} = w_t N_t + D_t$$

and finally, using the expression for dividends in (16), we obtain

$$C_t + G_t + I_t + I_t S\left(\frac{I_t}{I_{t-1}}\right) + \xi L_{t-1} = Y_t$$

which is the goods market clearing condition.

D Appendix to section 4

D.1 Calibration of the fiscal rule

Our calibration of the fiscal rule parameter ψ in (20) is informed by existing estimates from the fiscal rule literature, following Leeper (1991)'s seminal paper. We calibrate rather than estimate this parameter, because our model outcomes are insensitive to the value of ψ within a wide range, as we show below.

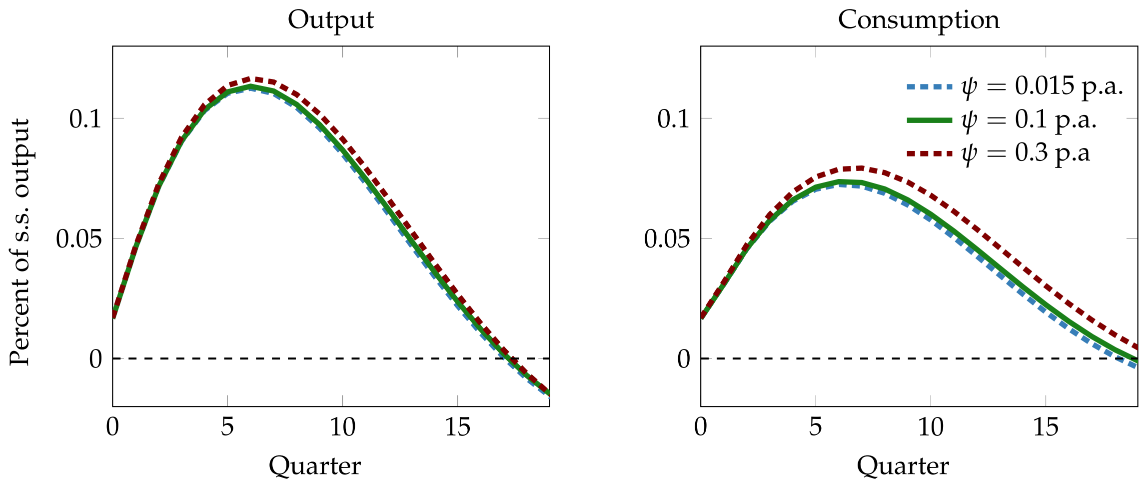
There exists a wide range of estimates for ψ, all of which tend to imply that the fiscal adjustment to shocks is delayed. Two representative examples from the literature are Davig and Leeper (2011) and Auclert and Rognlie (2018).

Davig and Leeper (2011) regress the ratio of federal receipts net of federal transfers to GDP on the debt-to-GDP ratio $\frac{q^{ss} B_{t-1}}{Y_t}$. Their estimate corresponds to an annualized value of $\psi = 0.28$. However, this is only an estimate for the active fiscal regime, which they estimate to be in place for half of their 1949:Q1 to 2008:Q4 sample (the estimate for the passive fiscal regime is $\psi = -0.1$ annually). Moreover, their numbers do not directly correspond to our specification in (20), which divides the face value of debt $q^{ss} B_{t-1}$ by steady-state rather than current GDP.

Auclert and Rognlie (2018)'s specification is closer to ours, since their regressor is the face value of debt divided by potential GDP, $\frac{q^{ss} B_{t-1}}{Y_t^{pot}}$. Combining their estimates for government spending and deficits, we obtain $\psi = -0.015 + 0.0288 \approx 0.015$ at an annual level. The implied estimates for ψ from Fernández-Villaverde, Guerrón-Quintana, Kuester and Rubio-Ramírez (2015) and Bianchi and Melosi (2017) lie somewhere between $\psi = 0.015$ and $\psi = 0.3$.

Altogether, we take $\psi = 0.015$ and $\psi = 0.30$ to be extreme points from the literature, and therefore pick $\psi = 0.1$ as our baseline calibration value.

Figure D.1: Impulse responses of output and consumption for various calibrated values of ψ



Note. This figure shows the impulse responses of output and consumption for different calibrated values of ψ. Our central model estimates are for $\psi = 0.1$ (solid green line). We then hold all other parameters fixed and recompute impulse responses for our extreme values of $\psi = 0.015$ and $\psi = 0.3$.

Robustness to alternative calibrations of ψ. Figure D.1 illustrates that our model outcomes are very insensitive to the value of ψ, given the existing range from the literature. Starting from our central estimates with our calibrated value of $\psi = 0.1$, we recompute impulse responses for the extreme values of $\psi = 0.015$ and $\psi = 0.3$ discussed above. We find that the impulse responses are almost identical, irrespective of ψ.

Estimating ψ. The results above suggest that there is little information in our impulse responses that can help identify ψ. However, in a further robustness exercise, we add nominal federal government current tax receipts divided by nominal GDP to our list of observables, estimate the impulse response of tax revenue to a monetary policy shock, and use this together with the other impulse responses in the main text to produce an estimate of ψ. This exercise yields $\psi = 0.13 \pm 0.18$. The point estimate suggests that our calibrated value of ψ is reasonable. The confidence bands are very wide, however, because with long-term debt, the fiscal impact of a monetary shock is in practice not large enough to provide sufficient identifying variation.

D.2 Model DAG and sequence-space Jacobian solution method

Figure D.2 displays the blocks for our model as a directed acyclic graph (DAG).

As discussed in Auclert et al. (2019), DAGs are useful devices to summarize how the model is computed and obtain impulse responses by chaining Jacobians. Endogenous sequences that are not the output of any block are at the left of the DAG, labeled "unknowns". Stacking these sequences in a vector $\mathbf{U}$, and stacking any exogenous sequences in $\mathbf{Z}$, we can evaluate each

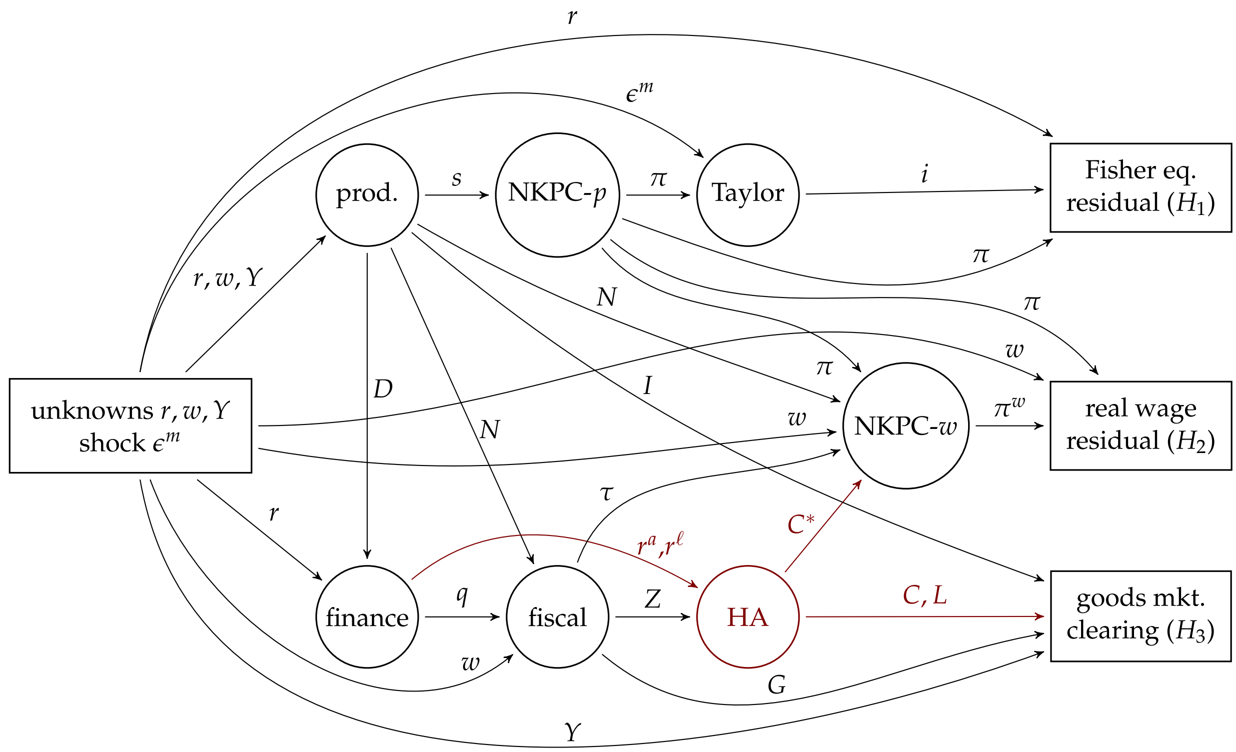


Figure D.2: Directed Acyclic Graph of the model

block in suitable order along the DAG to obtain every other endogenous sequence, including several—labeled H_1, H_2, and H_3 in the figure—that must be zero in equilibrium, and which we call "targets". Overall, then, the DAG represents a mapping

$$\mathbf{H}(\mathbf{U}, \mathbf{Z}) = 0 \tag{58}$$

As Figure D.2 shows, we set up our model so that the unknowns are $U_t \equiv (r_t, w_t, Y_t)$, the sequences of real interest rates, wages, and output. Corresponding to these are our three targets: first, the Fisher equation residual under perfect foresight,

$$H_{1t} = 1 + r_t - (1 + i_t) \frac{P_{t+1}}{P_t}$$

which in equilibrium, when $H_{1t} = 0$, corresponds to equation (10). Second, the real wage residual,

$$H_{2t} = \log \left(\frac{w_t}{w_{t-1}} \right) - (\pi_t^w - \pi_t)$$

which in equilibrium, when $H_{2t} = 0$, imposes consistency between the definition of the real wage

$w_t = \frac{W_t}{P_t}$, wage inflation $\pi_t^w = \log\left(\frac{W_t}{W_{t-1}}\right)$, and price inflation $\pi_t = \log\left(\frac{P_t}{P_{t-1}}\right)$. Finally, we have the goods market clearing residual

$$H_{3t} = C_t + G_t + I_t + I_t S\left(\frac{I_t}{I_{t-1}}\right) + \xi L_{t-1} - Y_t$$

which in equilibrium ensures goods market clearing at all times.

Our procedure solves equation (58) for $\mathbf{U}$ to first order around the steady state, as follows. Each block in the DAG is characterized to first order by the Jacobian matrices $\mathcal{J}^{o,i}$ for input sequences i and output sequences o. We combine these $\mathcal{J}$ using the chain rule to obtain the Jacobians $\mathbf{H_U}$ and $\mathbf{H_Z}$ of (58). This then provides a linear map from exogenous shocks to unknowns,

$$d\mathbf{U} = -\mathbf{H}_{\mathbf{U}}^{-1}\mathbf{H_Z}d\mathbf{Z}$$

Finally, we obtain all other sequences to first order given $d\mathbf{U}$ and $d\mathbf{Z}$, by applying $\mathcal{J}$'s along the DAG an extra time.

D.3 Details on solution method with informational rigidities

Deriving the recursion for sticky expectations. As discussed in section 4.3, if $\tau \leq s$, the impulse response of a household learning at date τ about a date-s change in input i is the same as the impulse response of a household who learns at date 0 about a date-$(s-\tau)$ change in i, shifted by τ periods. Both are the impulse response to a news shock about the value of i, $(s-\tau)$ periods in the future. This can be written as

$$\mathcal{J}_{t,s}^{o,i,\tau} = \mathcal{J}_{t-1,s-1}^{o,i,\tau-1} = \cdots = \mathcal{J}_{t-\tau,s-\tau}^{o,i,0} \tag{59}$$

If $\tau > s$, on the other hand, then we have $\mathcal{J}_{t,s}^{o,i,\tau} = \mathcal{J}_{t,s}^{o,i,s}$ for all t: we assume that the household is aware at date s of all inputs i to its problem at date s, so if not prior to s, τ is irrelevant.

These two observations allow us to simplify (23) for a given s, writing

$$\mathcal{J}_{t,s}^{o,i} = \theta^s \mathcal{J}_{t,s}^{o,i,s} + (1-\theta)\sum_{\tau=0}^{s-1}\theta^\tau \mathcal{J}_{t,s}^{o,i,\tau} \tag{60}$$

Applying (59) to each term of (60) except where $\tau = 0$, we can write for any $t, s > 0$

$$\begin{aligned}
\mathcal{J}_{t,s}^{o,i} &= \theta^s \mathcal{J}_{t-1,s-1}^{o,i,s-1} + (1-\theta)\sum_{\tau=0}^{s-2}\theta^{\tau+1}\mathcal{J}_{t-1,s-1}^{o,i,\tau} + (1-\theta)\mathcal{J}_{t,s}^{o,i,0} \\
&= \theta\mathcal{J}_{t-1,s-1}^{o,i} + (1-\theta)\mathcal{J}_{t,s}^{o,i,0} \qquad (61)
\end{aligned}$$

where the second step consolidates the first two terms in the previous line using (60).

For $s = 0$, (60) simplifies to just $\mathcal{J}_{t,0}^{o,i} = \mathcal{J}_{t,0}^{o,i,0}$. For $t = 0$ and $s > 0$, there is no response unless

$\tau = 0$, so $\mathcal{J}_{0,s} = (1-\theta)\mathcal{J}^0_{0,s}$. Combining all results, we obtain

$$\mathcal{J}^{o,i}_{t,s} = \begin{cases} \theta \mathcal{J}^{o,i}_{t-1,s-1} + (1-\theta)\mathcal{J}^{o,i,0}_{t,s} & t > 0,\ s > 0 \\ \mathcal{J}^{o,i,0}_{t,s} & s = 0 \\ (1-\theta)\mathcal{J}^{o,i,0}_{t,s} & t = 0,\ s > 0 \end{cases}$$

But $\mathcal{J}^{o,i,0}_{t,s}$, the Jacobian for households that learn at date $\tau = 0$ about shocks, is also, by definition, the full-information Jacobian $\mathcal{J}^{o,i,FI}_{t,s}$, so (24) follows.

Implementation for other behavioral or informational frictions. Here, to illustrate the method's generality, we use an analogous approach to derive the transformation of the full-information Jacobian associated with some other frictions.

Cognitive discounting. Under Gabaix (2016)'s "cognitive discounting" friction, at the micro level agents' expectations of disturbances k periods in the future shrink by a factor of $\bar{m}^k$ relative to rational expectations, where $\bar{m} \in [0,1]$ is a parameter capturing cognitive discounting.

If at date 0 there is a news shock about some change to input i at date s, agents subject to cognitive discounting perceive this instead as a series of news shocks: they learn about a fraction $\bar{m}^s$ of the change at date 0, a fraction $\bar{m}^{s-1} - \bar{m}^s$ of the change at date 1, and so on, up until they learn about the final fraction $1 - \bar{m}$ when the change actually happens at date s.

Using the same notation, the analog of (60) here is then

$$\mathcal{J}^{o,i}_{t,s} = \bar{m}^s \mathcal{J}^{o,i,0}_{t,s} + (\bar{m}^{s-1} - \bar{m}^s)\mathcal{J}^{o,i,1}_{t,s} + (\bar{m}^{s-2} - \bar{m}^{s-1})\mathcal{J}^{o,i,2}_{t,s} \ldots + (1-\bar{m})\mathcal{J}^{o,i,s}_{t,s} \tag{62}$$

Applying (59) to each term of (62) except the first, we can write for any $t, s > 0$

$$\begin{aligned} \mathcal{J}^{o,i}_{t,s} &= \bar{m}^s \mathcal{J}^{o,i,0}_{t,s} + (\bar{m}^{s-1} - \bar{m}^s)\mathcal{J}^{o,i,0}_{t-1,s-1} + (\bar{m}^{s-2} - \bar{m}^{s-1})\mathcal{J}^{o,i,1}_{t-1,s-1} \ldots + (1-\bar{m})\mathcal{J}^{o,i,s-1}_{t-1,s-1} \\ &= \bar{m}^s(\mathcal{J}^{o,i,0}_{t,s} - \mathcal{J}^{o,i,0}_{t-1,s-1}) + \bar{m}^{s-1}\mathcal{J}^{o,i,0}_{t-1,s-1} + (\bar{m}^{s-2} - \bar{m}^{s-1})\mathcal{J}^{o,i,1}_{t-2,s-2} + \ldots + (1-\bar{m})\mathcal{J}^{o,i,s-1}_{t-1,s-1} \\ &= \bar{m}^s(\mathcal{J}^{o,i,0}_{t,s} - \mathcal{J}^{o,i,0}_{t-1,s-1}) + \mathcal{J}^{o,i}_{t-1,s-1} \end{aligned} \tag{63}$$

For $s = 0$, (62) simplifies to just $\mathcal{J}^{o,i}_{t,0} = \mathcal{J}^{o,i,0}_{t,0}$, and for $t = 0$ and $s > 0$, $\mathcal{J}^{o,i,\tau}_{t,s} = 0$ for all $\tau > 0$, so that $\mathcal{J}^{o,i}_{0,s} = \bar{m}^s \mathcal{J}^{o,i,0}_{0,s}$. Combining all results and writing $\mathcal{J}^{o,i,FI}_{t,0} = \mathcal{J}^{o,i,0}_{t,0}$ for the full-information Jacobian, we have the recursion

$$\mathcal{J}^{o,i}_{t,s} = \begin{cases} \bar{m}^s(\mathcal{J}^{o,i,FI}_{t,s} - \mathcal{J}^{o,i,FI}_{t-1,s-1}) + \mathcal{J}^{o,i}_{t-1,s-1} & t > 0,\ s > 0 \\ \mathcal{J}^{o,i,FI}_{t,s} & s = 0 \\ \bar{m}^s \mathcal{J}^{o,i,FI}_{t,s} & t = 0,\ s > 0 \end{cases} \tag{64}$$

which can transform the full-information Jacobian $\mathcal{J}^{o,i,FI}_{t,s}$ into the Jacobian $\mathcal{J}^{o,i}_{t,s}$ with cognitive discounting with only a single evaluation for each Jacobian entry.

Noisy information about shocks. An alternative approach to model informational frictions is to have agents learn from noisy information about inputs. A particularly clean version of this approach, for our purposes, is laid out in Appendix E of Angeletos and Huo (2019), and we follow a special case of it here. In particular, for a date-0 shock $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ that causes expected future inputs i to change, suppose that at each date $t \geq 0$, all agents receive independent private signals $\epsilon + \nu_t$ about the shock, where $\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$. Agents receive no other information about the shock.[50] Then, applying standard Bayesian updating, the average belief about ϵ at date j is

$$\bar{\mathbb{E}}_j \epsilon = \underbrace{\frac{(j+1)\tau_\nu}{\tau_\epsilon + (j+1)\tau_\nu}}_{\equiv a_j} \epsilon \tag{65}$$

where $\tau_\epsilon \equiv 1/\sigma_\epsilon^2$ and $\tau_\nu \equiv 1/\sigma_\nu^2$ are the precisions of the shock and signal, respectively. This is because by date j, each agent has combined the prior on ϵ with $j+1$ noisy signals $\epsilon + \nu_0, \ldots, \epsilon + \nu_j$.

On average, then, agents receive a news shock about ϵ of $a_0 \epsilon$ at date 0, $(a_1 - a_0)\epsilon$ at date 1, and so on. They receive proportional news shocks about the changes in inputs i, until they learn fully about the change in input i at date s once date s actually arrives.

The analog of (60) is then

$$\mathcal{J}_{t,s}^{o,i} = (1 - a_{s-1}) \mathcal{J}_{t,s}^{o,i,s} + \sum_{\tau=0}^{s-1} (a_\tau - a_{\tau-1}) \mathcal{J}_{t,s}^{o,i,\tau} \tag{66}$$

where we take $a_{-1} = 0$. Applying (59) to this to reduce all superscripts τ to 0, and also using $\mathcal{J}_{t,s}^{o,i,\tau} = 0$ for $t < s, \tau$, along with $\mathcal{J}_{t,0}^{o,i,FI} = \mathcal{J}_{t,0}^{o,i,0}$, we get

$$\mathcal{J}_{t,s}^{o,i} = \begin{cases} (1 - a_{s-1}) \mathcal{J}_{t-s,0}^{o,i,FI} + \sum_{\tau=0}^{s-1} (a_\tau - a_{\tau-1}) \mathcal{J}_{t-\tau,s-\tau}^{o,i,FI} & t \geq s \\ \sum_{\tau=0}^{t} (a_\tau - a_{\tau-1}) \mathcal{J}_{t-\tau,s-\tau}^{o,i,FI} & t < s \end{cases} \tag{67}$$

Since the a_τ do not decay exponentially, it is impossible to simplify this further into a recursive form as in the prior examples. Still, directly applying (67) to calculate $\mathcal{J}^{o,i}$ from $\mathcal{J}^{o,i,FI}$, when both are $T \times T$ matrices, only takes $\mathcal{O}(T^3)$ operations, the same as matrix multiplication and inversion—which are already done many times as part of the Auclert, Bardóczy, Rognlie and Straub (2019) solution method. Even here, therefore, the additional computational burden from converting the full-information Jacobian to the frictional Jacobian is slight.[51]

[50] In particular, agents do not extract information about the shock ϵ from changes in variables like i, once they are actually observed. If agents did, then without noise from additional individual-level shocks, they would be able to back out ϵ perfectly. Though we conjecture that our methods should still apply in a model augmented with such additional shocks—with somewhat greater complexity due to the endogeneity of the observed variables—this is beyond the scope of the current paper.

[51] Since the sums in (67) are convolutions of the sequence $\{a_\tau - a_{\tau-1}\}$ with the diagonals of the $\mathcal{J}^{o,i,FI}$ matrix, it is possible to use the Fast Fourier Transform to speed up computation to $\mathcal{O}(T^2 \log T)$, but since it is already not a

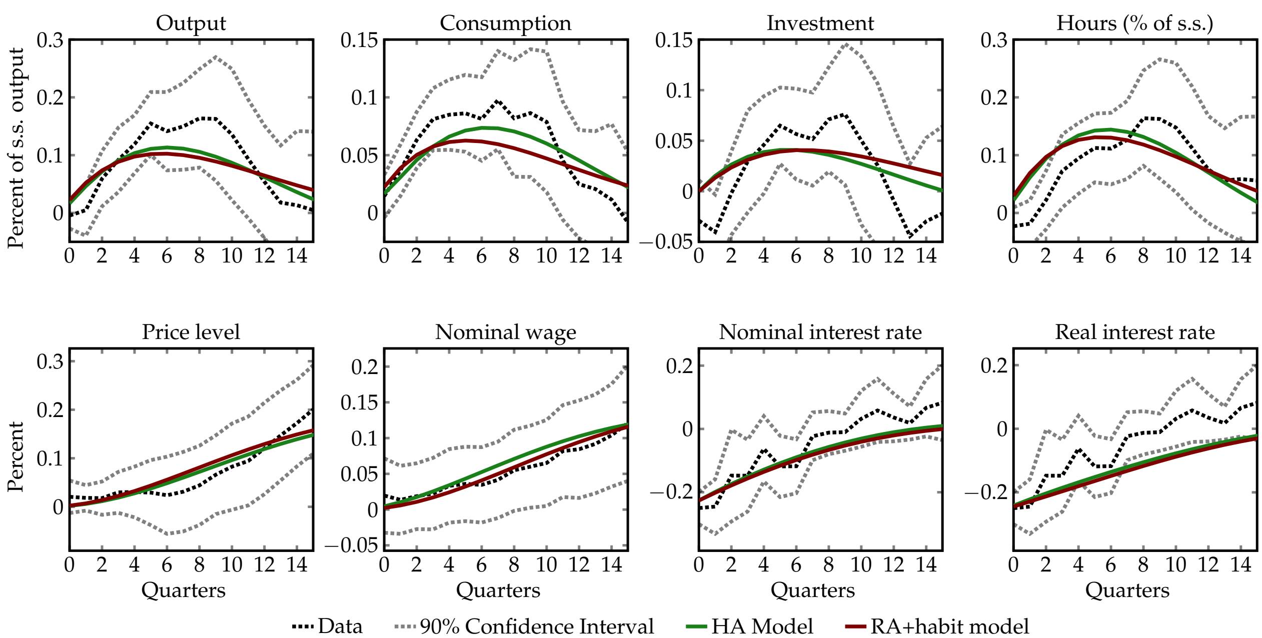
D.4 Estimated RA-habit model

Here we estimate the RA-habit model, using the procedure described in section 4.3 on the set of impulse responses described in section 4.2. Table D.1 displays the estimated parameters, figure D.3 shows the fit compared to that of our estimated HA model.

Table D.1: Estimated parameters for RA model.

Parameter		Value	std. dev.
γ	Household habit parameter	0.878	(0.012)
ϕ	Investment adjustment cost parameter	14.851	(4.016)
ζ_p	Calvo price stickiness	0.880	(0.042)
ζ_w	Calvo wage stickiness	0.946	(0.025)
ρ^m	Taylor rule inertia	0.904	(0.008)
σ^m	Std. dev. of monetary shock	0.057	(0.005)

Figure D.3: Impulse response to a monetary policy shock vs. model fit of HA and RA



Note. This figure shows our estimated set of impulse responses to an identified Romer and Romer (2004) monetary policy shock (dashed black, with gray confidence intervals). The solid lines are the impulse responses implied by our estimated inattentive heterogeneous-agent model (green) and a representative-agent model (red).

bottleneck, in practice this seems unnecessary.

E Appendix to section 5

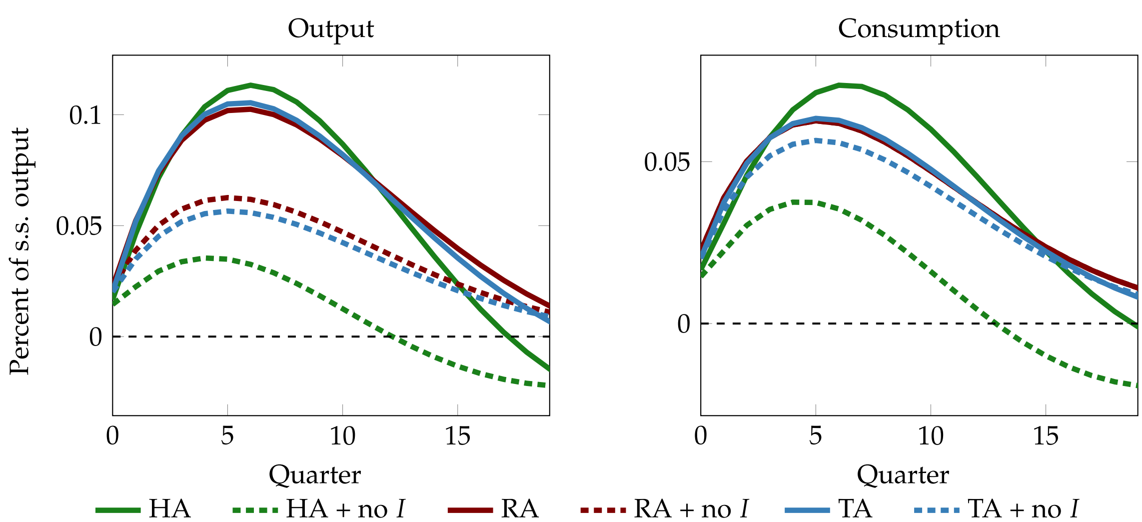
E.1 Investment counterfactual in TA-habit

Here we set up a two-agent version of our RA-habit model. The model is identical to the RA-habit model described in the main text, except that it features a share μ of hand-to-mouth households who consume their net-of tax income, $C_t^{HTM} = Z_t$. We choose $\mu = 0.20$ in line with the average MPC in figure 2. Table E.1 shows the estimated parameter values. Figure E.1 repeats the investment counterfactual of section 5, but using the estimated TA model as baseline. Evidently, in the TA model, shutting off investment has a much weaker effect on consumption than in our estimated HA model. This is in line with the recent results in Bilbiie, Känzig and Surico (2019) and can be explained by the fact that the TA model does not deliver a good match to the intertemporal MPCs shown in figure 2.

Table E.1: Estimated parameters for TA model.

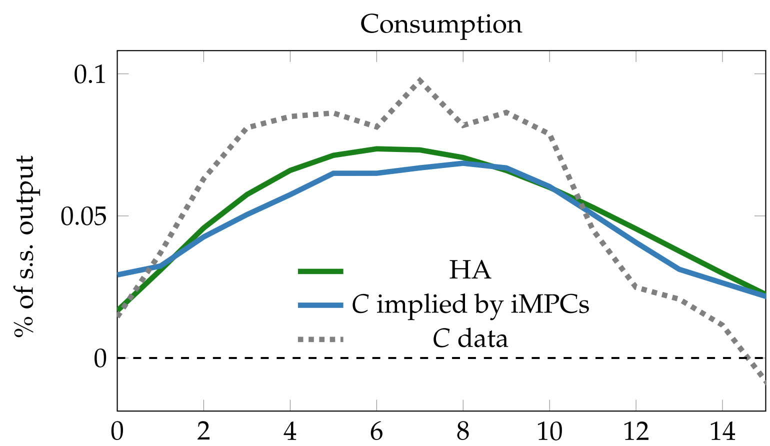
Parameter		Value	std. dev.
γ	Household habit parameter	0.884	(0.012)
ϕ	Investment adjustment cost parameter	13.150	(3.385)
ζ_p	Calvo price stickiness	0.898	(0.031)
ζ_w	Calvo wage stickiness	0.931	(0.024)
ρ^m	Taylor rule inertia	0.902	(0.008)
σ^m	Std. dev. of monetary shock	0.057	(0.005)

Figure E.1: Role of investment with estimated TA habit model



Note. This figure shows the general equilibrium paths of output and consumption in: our estimated HA model (green), an RA model with habits (red), and a TA model with habits (blue). Dashed lines correspond to an investment adjustment cost parameter $\phi = \infty$.

Figure E.2: Consumption implied by iMPCs and the output response to the monetary policy shock



Note. This figure shows the estimated model (green) and data (gray dashed) responses to the monetary policy, as well as the implied consumption response (blue) if agents were only to receive the income stream $(1-\alpha)Y_t^{data}$ where Y_t^{data} is the empirical impulse response to the monetary policy shock. This consumption response only depends on intertemporal MPCs.

E.2 iMPCs and the path of income

In figure E.2 we perform a simple experiment, which is independent of our supply-side calibration and depends only on the pattern of intertemporal MPCs. In this experiment, we suppose that households' aggregate before-tax labor income is given by $(1-\alpha)Y_t^{data}$, where Y_t^{data} is the empirical impulse response of output to the monetary shock, and then feed in this labor income shock—*and no other shocks*—to the full-attention household sector.

The blue line shows the resulting consumption impulse, which is already quite large, both relative to the estimated model consumption response (green) as well as the empirical consumption response (gray, dashed). There is no room for intertemporal substitution to add to consumption in the first few quarters: the entire impulse is explained by the consumption response to labor income alone. Some friction, therefore, must be dampening the overall consumption response, especially on impact. In our model, this friction is inattention.

F Appendix to section 6

F.1 Estimated Habit-RA model

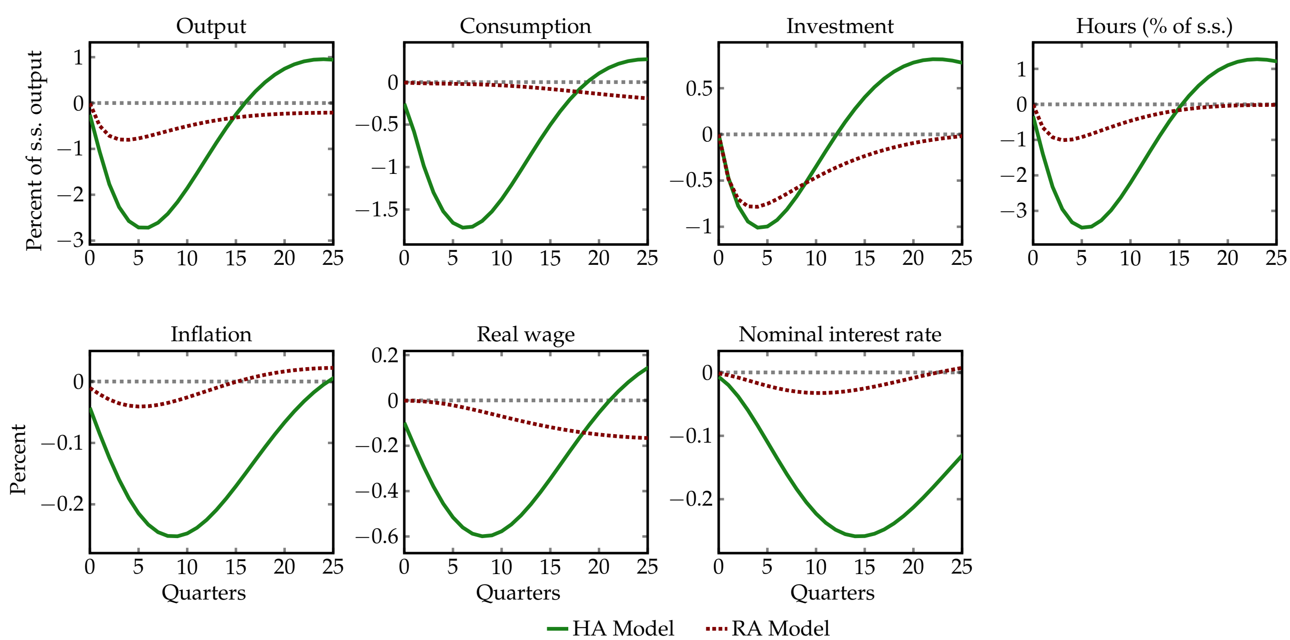
Table F.1: Priors and posteriors for the representative-agent model

Supply shock		Prior distribution	Posterior Mode	Posterior std. dev	Demand shock		Prior distribution	Posterior Mode	Posterior std. dev
TFP Θ_t	s.d.	Invgamma(0.1, 2)	0.330	(0.016)	Mon. policy ϵ_t^m	s.d.	Invgamma(0.1, 2)	0.215	(0.010)
	AR	Beta(0.5, 0.2)	0.970	(0.015)		AR	Beta(0.5, 0.2)	0.139	(0.051)
w markup $\epsilon_{w,t}$	s.d.	Invgamma(0.1, 2)	0.415	(0.028)	G shock G_t	s.d.	Invgamma(0.1, 2)	0.313	(0.015)
	AR	Beta(0.5, 0.2)	0.690	(0.132)		AR	Beta(0.5, 0.2)	0.884	(0.031)
	MA	Beta(0.5, 0.2)	0.647	(0.155)	C shock ϵ_t^C	s.d.	Invgamma(0.1, 2)	4.253	(1.067)
p markup $\epsilon_{p,t}$	s.d.	Invgamma(0.1, 2)	0.246	(0.016)		AR	Beta(0.5, 0.2)	0.759	(0.040)
	AR	Beta(0.5, 0.2)	0.266	(0.129)	I shock ϵ_t^I	s.d.	Invgamma(0.1, 2)	31.746	(7.832)
	MA	Beta(0.5, 0.2)	0.393	(0.095)		AR	Beta(0.5, 0.2)	0.528	(0.044)

Note. For an ARMA(1,1) process of the form $x_{t+1} - \rho x_t = \epsilon_{t+1} - \theta \epsilon_t$, "AR" refers to ρ, "MA" refers to θ. To be comparable with Smets and Wouters (2007) we scale the markup shocks such that $\epsilon_{w,t}, \epsilon_{p,t}$ appear with a coefficient of 1 in the Phillips curves (18) and (13).

F.2 Impulse response functions for HA and RA

Figure F.1: Impulse responses to 1-sd investment shock



Note. This is a positive shock to risk premia ϵ_t^I.

Figure F.2: Impulse responses to 1-sd consumption shock

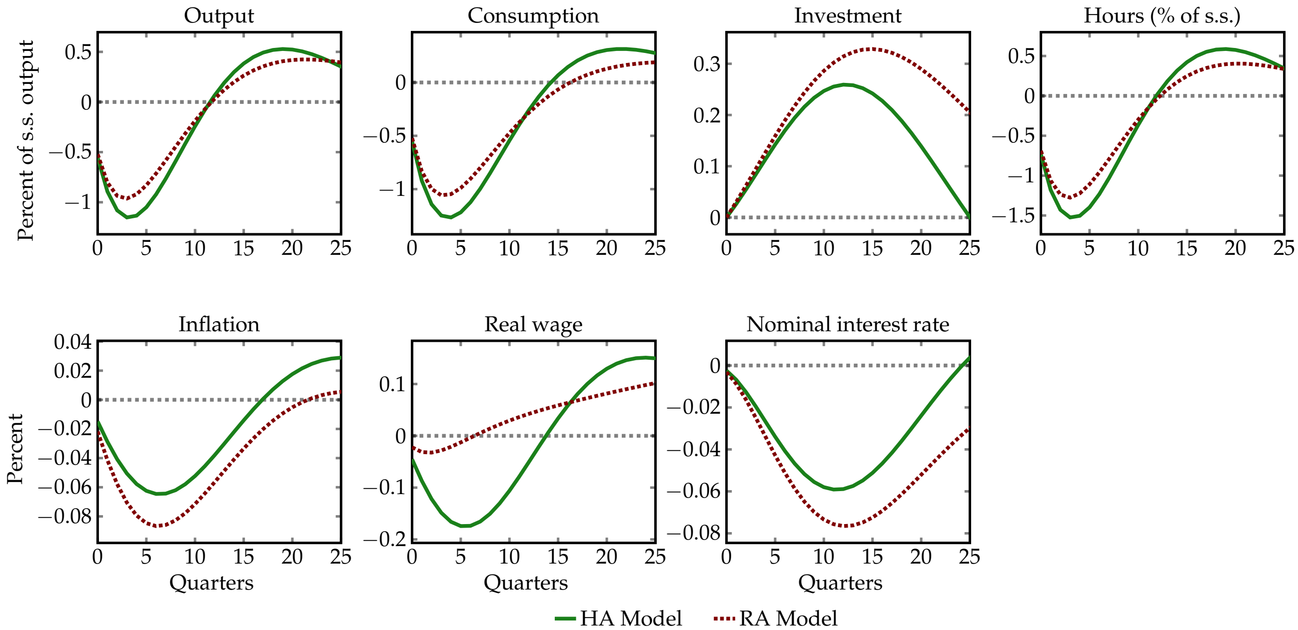


Figure F.3: Impulse responses to 1-sd government spending shock

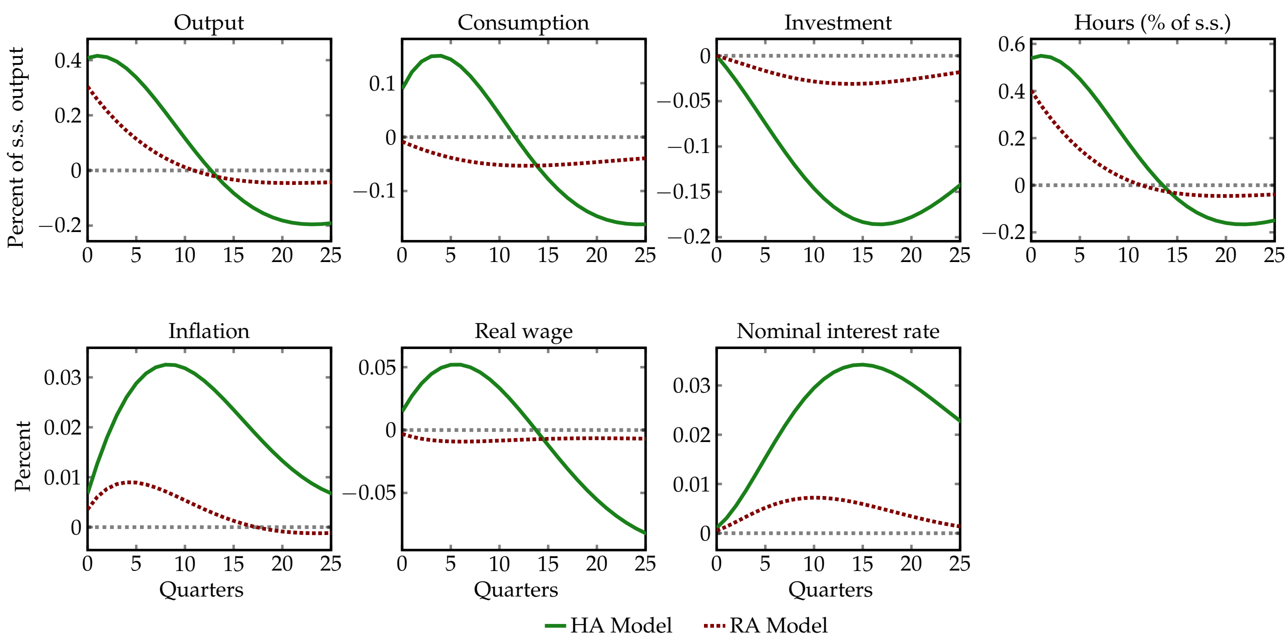
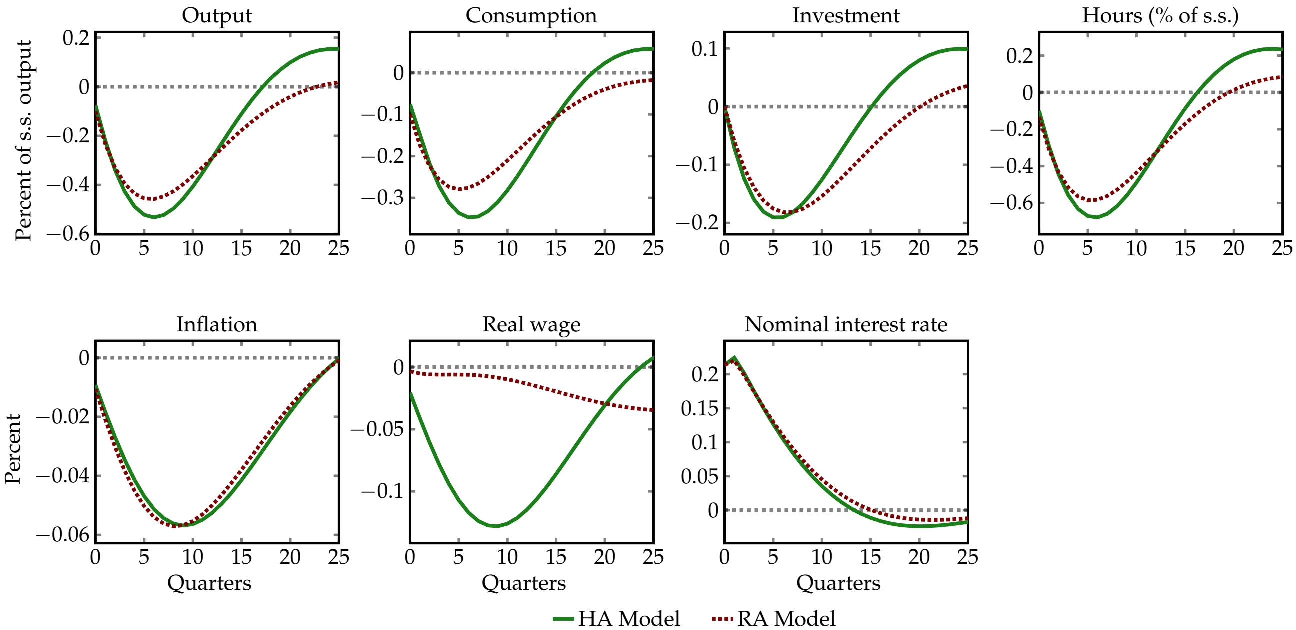


Figure F.4: Impulse responses to 1-sd monetary policy shock



Note. This is a positive shock to nominal rates ϵ_t^m.

Figure F.5: Impulse responses to 1-sd wage markup shock

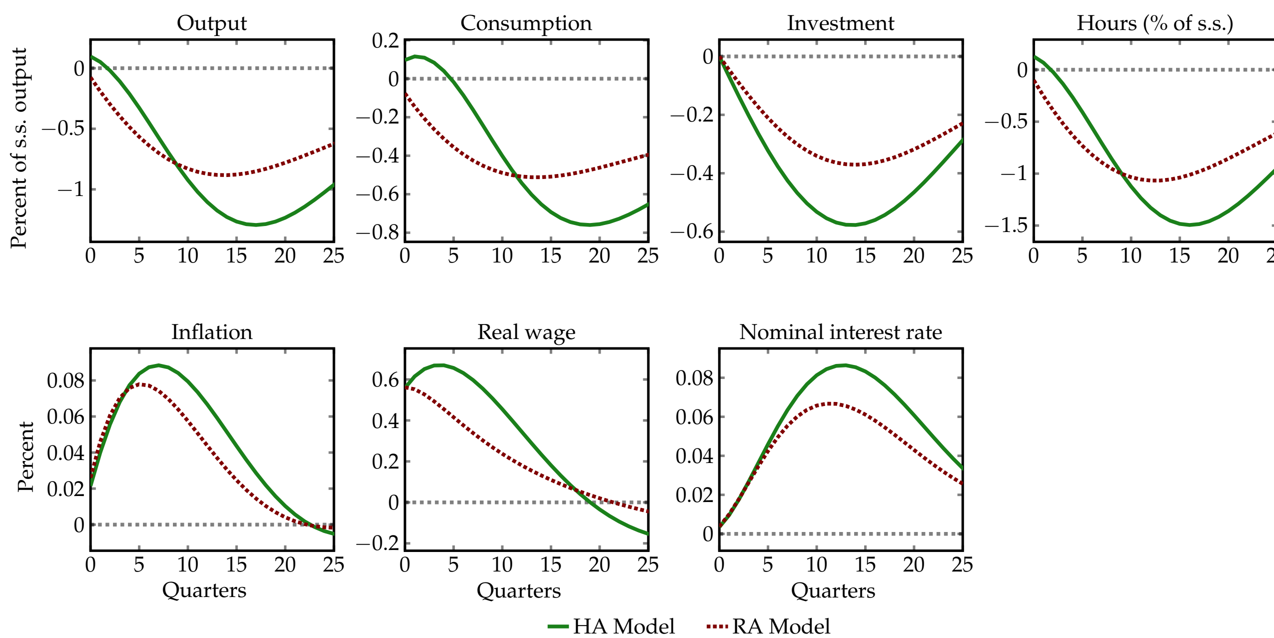


Figure F.6: Impulse responses to 1-sd productivity shock

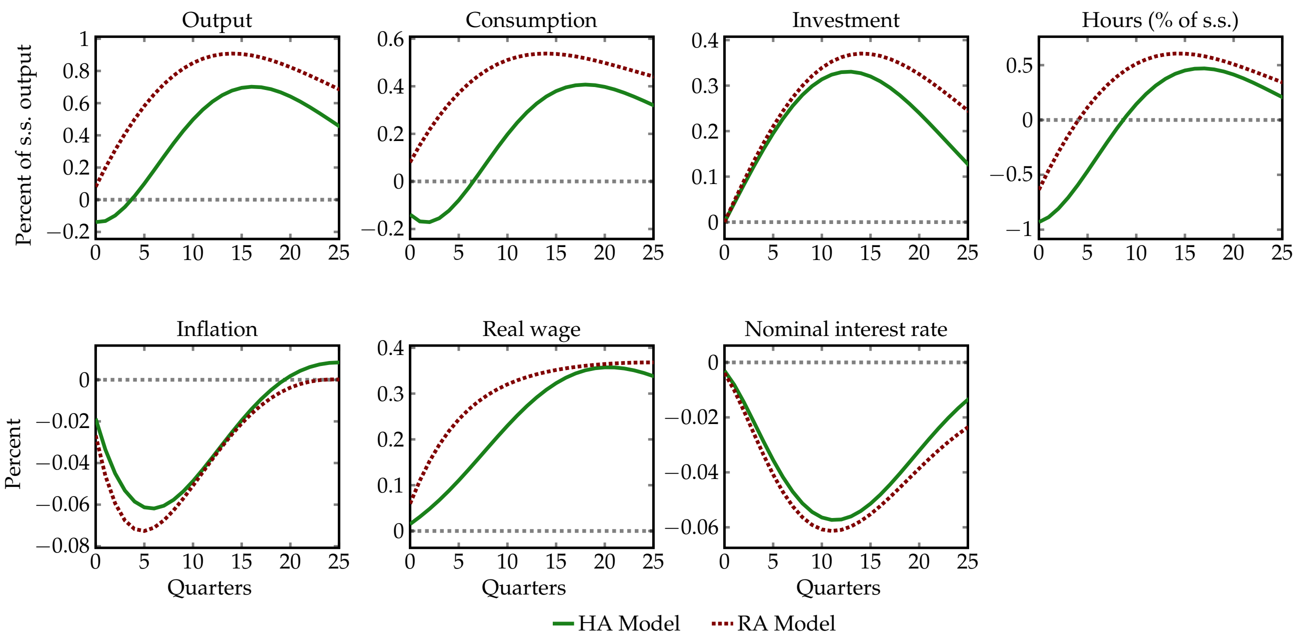
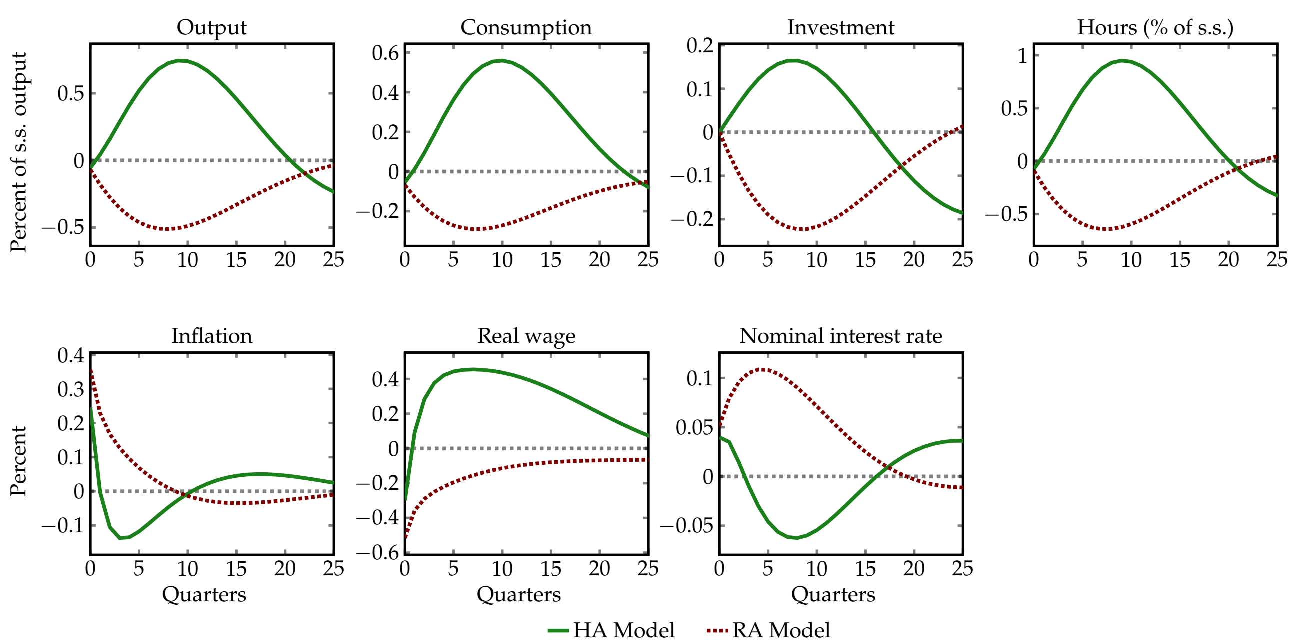
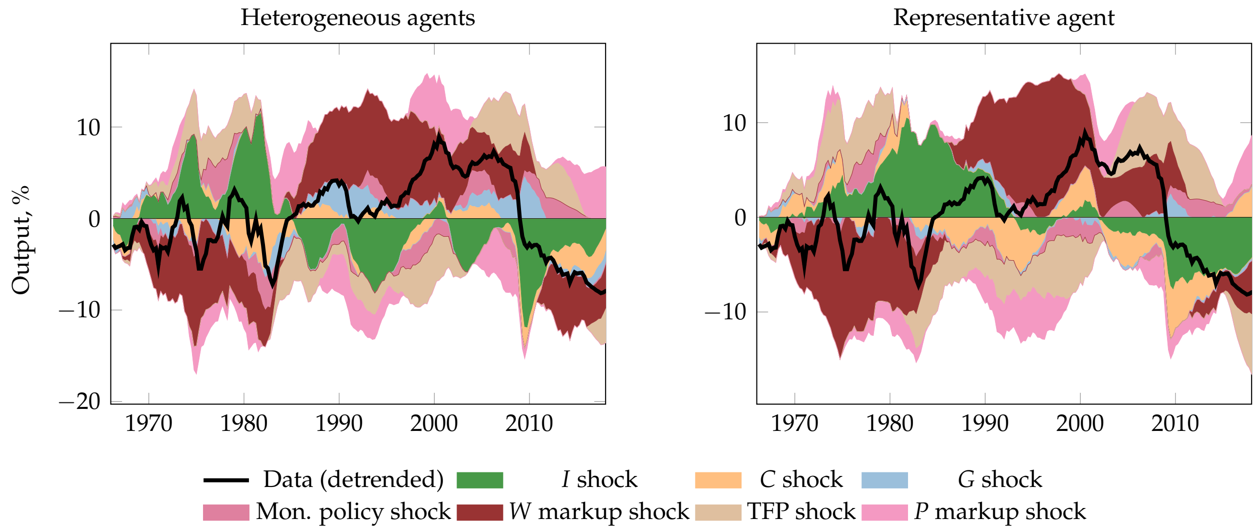


Figure F.7: Impulse responses to 1-sd price markup shock



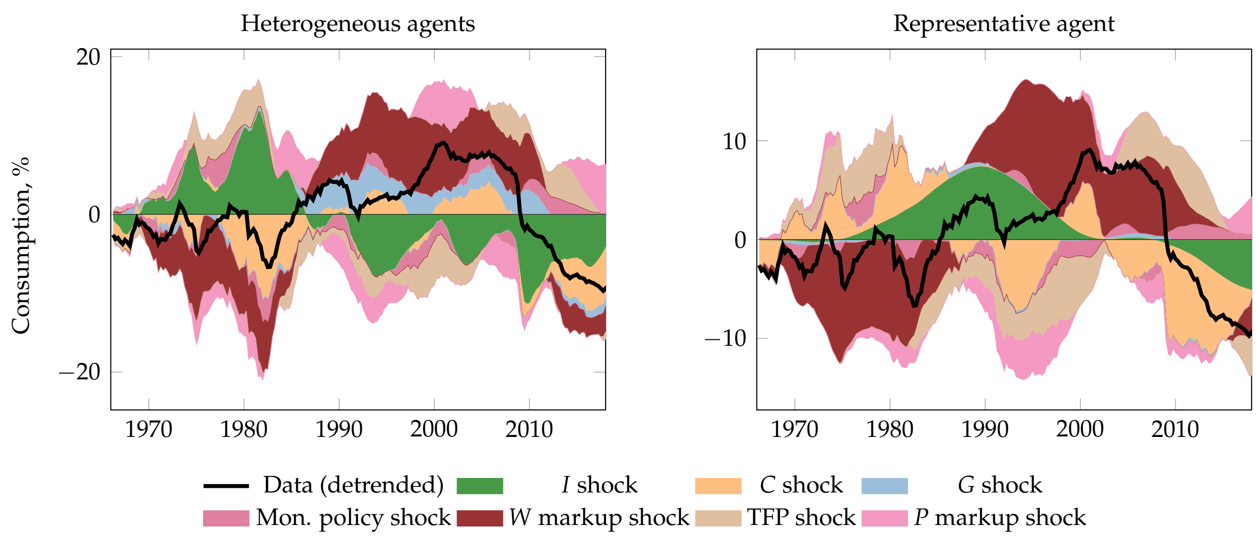
F.3 Historical shock decompositions of output and consumption

Figure F.8: Shock decompositions for output



Note. This figure decomposes the observed (linearly detrended real) output path Y_t into components driven by the seven shocks in the RA and HA models.

Figure F.9: Shock decompositions for consumption



Note. This figure decomposes the observed (linearly detrended real) consumption path C_t into components driven by the seven shocks in the RA and HA models.