Micro Jumps, Macro Humps:
Monetary Policy and Business Cycles in an Estimated HANK model

Adrien Auclert* Matthew Rognlie† Ludwig Straub‡

August 2019
Preliminary and Incomplete

Abstract

We estimate a Heterogeneous-Agent New Keynesian model that matches existing microeconomic evidence on marginal propensities to consume (“micro jumps”) and macroeconomic evidence on the impulse response to a monetary policy shock (“macro humps”). We rule out habit formation as an explanation for the hump shape of output, but show that sticky information in the sense of Mankiw and Reis (2002) can rationalize both the micro and the macro data. Our estimated model implies a central role for investment in the monetary transmission mechanism.

*Stanford University and NBER. Email: aauclert@stanford.edu.
†Northwestern University and NBER. Email: matthew.rognlie@northwestern.edu.
‡Harvard University. Email: ludwigstraub@fas.harvard.edu.
This research is supported by the National Science Foundation grant SES-1851717.
1 Introduction

A large empirical literature documents persistent and hump-shaped responses of economic activity to changes in monetary policy. The dynamics of macroeconomic variables in response to identified monetary policy shocks has inspired a large theoretical literature that, starting with Rotemberg and Woodford (1995) and Christiano, Eichenbaum and Evans (2005), has estimated representative-agent dynamic stochastic models that are able to replicate these impulse responses. The powerful argument made in this literature is that matching the impulse response is a prerequisite for conducting policy counterfactuals. Habit formation in consumption has been a key building block for this literature, since it provides a natural reason for the hump shape in the path of consumption—a key component of the monetary policy response.

A more recent literature has recognized that moving beyond aggregate outcomes is important to understand monetary policy. This literature has established that marginal propensities to consume (MPCs) and their dynamics shape the monetary transmission mechanism (see, for example, Auclert 2019, Kaplan, Moll and Violante 2018, and Auclert, Rognlie and Straub 2018). Since estimated representative-agent models are inconsistent with the MPCs observed in the data, their usefulness for policy counterfactuals may be limited after all. However, existing heterogeneous-agent models cannot be a substitute, since they are inconsistent with the empirical monetary policy impulse response—in particular, these models imply that the response to monetary shocks peaks on impact, which is clearly inconsistent with the long lags observed in the data. These models are also typically difficult and slow to use, which has prevented their estimation as well as their widespread adoption in policy institutions.\footnote{See Challe, Matheron, Ragot and Rubio-Ramirez (2017) for an early attempt at estimating a heterogeneous-agent model, but with restricted heterogeneity, and Hagedorn, Manovskii and Mitman (2018) for a monetary policy impulse-matching exercise in a heterogeneous agent model.}

In this paper, we take the natural next step. We propose and estimate a heterogeneous-agent New Keynesian model that matches both existing microeconomic evidence on marginal propensities to consume and macroeconomic evidence on the impulse response to a monetary policy shock. This unique combination makes our model suitable for policy counterfactuals. Our estimation procedure is also extremely flexible and fast to execute, so that variants of our model could easily be adapted for use in policy institutions.

We begin by showing that habit formation at the aggregate level, while consistent with the macro hump shape, is inconsistent with known patterns of MPCs and other microeconomic statistics. Hence, the reason for the hump shape of aggregate outcomes must lie elsewhere. We follow the lead of Gabaix and Laibson (2001), Mankiw and Reis (2002) and Carroll, Crawley, Slacalek, Tokuoka and White (2018), and propose inattention as an alternative explanation. We model inattention a la Calvo: following a shock, households only learn about the future path of macroeconomic aggregates with probability $1 - \theta$, independent across time and households. Households who have not learned about the new path continue to behave as if the future was the previous steady state, while still respecting their budget constraints. As a result, they initially do not fully
anticipate the exact future paths of interest rates and output. Thus, while inattention does not alter households’ responses to unanticipated income shocks—allowing the model to be consistent with micro evidence on MPCs—it introduces sluggishness at the aggregate level—allowing the model to be consistent with macro evidence on hump-shaped consumption responses.

These results suggest that inattention is a natural starting point to estimate heterogeneous-agent models. We turn to such a model next. The key features of our model are nominal price and wage rigidities, liquid and illiquid assets for households, as well as capital adjustment costs and inattention on the part of households and firms. The combination of capital adjustment costs and inattention enables our model to also generate sluggish behavior of aggregate investment, without resorting to investment adjustment costs as in Christiano et al. (2005) (see also Zorn 2018).

We estimate our model to match the impulse responses to an identified monetary policy shock. For our baseline exercise, we identify monetary policy shocks using Jordà (2005) projections with Romer and Romer (2004) shocks as instruments in their original sample. This combination was proposed by Ramey (2016). In line with her findings, and the general consensus in the vast literature estimating responses to monetary policy shocks, our impulse responses suggest hump-shaped responses of output, consumption, investment and hours, and a delayed response of inflation.

We match the evidence with our heterogeneous-agent model by following a two-step procedure. In a first step, we calibrate parameters that are relevant for the steady state by following the standard procedure in the heterogeneous-agent literature. These parameters include the steady state real interest rate, the depreciation rate, the process of idiosyncratic productivity shocks, risk aversion, the discount factor, and the share of liquid assets. Crucially, we choose the latter two to match the aggregate wealth-to-income ratio and an annual MPC of 0.50, in line with evidence from a vast literature documenting sizable MPCs.

In a second step, we estimate parameters that do not affect the steady state but affect the transition dynamics after shocks to match the monetary policy impulse response. These parameters include the inattention parameters for households and firms, the slopes of the price and wage Phillips curves, and the magnitude and persistence of monetary policy shocks. We find the parameters that minimize the squared distance between our impulse responses and the data impulse responses, with weights equal to the inverse sample variances of the empirical impulse responses (as in, e.g., Christiano et al. 2005).

Our model matches the empirical impulse responses well. Output, hours, and consumption all display a hump shape and peak around quarter 6. The price level initially responds very slowly before picking up later on, implying a hump-shaped path for inflation as well. To the best of our knowledge, ours is the first heterogeneous-agent model that replicates the impulse response to a monetary policy shock, while featuring (intertemporal) MPCs that are as high as they are in the data. Inattention is chiefly responsible for this success. To give a sense of magnitudes, our estimate of household inattention suggests that about half of all households become attentive by quarter 5.

To illustrate how inattention causes a hump-shaped response in consumption, we split the
general equilibrium consumption response into two separate parts: a direct effect of real interest rates on consumption, driven by income and substitution effects; and an indirect effect of greater after-tax labor incomes and transfers from the illiquid account, driven by MPCs. Since agents observe their current income, inattention does not dampen MPCs, and thus barely weakens the strength of indirect effects. However, it does significantly dampen the strength of direct effects. In that sense, our estimated degree of inattention informs our model of the relative magnitudes of direct and indirect effects.

We use our estimated model to revisit the monetary policy transmission mechanism. We find that investment plays an outsized role. If we shut off the response of investment by taking the limit of infinite adjustment costs—not affecting the steady state, only the transition—we find that the cumulative impulse response of output over the first 20 quarters shrinks by about 90%. That this reduction is far larger than merely the direct contribution of investment is evidenced by the fact that the cumulative impulse response of consumption also shrinks significantly—by about 85%. Thus, investment matters not just for output directly, but also indirectly through consumption.

This stark finding, which is new to the literature on monetary policy, is an implication of the fact that we jointly match the impulse responses of aggregate variables as well as sizable MPCs. Because we match the impulse responses, our estimated parameters, specifically the degree of firm inattention, generates the correct overall size of investment. Because we match MPCs and the hump-shape of consumption, our estimated level of inattention suggests that households are not very responsive to changes in interest rates; rather, they are spending mostly out of higher incomes, which are partly driven by investment. This generates an amplification mechanism, whereby the initial response of investment increases aggregate incomes, therefore stimulating greater household spending, which feeds back into incomes. This way, investment takes up the central role in the monetary transmission mechanism.

We show that this finding has two important practical implications. First, it is often argued that limited pass-through of the federal funds rate to consumer interest rates in the proximity of the zero lower bound limits the potency of monetary policy. Our findings suggest that this type of pass-through is not critical, even for the effects of monetary policy on consumption, as consumption is mostly indirectly driven by investment. Second, to the extent that capital investment is at least partly irreversible, our findings imply a strong state dependence of monetary policy based on the amount of capital that was accumulated. If a recession was preceded by an investment boom, investment will not be very responsive to monetary policy accommodations, limiting the aggregate impact of a given monetary stimulus.

We make use of our estimated model to speak to several open debates in monetary economics. First, we investigate the importance of interactions between monetary and fiscal policy, which Kaplan et al. (2018) have moved to center stage. We find that, when government debt is long-term—the assumption we make in our baseline model—income effects on the government balance

\footnote{See Auclert et al. (2018) for related amplification mechanisms in response to government spending and deleveraging shocks.}
Sheet from movements in real rates are much more muted than when government debt is short-term. Still, we confirm the message in Kaplan et al. (2018): fiscal rules adjusting lump-sum taxes running balanced-budget policies contribute to a greater output sensitivity to interest rates.

Second, we revisit the effects of forward guidance in our model. While we find that the infamous “forward guidance puzzle” (Del Negro, Giannoni and Patterson 2013, McKay, Nakamura and Steinsson 2016) is worsened in heterogeneous-agent models once investment is introduced, our addition of inattention resolves this issue and mitigates the puzzle (see also Gabaix 2016 and Angeletos and Lian 2018).

Literature. Our paper rests on the shoulders of three giant literatures.

First and foremost, it relates to a recent and very active literature studying heterogeneous-agent models with nominal rigidities. In this literature, our paper is closest to the work by Gormann, Kuester and Nakajima (2012), McKay et al. (2016), Werning (2015), Kaplan et al. (2018) and Auclert (2019), in that we also focus on monetary policy, and the central role played by heterogeneity and MPCs. Different from these papers, we estimate our model to also match the dynamics of aggregates and use our results to shed light on the crucial interaction between MPCs at the micro level and investment at the macro level. This interaction is related, at an abstract level, to the interaction between MPCs and deficit-financed fiscal policies, which we documented in Auclert et al. (2018).

Second, our paper builds on the voluminous literature estimating representative-agent models (without inattention). In this literature, our paper is most closely related to the work using limited information estimation techniques to estimate models, e.g. Rotemberg and Woodford (1997), Christiano et al. (2005) and Altig, Christiano, Eichenbaum and Lindé (2011). The contribution of our work is to apply those methods to models with heterogeneous agents, and to reveal new insights about the transmission mechanism.

Third, our paper benefited from the literature on bounded rationality in macroeconomics. Our approach of building on inattention to match aggregate impulse responses has its roots in Mankiw and Reis (2002, 2007) and Maćkowiak and Wiederholt (2009, 2015), and was more recently used in Gabaix (2016) and Angeletos and Lian (2018) to solve the forward guidance puzzle (among other things). Common across these papers is not to rely on some of the typical ingredients being used to match aggregate dynamics, e.g. habits. Carroll et al. (2018) present a powerful case using an inattentive heterogeneous-agent model in partial equilibrium to argue that inattention is far more compatible with micro-level consumption dynamics than habits are. Due to its tractability, Carroll et al. (2018), like us, relies on the simple sticky information approach pioneered by Gabaix and Laibson (2001) and Mankiw and Reis (2002), and later included in a general equilibrium model in Mankiw and Reis (2007), though without investment.

---

3 Oh and Reis (2012), McKay and Reis (2016), Hagedorn, Manovskii and Mitman (2019) also study fiscal policy in heterogeneous-agent models with nominal rigidities.

4 See Havranek, Rusnak and Sokolova (2017) for a survey of empirical estimates of habits.
In our model, inattention not only helps match the sluggish behavior of aggregate consumption, but also that of aggregate investment to aggregate shocks. Zorn (2018) provides direct empirical evidence for the role of inattention in causing that sluggishness and builds a model with rational inattention to rationalize this fact.

To the best of our knowledge, the only other paper that, like ours, combines bounded rationality, heterogeneity and monetary policy is Farhi and Werning (2017), though they do so without investment. They use this setup to argue that this combination can resolve the forward guidance puzzle, which is still present in benchmark heterogeneous-agent models (Werning 2015).

**Layout.** The rest of the paper proceeds as follows. In section 2 we make the case for introducing inattention in heterogeneous-agent models. This sets us up to introduce our general equilibrium model in section 3, which we estimate in section 4. We discuss the key role of investment for the monetary transmission mechanism in section 5. We revisit current debates in monetary economics using our estimated model in section 6, before presenting a brief conclusion in section 7.

### 2 Inattention, not habits

A large literature in monetary economics relies on habit formation to make models consistent with the hump-shaped response of aggregate consumption to monetary policy shocks. In this section, we demonstrate that, while habits succeed in generating such responses, they do so for reasons that are counterintuitive and at odds with micro data patterns. We introduce inattention as a viable alternative and highlight why, in contrast with habits, it generates hump-shaped responses in a way that is both intuitive and in line with key moments of consumption behavior.\(^\text{5}\) Throughout the paper, we focus on sticky information as a form of inattention, as in Gabaix and Laibson (2001), Mankiw and Reis (2002), and Reis (2006), among others.

#### 2.1 The representative agent model with habits

We begin by studying habits in a standard deterministic representative agent model (henceforth habit-RA), as in Fuhrer (2000) and Christiano et al. (2005).\(^\text{6}\) The agent has an infinite horizon, earns a stream of labor income \(\{y_t\}\) and faces real interest rates \(\{r_t\}\). His value function at date \(t\) is given by

\[
V_t(a_{t-1}, c_{t-1}) = \max_{c_t, a_t} u(c_t - \gamma c_{t-1}) + \beta V_{t+1}(a_t, c_t)
\]

\[
c_t + a_t \leq (1 + r_t)a_{t-1} + y_t
\]

where \(\gamma \in [0, 1)\).

\(^\text{5}\)For a highly complementary comparison of habits and inattention, see Carroll et al. (2018).

\(^\text{6}\)We focus on internal habits. External habits have similar properties.
How does this model generate a slow response of consumption in response to a given path of incomes \( \{y_t\} \) and interest rates \( \{r_t\}\)? We answer this question by first studying three sets of moments characterizing optimal consumption behavior, which have not received much attention in the large literature on habits. Then, we argue that these moments jointly imply a slow response of consumption. We focus on moments defined around a steady state in which \( c, y, r \) are constant and \( \beta(1 + r) = 1 \).

**Fact 1.** Habits dampen responses to \( \{r_t\} \), but less so at long horizons.

The left panel in figure 1 shows the initial consumption response to interest rates at various horizons, \( (1 + r)^t \partial c_0 / \partial r_t \). The plot makes apparent that, compared to a no-habit economy (\( \gamma = 0 \); black), habits dampen the initial consumption response to interest rates at all horizons (\( \gamma > 0 \); red), but less so at long horizons.

**Fact 2.** Habits dampen responses to \( \{y_t\} \), equally at all horizons.

The middle panel in figure 1 shows the initial consumption response to income \( (1 + r)^t \partial c_0 / \partial y_t \) as the horizon \( t \) varies. As one can easily verify using pen and paper (see appendix B.1), this derivative is given by \( (1 - \beta)(1 - \beta \gamma) \) and is therefore constant as a function of \( t \). It also falls with \( \gamma \): larger habit persistence implies lower MPCs.

**Fact 3.** Habits imply intertemporal MPCs that increase over time.

The right panel in figure 1 shows the dynamic consumption response to a current income shock \( \partial c_t / \partial y_0 \), an important set of intertemporal MPCs.\(^8\) As is well known, without habits, \( \gamma = 0 \), that is,

---

\(^7\)The factor \( (1 + r)^t \) normalizes the response to present value terms.

\(^8\)Auclert et al. (2018) introduced intertemporal MPCs and showed that they serve as useful moments for discriminating between models of consumption.
the dynamic consumption response is flat, as per the permanent income hypothesis. Instead, with positive habits, $\gamma > 0$, the consumption response builds up over time, initially starting at an (even) lower level than the no-habit response.

How do facts 1–3 cause a hump-shaped response of consumption? This can be seen by imagining that $\{y_t\}$ and $\{r_t\}$ correspond to the empirical responses to a monetary policy shock. As is well known, a model without habits implies that, in response to such income and interest rate paths, consumption should decline over time. Allowing for facts 1–3 dampens the first few consumption observations considerably, thus generating a hump-shaped response.

Discussion. In sum, the habit-RA model implies consumption behavior characterized by facts 1–3 which allow it to generate a hump-shaped consumption profile. How reasonable is this behavior?

Fact 1 may seem very counterintuitive, and is in fact the key subject in a large debate on the “forward guidance puzzle”, according to which the model without habit already implies far too large of an effect of future rates compared to current rates (see for example Del Negro et al. 2013, McKay et al. 2016, and many others). This is exacerbated by habits.

Fact 2 is problematic for two reasons. First, typical macroeconomic models already have trouble matching realistically high marginal propensities to consume (MPCs) $\partial c_0 / \partial y_0$. Because they lower MPCs, habits make that problem much worse. Second, standard models of consumption imply too much of an anticipatory effect, compared to the evidence from the recent work of Fuster, Kaplan and Zafar (2018) and Baugh, Ben-David, Park and Parker (2018) (see also Jappelli and Pistaferri 2010). The habit-RA model is also subject to that critique, since responses to anticipated income shocks are equally strong as responses to unanticipated income shocks.

Fact 3 is clearly at odds with the much higher, downward sloping intertemporal MPCs in the data. The RA model without habits is already rejected, but an increasing profile of makes the habit-RA model look even more at odds with the data (see the evidence in Fagereng, Holm and Natvik 2018 and the discussion in Auclert et al. 2018).

We next introduce a version of the RA model with inattention, arguing that it changes things to the better.

2.2 The inattentive representative-agent model

Our model of inattention is based on the idea that agents are not always paying attention to the paths of aggregate variables $\{y_t, r_t\}$ and only infrequently update their information sets. In particular, we assume that agents are either attentive or inattentive about aggregates. All agents are initially inattentive and, as such, behave as if all future aggregate variables are still at their steady state values. As in Gabaix and Laibson (2001) and Mankiw and Reis (2002), they face a constant probability $\theta$ of becoming attentive each period. Attentive agents know the entire future paths of aggregates. Thus, all agents eventually become attentive to aggregate economic developments, but it takes some of them a while.
Since attentive households have knowledge of the full equilibrium path of \( \{ y_t, r_t \} \), they solve the standard Bellman equation

\[
V^A_t (a_{t-1}) = \max_{c_t, a_t} u(c_t) + \beta V^A_{t+1} (a_t) \\
c_t + a_t \leq (1 + r_t)a_{t-1} + y_t
\]  

(1a)

By contrast, inattentive households make decisions believing that future aggregate variables are at their steady state values

\[
V^I_t (a_{t-1}) = \max_{c_t, a_t} u(c_t) + \beta V^{ss} (a_t) \\
c_t + a_t \leq (1 + r_t)a_{t-1} + y_t
\]  

(1b)

where \( V^{ss} \) is the steady state value function that results from setting \( y_t = y_{ss} \) and \( y_t = r_{ss} \) at all times when solving (1a).

There are many existing formulations of a problem of inattentive consumers in the literature (see Reis 2006, Mankiw and Reis 2006, Gabaix 2013, Carroll et al. 2018.) All share a common objective of relaxing rational expectations in favor of slow adjustment of information. Our formulation is based on the “sticky information” models of Gabaix and Laibson (2001), Mankiw and Reis (2002), and Mankiw and Reis (2007), and therefore has two useful properties. First, uninformed households correctly perceive their current flow budget constraint at every point in time and thus never violate it. This is particularly important in models with borrowing constraints, which we use below. Second, the approach is highly tractable, and remains so even if idiosyncratic uncertainty is introduced to the model as is done in the next section.

To illustrate how the inattentive-RA model works, we solve for this model explicitly in appendix A.1, when the real interest rate \( r \) is constant but income \( \{ y_t \} \) varies. We summarize the main properties of the model here. How does the inattentive RA model compare with the habit-RA model, both in the key moments it generates as well as its ability to generate a sluggish, hump-shaped consumption response? We reconsider the three facts from before.

**Fact 4.** Inattention dampens responses to \( \{ r_t \} \), equally so at all horizons.

We illustrate this fact in the left panel of figure 1 (orange line). The initial consumption response to \( r_t \) is scaled down by \( \theta \) for any horizon \( t \), as only a fraction \( \theta \) of agents is attentive to changes in current or future interest rates.

**Fact 5.** Inattention dampens responses to \( \{ y_t \} \) only at long horizons (i.e. anticipated responses).

We illustrate this fact in the middle panel of figure 1 (orange line). The initial consumption response to \( y_t \) is only dampened (again by \( \theta \)) if \( t > 0 \), that is, for future changes in income. This is because while inattentive agents respond to current income changes, they are still unaware of changes to future incomes.

**Fact 6.** Inattention does not affect intertemporal MPCs.
Figure 2: Three important consumption moments in a heterogeneous-agent economy.

Note. The left panel shows the response of initial consumption to interest rates, $\frac{\partial c_0}{\partial r_t}$; the middle panel shows the response of initial consumption to income, $\beta^{-1}\frac{\partial c_0}{\partial y_t}$; the right panel shows the dynamic response of consumption to a current income shock $\frac{\partial c_t}{\partial y_0}$. The black line is an HA economy without habit; the blue line an HA economy with habit; the green line an HA economy with inattention.

We illustrate this fact in the right panel of figure 1 (orange line). Period-0 income shocks are immediately observed by attentive and inattentive agents alike, implying that inattention does not change the pattern of intertemporal MPCs.

Discussion. While giving rise to a similar overall consumption behavior, facts 4–6 paint a picture of the underlying moments that is very different from facts 1–3. Inattention avoids an exacerbated “forward guidance puzzle”, since it dampens $r_t$ responses equally at all horizons (fact 4), it does not dampen MPCs but instead dampens anticipated responses (fact 5), and it does not imply counterfactually increasing intertemporal MPCs (fact 6).

In sum, the inattentive RA model can generate consumption responses that, overall, look similar to those generated by a habit-RA model. Compared to adding habits, however, adding inattention does not make the RA model’s micro moments less consistent with micro data. We now turn to a heterogeneous-agent model and argue that the exact same lessons apply.

2.3 The inattentive heterogeneous-agent model

The inattentive heterogeneous-agent (HA) model is like an inattentive RA model, except that agents are also subject to idiosyncratic income risk, which we assume that agents are fully informed about.\textsuperscript{9} We model income risk by as a state-dependent labor productivity $e(s)$ where the income state $s$ evolves according to a Markov chain. Allowing for state-dependent value func-

\textsuperscript{9}Given that idiosyncratic uncertainty at the individual level is much larger than at the aggregate level, for any individual agent the cost of aggregate mistakes is never large: see, for example, Carroll et al. (2018).
Figure 3: Intertemporal MPCs in the data and in four models

Note. This figure shows the dynamic consumption responses to an unanticipated income shock $\frac{\partial c_t}{\partial y_0}$, i.e. intertemporal MPCs, across four models vs. the data. The data was obtained by smoothly interpolating the cumulative spending response in Fagereng et al. (2018) from their annual to the quarterly level.

The system of Bellman equations is then

$$V^A_t(a_{t-1};s_t) = \max_{c_t,a_t} u(c_t) + \beta E \left[ V^A_{t+1}(a_t;s_{t+1}) | s_t \right]$$

$$V^I_t(a_{t-1};s_t) = \max_{c_t,a_t} u(c_t) + \beta E \left[ V^{ss}_{t}(a_t;s_{t+1}) | s_t \right]$$

where both are subject to the flow budget constraint

$$c_t + a_t \leq (1 + r_t)a_{t-1} + y_t e(s_t)$$

and the borrowing constraint $a_t \geq 0$. How does the addition of inattention change the micro behavior of the HA model? As it turns out, facts 4–6 still hold true in this context: inattention dampens initial consumption responses to interest rate changes at all horizons, dampens anticipatory spending responses to future income changes, and leaves intertemporal MPCs unchanged. We illustrate all three in figure 2, which also shows the respective responses in a HA model with habit (for details, see Appendix B.2). As we discussed above, inattention changes the implications of the HA model in a realistic direction, supported by micro evidence, whereas habits do not. As a striking case in point, the right panel in figure 2 and figure 3 show that, even in a HA model, habits lead to initially upward-sloping intertemporal MPCs—a stark contrast to empirical evidence (Fagereng et al. 2018, Auclert et al. 2018).\textsuperscript{10}

\textsuperscript{10}Inattention also has a computational advantage, in that it does only requires an additional binary state variable (informed vs. uninformed) but not an additional continuous state variable, as is the case with habits (past consumption). We discuss this further in section 4.3.
3 A general equilibrium inattentive-HA model

We now embed the inattentive heterogeneous-agent household side introduced above in a general equilibrium model. The model will consist of three inattentive sectors: households, firms and unions. Section 2 explained why inattentive households avoids the need for habit formation. In a similar vein, we explain why inattention on the part of firms generates sluggishness in both investment and price inflation, avoiding the need for investment adjustment costs and price indexation, while inattention on the part of unions generates sluggishness in wage inflation, avoiding the need for wage indexation. These features of DSGE models are often criticized because they are inconsistent with micro data on price and capital adjustment (see e.g. Chari, Kehoe and McGrattan 2009). We find that one single friction—inattention—is able to generate the sluggishness that these three separate frictions are designed to mimic. \(^{11}\) After describing the inattentive sectors, we describe the financial sector and government policy.

3.1 Inattentive households

The economy is populated by a unit mass of ex-ante identical agents who face face idiosyncratic uncertainty, but no aggregate uncertainty. Households transition between ability states \(\{e_i\}\) according to a Markov process with fixed transition matrix \(\Pi\). The mass of worker type \(i\) in idiosyncratic state \(e_i\) is always equal to \(\pi(e_i)\), the probability of \(e_i\) in the stationary distribution of \(\Pi\). We normalize ability levels to average one, so \(\sum e_i \pi(e_i) = 1\).

Households have time-0 utility over consumption and labor supply plans given by separable preferences

\[
E \left[ \sum_{t \geq 0} \beta^t \{ u(c_{it}) - v(n_{it}) \} \right]
\]

Each period \(t\), an agent with realized earnings ability \(e_{it}\) enjoys the consumption of a generic consumption good \(c_{it}\) and gets disutility from working \(n_{it}\) hours. Because of sticky wages, hours \(n_{it}\) are determined by union labor demand. As described in section 3.3, unions allocate labor equally across all agents, so that for every agent \(i\), \(n_{it} = N_t\) where \(N_t\) is aggregate labor demand. Individual after-tax labor income in period \(t\) is therefore given by \(z_{it} \equiv Z_t e_{it}\), where \(Z_t \equiv (1 - \tau_t) W_t / P_t N_t\) is aggregate labor income.

Households have access to the simplified two-asset market structure described in Auclert et al. (2018). They can trade in a liquid asset \(a_{it}\) with return \(r_t\) subject to a zero-borrowing constraint, and also hold an entirely illiquid asset \(a_{illiq, ss}^{i}\), which we assume to be the same amount \(a_{illiq, ss}\) for each household in the steady state of the economy. Both assets are liabilities of a representative mutual fund, which is introduced in section 3.4 below. In the transition, illiquid asset positions may differ across households as households may become aware of valuation changes in their

\(^{11}\)For general equilibrium models with multiple sources of inattention, but without investment, see Mankiw and Reis (2007) (sticky information) and Maćkowiak and Wiederholt (2015) (rational inattention). For a model in which rational inattention generates a hump-shape in investment, see Zorn (2018).
illiquid account at different times. We assume that as long as agents are inattentive, they do not adjust their transfers from the illiquid account; attentive agents, by contrast, withdraw the annuity value of their illiquid account.\footnote{To ensure stationarity, we assume that agents withdraw the annuity value plus $\epsilon (a_{illiq}^{ss} - a_{illiq}^{III})$ for an arbitrarily small $\epsilon > 0$.}

Our approach to modeling the two-asset market structure allows our model to simultaneously match high average MPCs and a high level of aggregate wealth. In equilibrium, households therefore face the following set of constraints:

\begin{align}
    c_{it} + a_{it} &= Z_t e_{it} + (1 + r_t) a_{it-1} + d_{it}^m \quad (4) \\
    a_{it} &\geq 0 \quad (5)
\end{align}

Every period, households choose consumption $c_{it}$ and assets $a_{it}$ subject to (4) and (5) under inattention about aggregates. During the time they are inattentive, which continues with probability $\theta_h$ each period, they make decisions with full knowledge of $r_t$ and $Z_t$, and by withdrawing the steady state quantity from their illiquid account, $d_{it}^m = d^{ss}$. They face constraints (4)–(5) and assume that $r_{t+k}$, $Z_{t+k}$, and $d_{it+k}^m$ will be constant at their steady state values for every $k \geq 1$. Once they become attentive, they are fully aware of the full paths of $r_t$ and $Z_t$, choose a transfer $d_{it}^m$ equal to the annuity value of their current illiquid wealth $a_{illiq}^{III}$. The behavior of the model in response to changes in income and interest rates is thus as described in section 2.3.

### 3.2 Inattentive firms

We assume a standard two-tier production structure. Intermediate goods are produced by a mass one of identical monopolistically competitive firms, whose shares are traded and owned by the mutual fund. All firms have the same production technology, which we assume to be Cobb-Douglas in labor and capital

$$ y_t = F(k_{t-1}, n_t) = k_{t-1}^a n_t^{1-a} $$

Final goods firms aggregate intermediate goods with a constant elasticity of substitution $\mu / (\mu - 1) > 1$. Capital is subject to quadratic capital adjustment costs, so that the costs arising from choosing capital stocks $k_{t-1}$ and $k_t$ in any period $t$ are given by $\zeta \left( \frac{k_t}{k_{t-1}} \right) k_{t-1}$, with

$$ \zeta(x) \equiv x - (1 - \delta) + \frac{1}{2\delta \epsilon} (x - 1)^2 $$

where $\delta > 0$ denotes depreciation and $\epsilon > 0$ is the firm-level sensitivity of net investment to Tobin’s $Q$. We also assume that investment is irreversible,

$$ k_t \geq (1 - \delta) k_{t-1} \quad (6) $$
As we mostly focus on linearized behavior around the deterministic steady state, this assumption is only relevant in our analysis of the state-dependent effects of monetary policy.

Finally, any firm chooses a price $p_t$ in period $t$ subject to Rotemberg (1982) adjustment costs. We assume that at time $t$, firms choose their price for the subsequent period. This is not essential, but simplifies the analysis by treating prices and capital as choices with a one-period delay. The adjustment cost for prices takes the form

$$\xi(p_{t+1}, p_t) \equiv \frac{1}{2 \kappa^\eta (\mu-1)} \left( \frac{p_{t+1} - p_t}{p_t} \right)^2$$

where $\kappa^\eta > 0$. An attentive intermediate goods firm therefore maximizes its value

$$J_t(k_{t-1}, p_t) = \max_{p_t, k_{t-1}} \left\{ \frac{p_t}{P_t} F(k_{t-1}, n_t) - \frac{W_t}{P_t} n_t - \xi \left( \frac{k_t}{k_{t-1}} \right) k_{t-1} - \xi \left( p_{t+1}, p_t \right) Y_t + \frac{1}{1 + r_{t+1}} J_{t+1}(k_t, p_{t+1}) \right\}$$

subject to the requirement that it satisfies final goods firms’ demand in each period at its current price,

$$F(k_{t-1}, n_t) = Y_t \left( \frac{p_t}{P_t} \right)^{-\mu/(\mu-1)}$$

By contrast, an inattentive firm solves a problem analogous to (7)–(8) where $J_{t+1}(k_t, p_{t+1}) / (1 + r_{t+1})$ is replaced by $f^{ss}(k_t, p_{t+1}) / (1 + r^{ss})$. In other words, inattentive firm $j$ chooses its price $p_{jt}$, capital for next period $k_{jt}$ and labor $n_{jt}$ with full knowledge of $W_{t+k} / P_{t+k}, r_t, P_t$ and $Y_t$, but assuming that $P_{t+k}$, $r_{t+k}$, $P_{t+k}$ and $Y_{t+k}$ remain constant at their steady-state level for $k \geq 1$. Inattentive firms turn attentive with probability $1 - \theta^f$ each period.

In appendix A.2, we show that this setup delivers the following properties: inattentive firms choose a constant price $p_{t+1} = P_{ss}$ and a constant capital stock $k_t = k_{ss}$ for the following period and passively satisfy current demand by adjusting labor. Attentive firms set prices according to a dynamic equation similar to a Rotemberg price Phillips curve, and their investment decisions are governed by $q$-theory dynamics. In the aggregate, inattention therefore generates sluggishness in both the adjustment of the price level and in aggregate investment. Finally, given its policy, firm $j$ pays dividend

$$d_{jt} = \frac{p_{jt}}{P_t} F(k_{jt-1}, n_{jt}) - \frac{W_t}{P_t} n_{jt} - \xi \left( \frac{k_{jt}}{k_{jt-1}} \right) k_{jt-1} - \xi \left( p_{jt+1}, p_{jt} \right) Y_t$$

### 3.3 Inattentive unions

We follow standard practice in the New Keynesian sticky-wage literature and assume that labor hours $n_{ij}$ are determined by union labor demand (Erceg, Henderson and Levin 2000). Specifically, we assume that every worker $i$ provides $n_{ij}$ hours of work to each of a continuum of unions indexed by $j \in [0, 1]$, so that total labor effort for person $i$ is $n_{it} = \int_j n_{ij}dj$.

Each union $j$ aggregates efficient units of work into a union-specific task $N_{jt} = \int e_{it}n_{ij}di$. A competitive labor packer then packages these tasks into aggregate employment services using the

---

13The impulse responses we compute are linearized to first order in aggregates. Hence, Rotemberg adjustment costs are equivalent to price setting à la Calvo (1983).
constant elasticity of substitution technology

\[ N_t = \left( \int j N_{jt}^{\epsilon - 1} \, dj \right)^{\frac{1}{\epsilon - 1}} \]

and sells these services at price \( W_t \). Union \( j \) faces quadratic utility costs of adjusting its nominal wage \( W_{jt+1} \) for the following period, which we assume generates an extra additive disutility term

\[ \psi j \left( \frac{W_{jt+1}}{W_j} - 1 \right)^2 \, dj \]

in household utility (3). In every period, unions call upon its members to supply hours according to a uniform rule, \( n_{ijt} = N_{jt} \). The union sets \( W_{jt+1} \) to maximize the average utility of its members given this rule.

A union that has become attentive knows the full path of future employment \( N_t \), the distribution of consumption \( c_{it} \), as well as the marginal tax rate \( \tau_t \) which impacts the rewards to working at each point in time. An inattentive union assumes that those quantities remain at their steady state going forward. As we show in appendix A.3, attentive unions’s wage dynamics follows a path generated by a Rotemberg Phillips curve with slope \( \kappa w \). By contrast, inattentive unions do not adjust their wage until they become attentive. This generates sluggishness in the adjustment of aggregate wages.

3.4 Mutual fund

A representative mutual fund issues the liquid deposits \( a_t \) and the illiquid deposits \( a_{illiq}^m \). The mutual fund owns the intermediate goods firms and government bonds, which we model as long-term bonds with an exponentially decaying coupon. The mutual fund is perfectly attentive, that is, inattention does not interfere with market efficiency in our setup. It receives the dividends \( d_t = \int d_{jt} \, dj \) from all firms and deposits a transfer of \( d_{m}^m \) into every illiquid account each period. Denoting by \( v_{t-1} \) its holding of shares (equal to 1 in equilibrium), and \( b_{t-1} \) its holdings of government bonds, its flow of funds constraint is therefore:

\[ d_{illiq}^m + (1 + r_t) a_{illiq} + q_t b_t + p_t v_{t} = a_{illiq} + (1 + \delta q_t) b_{t-1} + (p_t + d_t) v_{t-1} \]

Agent \( i \)'s illiquid budget constraint therefore evolves as

\[ a_{illiq} + d_{illiq}^m \leq (1 + r_t) a_{illiq} + d_{illiq}^m \]

The mutual fund’s optimal allocation of assets equalizes the ex-post rates of return on bonds and equity, so that in every period \( t \geq 0 \),

\[ 1 + r_t = \frac{p_{t+1} + d_{t+1}}{p_t} = \frac{1 + \delta q_{t+1}}{q_t} \]

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), Gürkay-
nak, Sack and Swanson (2005), and Nakamura and Steinsson (2013) for monetary policy.

3.5 Government

The government issues long-term bonds $B^*_t$, collects taxes and spends on goods and services $G_t$. Its budget constraint is

$$q_t B^*_t + \tau_t \frac{W_t}{P_t} N_t = G_t + (1 + \delta q_t) B^*_t \frac{t-1}{1}$$

Given the lack of Ricardian equivalence in our model, the fiscal rule of the government is important. We assume that the government has the following “automatic stabilizer” rule for the tax rate

$$\tau_t - \tau_{ss} = \psi q_{ss} \frac{B^*_t \frac{t-1}{1} - B^*_t \frac{ss}{1}}{y_{ss}}$$

Equation (9) implies that the government raises the tax rate when debt is above its long-run level, slowly bringing the debt level back down.

We assume that monetary policy follows a rule for the nominal interest rate. In our benchmark, we assume that it sets the nominal interest rate such that the real interest rate is equal to some target path. In our estimation procedure we will take this path directly from the data.

Given these elements, the definition of equilibrium is standard.

4 Estimation

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 inattentive HA model, very similar to the one in Christiano et al. (2005). In the first step, we calibrate parameters that are relevant for the steady state of our model. In the second, we estimate the remaining parameters by matching impulse responses from an identified monetary policy shock.

First step: Calibration. Table 1 summarizes the steady state parameters. Households have constant CES utility over consumption $u(c) = \frac{c^{1-\nu} - 1}{1-\nu}$ with an EIS of $\nu = 1$, and a power disutility from labor $v(n) = v_0 n^{1+\phi^{-1}} / (1 + \phi^{-1})$ with Frisch elasticity $\phi = 0.5$. We set government spending to $G_Y = 20\%$ of output, and use $v_0$ to normalize steady-state output. We assume that steady-state inflation is $\pi = 0$ and that the annual steady state real interest rate is $r = 4\%$. We jointly choose $\beta$ and the fraction of illiquid bonds $B^{illiq} / B$ to match a 4-quarter cumulative MPC of 0.50 and an aggregate wealth to income ratio of 3.6, which is implied by the supply side of our model. This yields a ratio of liquid bonds to income of 0.21.

\[14\] This is equivalent to a Taylor rule with a coefficient of 1 on expected inflation and shocks to the intercept of the Taylor rule.
Table 1: Calibrated and estimated parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Panel A: Calibrated parameters</th>
<th>Panel B: Estimated parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$ EIS</td>
<td>1</td>
<td></td>
<td>$\theta^h$ Household inattention</td>
</tr>
<tr>
<td>$\phi$ Frisch</td>
<td>0.5</td>
<td>$\theta^f$ Firm inattention</td>
<td>0.956 (0.011)</td>
</tr>
<tr>
<td>$\rho_e$ Log $e$ persistence</td>
<td>0.91</td>
<td>$\theta^u$ Union inattention</td>
<td>0.000 —</td>
</tr>
<tr>
<td>$\sigma_e$ Log $e$ std. dev.</td>
<td>0.92</td>
<td>$\kappa^p$ Price Phillips Curve slope</td>
<td>0.071 (0.033)</td>
</tr>
<tr>
<td>$G/Y$ Spending-to-GDP</td>
<td>0.2</td>
<td>$\kappa^w$ Wage Phillips Curve slope</td>
<td>0.004 (0.001)</td>
</tr>
<tr>
<td>$B/Y$ Liquid assets to GDP</td>
<td>0.21</td>
<td>$\sigma_r$ Standard deviation of $r_t$</td>
<td>0.109 (0.013)</td>
</tr>
<tr>
<td>$\beta$ Discount factor</td>
<td>0.82</td>
<td>$\rho_r$ Persistence of $r_t$</td>
<td>0.820 (0.023)</td>
</tr>
<tr>
<td>$\delta$ Decay rate of government bonds</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the income process in our model, we follow standard practice in the literature and assume that gross income follows an AR(1) process. We use Floden and Lindé (2001)’s estimates of the persistence of the US wage process, equal to 0.91 yearly, set the variance of innovations to match the standard deviation of log gross earnings in the US of 0.92 as in Auclert and Rognlie (2018), and discretize this process as a 9-point Markov chain. Following McKay et al. (2016), we also assume that households cannot borrow, $\bar{a} = 0$.

**Second step: Estimation.** We estimate the remaining parameters by matching the impulse responses to an identified monetary policy shock, as in Christiano et al. (2005). To simulate a monetary policy shock in the model, we assume that the real rate follows an AR(1) process, with $r_t = (1 - \rho_r)r^* + \rho_r r_{t-1} + \epsilon_r^t$, with persistence $\rho_r$ and innovation $\epsilon_r^t \sim N(0, \sigma_r^2)$. The parameters to be estimated are then inattention of households $\theta^h$, inattention of firms $\theta^f$, inattention of unions $\theta^u$, the slope of the price Phillips curve $\kappa^p$, the slope of the wage Phillips curve $\kappa^w$, as well as the parameters of the real rate process, $\rho_r$ and $\sigma_r$. Collecting these parameters in $z = (\theta^h, \theta^f, \theta^u, \kappa^p, \kappa^w, \sigma_r, \rho_r)$, let $\Psi(z)$ denote the model-implied impulse responses and $\hat{\Psi}$ their empirical counterpart. Our estimator $\hat{z}$ solves

$$
\min_z (\Psi(z) - \hat{\Psi})' \Sigma^{-1} (\Psi(z) - \hat{\Psi})
$$

where $\Sigma$ is a diagonal matrix with the variances of the empirical impulse responses on the diagonal. We compute an estimator $\hat{\Sigma}$ for the asymptotic variance-covariance matrix of $\hat{z}$ as

$$
\hat{\Sigma} = \left( \frac{\partial \Psi}{\partial z}(\hat{z})' \Sigma^{-1} \frac{\partial \Psi}{\partial z}(\hat{z}) \right)^{-1}
$$

15This is the same value as in Kaplan et al. (2018), and somewhat higher than the value of 0.70 implied by Floden and Lindé (2001), in order to capture the “new view” of household idiosyncratic risk.
where $\frac{\partial \Psi}{\partial z}(z)$ is the Jacobian of $\Psi(z)$, evaluated at $z$. In our application, we include in $\Psi$ the responses of output $Y_t$, investment $I_t$, consumption $C_t$, real rates $r_t$, the price level $P_t$, real wages $W_t/P_t$, and hours $N_t$.

This procedure can be implemented for various empirical impulse responses to identified monetary policy shocks. We next present our baseline approach. In Appendix ?? we estimate our model to match other sets of impulse responses that have been proposed by the literature.

### 4.2 The empirical response to a monetary policy shock

As our baseline approach, we use Jordà (2005) projections using Romer-Romer shocks as instruments (Romer and Romer 2004) on the original Romer-Romer sample (1969m3–1996m12). This is one of Ramey (2016)'s preferred specifications (see her Figure 2, panel B). We collect monthly data on outcomes $Y_t$, including output, consumption, investment, hours, prices, real wages, the nominal interest rate and a measure of the real interest rate constructed by subtracting one-year ahead SPF inflation forecasts from the nominal interest rate.\footnote{Specifically, we use the logs of real chained GDP (GDPC1), real chained personal consumption expenditures (DPCERA3M086SBEA), real chained investment (GPDIC1), hours of all persons in nonfarm business sector (HOANBS), the consumer price index for all items (CPIAUCSL), average hourly earnings of private production employees divided by the CPI (AHETPI/CPIAUCSL), as well as the level of the Federal Funds rate (FEDFUNDS) and the difference between that rate and the one-year inflation forecasts from the Survey of Professional Forecasters. GDP, Investment and inflation expectations are only available at the quarterly frequency so we interpolate those to monthly frequency before running the regression.}

\[ Y_{t+h} = \gamma_{Y_h} \epsilon_t + X_t + \eta_t \]  

separately for horizons $h = 1 \ldots T$, up to $T = 48$ months, where $\epsilon_t$ is the Romer-Romer series. To control for the potential endogeneity of the this series, we control for lags of $Y_t$, as well as lags of industrial production, unemployment, the consumer price index and a commodity price index in all specifications. For output, consumption, investment, we run the regression in levels, dividing by the level of output in the period before the shock, so that the impulse response are interpretable as the percentage movement in these variables relative to output in a base period, and accounting identities are satisfied. We compute standard deviations using a Newey and West (1987) correction. We then aggregate back the impulse responses $\{\gamma_{Y_h}\}$ to the quarterly level.

The solid lines in figure 4 show these impulse responses. The nominal interest rate falls by 15 bps. Since inflation expectations, like actual inflation, do not respond much on impact, the real interest rate falls too and by a similar amount. Output, consumption, investment and hours follow the typical hump-shaped pattern also documented in many other studies. Consumer prices take time to respond—they do not rise initially and only start picking up around quarter 6.

### 4.3 Computational methodology

Our estimation procedure to match empirical impulse responses is very close to the industry standard popularized by Christiano et al. (2005). There is one crucial distinction, however, which is
Note. This figure shows our estimated set of impulse responses to an identified Romer and Romer (2004) monetary policy shock (black, with gray confidence intervals). The green dashed line are the impulse responses implied by our estimated inattentive heterogeneous-agent model.

that we estimate a model with heterogeneous agents, rather than a representative agent. This requires us to overcome the computational challenges associated with simulating heterogeneous-agent models many times when minimizing (10).

We achieve this by relying on the recent sequence space methodology that we propose in Auclert, Rognlie and Straub (2019). As we show in that paper, we can solve for the model-implied aggregate impulse responses $\Psi$ of a shock $\epsilon$ by solving a system of equations,

$$F(\Psi, \epsilon) = 0$$

to first order in $\epsilon$. The impulse response is then given by $F^{-1}_\Psi F_\epsilon$. For the Jacobian $F_\Psi$, we rely on the computational approach in Auclert et al. (2019), which builds on a representation of the model economy as a computational graph, as well as a “sandwich product” procedure to compute the household sector’s response to shocks. For details, we refer the reader to that paper. Importantly, the sequence space approach allows for a straightforward transformation of a full-attention economy, to an inattentive economy, with essentially no extra computational burden, by applying simple column operations to the full-attention Jacobians $F_\Psi$ and $F_\epsilon$.

How efficient is the sequence space approach in practice? On a 2018 MacBook Pro, after the steady state is solved for, the simulation of the impulse responses in response to a monetary policy shock takes around 60 milliseconds. The full impulse response matching therefore only takes a few
Table 1 shows the parameter estimates for our baseline impulse responses. The model-implied impulse responses are the dashed lines in Figure 4. Overall, the model produces a decent fit to the impulse responses, comparable with that of representative-agent models, with two exceptions: the model’s price level response is initially too fast and then does not accelerate enough; and as mirror image of this, it under predicts the initial increase in the real wage and over-predicts the eventual decline. The culprit for this is the fact that in an effort to keep our model as parsimonious as possible, we chose not to include price indexation in the two Phillips curves or to add a “cost channel” of monetary policy that causes deflation in the short run.

Our parameter estimates for the Phillips curve slopes suggest a significant degree of price and wage stickiness. The estimated slope of the Phillips curve, $\kappa_p$, is 0.071. The estimated slope of the wage Phillips curve, $\kappa^w$, is 0.004. Both lie in the ballpark of the estimates in the literature, and indicate a significant degree of nominal rigidity.\footnote{For example, Altig et al. (2011) estimate price and wage reset frequencies that correspond to slope parameters $\kappa_p = 0.014$ and $\kappa^w = 0.003$. The agnostic prior specification in Del Negro and Schorfheide (2008) finds frequencies that correspond to slope parameters $\kappa_p = 0.06$ and $\kappa^w = 0.02$.}

Our estimated inattention parameters show that while unions are not inattentive, $\theta^u = 0$, there is considerable inattention among households and firms, $\theta^h = 0.89$ and $\theta^f = 0.96$. To understand why unions are attentive in our estimation, note that union inattention and slope of the wage Phillips curve are highly substitutable, so either low $\kappa^w$ or high $\theta^u$ could cause nominal wage stickiness. Our estimation found that low $\kappa^w$ is a better description of the data. To get a sense for the magnitudes of inattention among households and firms, notice that a parameter estimate of $\theta^h = 0.89$ suggests that approximately half of all households are perfectly attentive by quarter 5.

\subsection*{4.4 The important role of inattention}

What role do household and firm inattention play for our estimated model? To investigate this question, we conduct two experiments.

\textbf{Switching off inattention.} In figure 5, we iteratively switch off inattention in our model (without re-estimating the other parameters). The green line shows our baseline impulse responses for output and consumption; the dashed blue line the same responses with household inattention switched off; the solid blue line the responses without firm inattention. We see that both types of inattention are necessary to avoid a counterfactual jump in consumption and investment. Thus, while inattention among firms may be enough to generate a hump-shaped response in investment, it does not imply a hump-shaped response of consumption.

\textbf{Household inattention and direct vs. indirect effects.} Why is inattention among households both necessary and sufficient for a hump-shaped response of consumption, despite the presence of
Figure 5: 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, but without re-estimation. Green uses our baseline estimates of household and firm inattention. Solid blue sets all inattention parameters to zero. Dashed blue only sets household inattention to zero.

large MPCs in our model? To understand this question, we decompose the aggregate consumption response in direct and indirect components, similar to the decomposition in Kaplan et al. (2018). As the aggregate consumption can be written as a function $C_t = C_t \left\{ r_s, Z_s, d_m \right\}$ of real rates, aggregate after-tax labor incomes, and aggregate transfers, we can decompose the aggregate consumption response into three terms, a direct and two indirect ones,

$$dc_t = \sum_s \frac{\partial C_t}{\partial r_s} dr_s + \sum_s \frac{\partial C_t}{\partial Z_s} dZ_s + \sum_s \frac{\partial C_t}{\partial d_m} d_{m,s}$$ (12)

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

Figure 6 shows how inattention affects the breakdown into direct and indirect effects, when the paths of $\left\{ r_s, Z_s, d_{m,s} \right\}$ are those observed in response to the monetary shock in figure 4. Since our inattentive HA model preserves high MPCs, the indirect effect is only dampened initially—due to smaller anticipatory effects of future labor income and transfers on initial consumption—but otherwise almost equally large. The direct effect, however, which is responsible for the initial jump of consumption in the fully attentive model, is strongly dampened by inattention, even more so than already in fully attentive HA models (Kaplan et al. 2018). This emphasizes that the initial jump relied on an immediate response of households’ beliefs about future interest rates to the monetary shock.

Thus, by dampening direct effects, household inattention can ensure that consumption follows
Figure 6: Decomposition of consumption

Note. This figure decomposes the consumption response of our estimated HA model in to direct and indirect effects as in Kaplan et al. (2018), see 12.

the same type of hump-shaped trajectory that can be observed in the aggregate macro data, while still being consistent with micro evidence on (intertemporal) MPCs.

5 Investment is the transmission mechanism

The prominent role of indirect effects for consumption in our estimated HA model fundamentally alters the monetary transmission mechanism. In particular, we argue that it implies that investment becomes the crucial driver of both output and consumption responses to monetary policy shocks.

5.1 Transmission channels in the estimated HA model

Conceptually, monetary policy in our model is transmitted to output through a complicated web of direct effects (red in figure 7) and indirect effects (green in figure 7). There are three kinds of direct effects: first, real interest rates directly change investment decisions by affecting the user cost of capital (the red line from “r” to “I”); second, real interest rates change consumption decisions as in (12), through substitution and income effects (the red line from “r” to “C”); third, real interest rates change the interest payments the government needs to make on its (newly issued) debt (the red line from “r” to “G”).

These “first round” direct effects then trigger a number of indirect effects, which capture the impact of the endogenous change in aggregate output $Y_t$ on the three sectors, as well as how these sectors affect each other. For instance, an increase in $Y_t$ increases aggregate labor income $Z_t$ and raises consumption (the green line from “$Y$” to “$C$”), especially so when MPCs are high. The green indirect effects, combined with the red direct effects shape the equilibrium response of macro aggregates—e.g. output, consumption and investment—to the monetary policy impulse.
In our estimated HA model, these channels are not all equally powerful. In fact, as we already explored in section 4.4, household inattention leads to a very limited role for the direct effect on consumption, and a much larger role for the indirect effects. Thus, one might say that consumption is not an active respondent to monetary policy, but a passive one. Is this partial equilibrium breakdown important for the general equilibrium transmission of monetary policy shocks? We argue that it is in the next subsection, by focusing on the active role of investment.

5.2 Investment drives the output response

Imagine investment were unresponsive to shocks in our estimated HA model. How would that affect the response of consumption and output to a monetary policy shock?

We conduct this experiment in figure 8. The green solid lines show the output and consumption responses of our estimated HA model (see figure 4). The green dashed line performs the same monetary policy experiment, just assuming that investment stays put at its steady state level and does not react. Clearly, this must reduce the output response as investment is a component of aggregate demand. However, somewhat less immediately clear is that it also reduces the consumption response (right panel). As can be seen, this effect is large and therefore contributes to a further decline in the output response, relative to the case with investment, which then further declines consumption and so on. This feedback effect is reminiscent of a “Keynesian cross” type logic, which was formalized and modernized for heterogeneous-agent models in Auclert et al. (2018). Taken together, shutting off the investment response shrinks the present values of the output and consumption responses (over the first 20 quarters) by 88% and 83% respectively.

Why is the effect of lower aggregate output on consumption this powerful? The reason is precisely the relative importance of indirect channels vs. direct ones. When the estimated consump-
Figure 8: Impulse responses with and without investment

Note. The panels show the output and consumption responses after a monetary policy shock, comparing responses with (solid) and without investment (dashed) in our inattentive HA model (green) and an estimated RA model with habits (red).

The output response consists mostly of indirect channels coming from aggregate output, consumption will collapse when output does.

To emphasize that this logic is not a mere consequence of matching the macro dynamics in figure 4 but instead crucially relies on matching micro moments such as MPCs on top of that, we also include in figure 8 what would happen in an estimated habit-RA model if investment were shut off (see appendix B.1 for details on the habit-RA model). Since consumption in that model is entirely driven by direct effects, it does not respond at all to changes in aggregate output, which is why the decline in the output response, when investment is absent, is much more modest (23% vs 88%).

An additive general equilibrium decomposition. Interpreted through the lens of the schematic plot in figure 7, this suggests that the direct effect of monetary policy on investment ends up causing the bulk of the general equilibrium output response in the estimated HA model (after taking into account the indirect knock-on effects through consumption), but not in the estimated habit-RA model.

To fully understand how the transmission works in the HA and RA models, it is thus important to decompose the general equilibrium output response, going beyond individual sectors’ partial equilibrium behavior. To achieve such a decomposition in a well-defined way, we must clearly define what it means to “only shock the household sector”, or “only shock investment”. This can be achieved by introducing three different interest rates into the model, appearing in three different places in the model’s equilibrium conditions, one for each of the direct effects we aim to disentangle.

In particular, we denote by \( \{ r^C_t \} \) the interest rate that agents use in their Euler equation. Such a shock is well-defined as it merely changes households’ behavior (not their income flow), repre-
Figure 9: Channels of monetary policy transmission

(a) Representative agent

(b) Heterogeneous agents

Note. This plot attaches weights to the direct effects from \( r \) to the three sectors, in line with the importance of a direct channel for the general equilibrium output response (see (13)). To compute the weights, we simulated the general equilibrium responses when only one of the direct channels was active at a time. For the indirect effects, we only show arrows pointing to a sector if that link is “important” (in a quantitative sense) for that sector. For example, the indirect links from “\( G \)” and “\( I \)” to “\( C \)” are absent in the RA model, since those income effects exactly cancel out and consumption, through the Euler equation, is entirely driven by the path of real interest rates.

senting the substitution effect on consumption. We denote by \( \{r^I_t\} \) the interest rate that enters the user cost of capital through the firms’ maximization problem (7). Such a shock to first order only affects the firms’ investment choices, but nothing else in the model. Finally all other occurrences of the interest rate relate to income effects from greater payments on government debt or deposits. We denote this interest rate by \( \{r^G_t\} \). When all three types of interest rates are shocked jointly and by the same amount, one recovers the original shock to \( r_t \) that we seek to decompose. In that sense, the general equilibrium output response is a mere sum of responses to the three individual shocks

\[
dY_t = \sum_s \frac{dY_t}{dr^C_s} dr^C_s + \sum_s \frac{dY_t}{dr^I_s} dr^I_s + \sum_s \frac{dY_t}{dr^G_s} dr^G_s
\]

This additive decomposition allows us to compute how important each channel is for the present value response of output (after 20 quarters). We do so in figure 9, where we only show the most important active channels in the two models. The numbers confirm our previous results: the direct response of investment to interest rates accounts for the bulk of the output response in the inattentive HA model. In the habit-RA model, the direct response of consumption (intertemporal substitution) accounts for the bulk of the general equilibrium output response.\(^{18}\)

Why should we care whether investment matters for the transmission mechanism or not? After all, the estimated RA and HA models produce similar aggregate responses. We next provide two important reasons why we should care: state dependence, and the path of natural rates.

\(^{18}\)There is no effect of \( r^G_t \) in the RA model due to Ricardian equivalence.
5.3 Implication 1: State dependence

Knowing which transmission channels are operative is informative for the state-dependent effects of monetary policy. To illustrate this in our case, consider a monetary intervention of the kind in figure 4 in three different economic environments. (1) The intervention happens at the steady state of the model. (2) The intervention happens while the economy suffers from an “housing bust”: a period of weak investment where investment is at the irreversibility constraint (6) and therefore does not respond.\(^{19}\) (3) The intervention happens during times of very low nominal interest rates, during which liquid interest rates no longer adjust, in line with recent evidence on sticky deposit rates (Drechsler, Savov and Schnabl 2017): we implement this by simply assuming that mutual funds do not pass through \(r_t\) to the liquid rates that consumers receive. Instead, they adjust the transfer \(d_m\) to the households’ illiquid accounts accordingly.

Figure 10 juxtaposes the responses of output (black, solid) and consumption (green, dashed) in these three environments. By design, the responses in panel (a) are the estimated ones from figure 4. The responses in panel (b) are equivalent to those in (8) (without investment) and highlight that investment may well be inactive during “housing bust” episodes. Finally, in panel (c) we see that an active response by consumers is not at all necessary for monetary policy to affect output, as long as investment responds.\(^{20}\) Indeed, as happens to be the case in our model, the fact that households are no longer experiencing lower interest rates in their liquid accounts, causes a sufficiently positive income effect to overturn the dampened substitution effect. This even increases the effect of the monetary policy shock on output.

In short, our estimated HA model provides a novel narrative for why monetary policy was “pushing on a string” in the recent financial crisis: investment could no longer be stimulated, which reduced the transmission to output by 88%—much more than the mere 23% that an estimated RA model would have suggested.

5.4 Implication 2: Natural rate

A state-dependent transmission of interest rates to output suggests that crises during which investment is constrained, such as housing busts, may require a much more pronounced reduction in interest rates to establish full employment than other types of crises, during which investment is not constrained. In other words, a crisis driven by a housing bust may imply much lower natural rates than a shock to consumption, say.

To explore this idea, we first simulate the impulse response to a simple AR(1) discount factor shock, \(d\beta_t = \rho\beta d\beta_{t-1} + \epsilon_t\). We choose \(\rho = 0.95\) (quarterly) and an initial shock size \(\epsilon_0 > 0\) so as to generate a large recession, with a trough at around \(-7\%\) relative to trend. This shock is our non-housing-bust shock, during which investment is still fully responsive to interest rates. We compare this to a housing bust shock, which, for this simple example, we model as an exogenous decline

\(^{19}\)We call it “housing bust” since residential investment and non-residential investment in structures are arguably the types of investment that are most responsive to monetary policy. See our discussion in section 5.5.

\(^{20}\)We discuss the potential role households are playing in making investment decisions (e.g. in housing) in section 5.5.
in the path of investment \( \{dI_t\} \). We choose the first 50 levels of the path \( \{dI_t\} \) to perfectly match the first 50 quarters of the recession in output caused by the discount factor shock. After that, we assume \( dI_t \) follows an AR(1) process back to its steady state level. This procedure ensures that both the discount factor shock (the “C shock”) as well as the housing-bust shock (the “I shock”) cause exactly the same type of recession, but that investment is unresponsive during an I shock episode.

As we show in the left panel of figure 11, both types of shocks indeed cause a recession of the same magnitude. But, the right panel makes clear that this does not mean that monetary policy should react the same way: the natural rate—here constructed as the flexible-price and flexible-wage equilibrium interest rate—falls by almost twice as much in the I shock episode. This stark result underscores that it is imperative for policymakers to understand the sources of a given business cycle shock and how such shocks may affect the responsiveness of investment before deciding on the path for policy rates.

5.5 Discussion

Our finding that investment is important for the transmission of monetary shocks raises several questions, which we now discuss.

What kind of investment? Investment can be broken down into several subcategories. Which of those categories might be most important for the transmission of monetary policy? Since, as we argued, the relevant channel is the direct effect of interest rates on investment, we expect this
Figure 11: The shock-dependence of the natural interest rate

Note. This figure contrasts two shocks, one to consumption, “C shock”, and one to investment, “I shock”. During the latter shock, investment follows a fixed, unresponsive path that is chosen such that both shocks imply equal paths for output absent any monetary policy intervention (left panel). The right panel shows the path of the natural rate—here defined as the flexible-wage and flexible-price equilibrium rate—under both shocks.

channel to be dominated by the most interest-elastic types of investment. These types are likely the ones that depreciate the slowest, as a lower rate of depreciation increases the interest elasticity of the user cost of capital: at a real interest rate $r = 4\%$, an asset that depreciates at $\delta = 4\%$ is twice as elastic to interest rates as an asset with $\delta = 12\%$. It therefore seems natural to us that the type of investment that is at the very core of the monetary transmission mechanism is residential investment and non-residential investment in structures, i.e. broadly speaking, investment in housing.

**How to think of housing and durables in the model?** In our model, housing investment is part of investment $I_t$. When interest rates fall, (housing) investment rises, financed by lower dividends, and ultimately therefore by lower transfers from the illiquid account. When thinking about residential investment, these lower transfers are a stand-in for greater mortgage payments. One may think of durable spending in our model as a kind of investment, financed in a similar fashion.\(^{21}\)

**What if borrowing constraints relax?** Aside from financing housing investment through lower transfers, households may also desire to borrow in order to raise consumption, rather than investment. This could in principle amplify the direct response of consumption to interest rates. In our baseline model in section 3, this “consumer credit” channel is absent, as we assume an interest-rate inelastic borrowing constraint (of zero). We relax this assumption in appendix 4?, where we re-estimate our model with a borrowing constraint $\bar{g}(r) \equiv -\lambda \frac{\bar{r}}{T}$. We set $\lambda$ equal to the average quarterly income, as in Kaplan et al. (2018). As we show in the appendix, when such a model is estimated to match the exact same set of impulse responses in figure 4, the greater direct response

\(^{21}\)In fact, it is not uncommon in this literature to count durables as investment, see e.g. Christiano et al. (2005) and Justiniano, Primiceri and Tambalotti (2010).
is largely dampened by greater inattention. The reason for this is exactly as in section 4.4: since the fully attentive direct effect is greater with elastic borrowing constraints and since inattention does not affect MPCs, a greater degree inattention is necessary to dampen the direct response. This is why when we perform our no-investment counterfactual in an economy with elastic borrowing constraints, we find a similarly large role of investment (for details, see appendix ??).

**How robust is this to other types of inattention?** Our model in section 3 assumes a particular kind of inattention—sticky information as in Mankiw and Reis (2002, 2007). In appendices ?? and ??, we repeat our estimation for two alternative ways to generate inattention, that is, slow updating of beliefs: a version of cognitive discounting, as in Gabaix (2016), and a version of incomplete information, as in Angeletos and Lian (2018). The results for both types of inattention broadly confirm our findings in sections 4 and 5: (a) there is significant inattention among households and firms; (b) the inattention largely dampens the direct effect of interest rates on consumption, rather than the indirect effects; (c) this leads to a very significant role for investment in the monetary transmission mechanism.

In all our implementations of inattention, we assume that households always perfectly observe what flows in and out of their liquid account. To us, this is a reasonable assumption, for three reasons. First, each household’s idiosyncratic income is very volatile, so that households naturally will want to track their income closely. Second, households face a possibly binding borrowing constraint in their liquid account. Violations of that borrowing constraint are impossible (in the model) and carry considerable costs, e.g. overdraft fees (in reality). And third, to the extent that some of the current income fluctuations are caused by movements along the extensive margin, i.e. more or fewer hours worked, households are well informed about them.22

**How robust is this to other empirical specifications?** Our specification in section 4.2 is very close to the Jordà (2005) version of Romer and Romer (2004), as proposed by Ramey (2016), with one exception; we interpolate variables that are only available at the quarterly level, to stay as close to the original (monthly) specification in Ramey (2016) as possible. This works since we are not interested in sub-quarterly fluctuations in these variables and instead aggregate our monthly impulse responses back to the quarterly level. Still, to ensure that our approach is by no means special, we estimate our model to the well-studied VAR responses in Altig et al. (2011) in appendix ??.

**Taylor rule vs. real rate rule.** Our model in section 3 assumes a rule for the real interest rate, equivalent to a Taylor rule with a coefficient of 1 on expected inflation. A real-rate rule is very instructive as it allows to cleanly track how monetary policy affects the economy (see, e.g., our

---

22 We do not assume that households have perfect information about all positions in the illiquid account. This is line with a large literature on infrequent portfolio adjustment and spending behavior in response to fluctuations in stock prices (e.g. Chodorow-Reich Nemov Simsek 2019).
decomposition in section 5.2). We re-estimate a version of our model with a Taylor rule in appendix ??  We find that the results are almost identical.

Our estimated HA model compared to HANK and TANK. Kaplan et al. (2018) proposed one of the first heterogeneous-agent New-Keynesian models, “HANK”, and demonstrated that consumption is determined in roughly equal parts by the direct effect, an indirect effect due to taxes, and an indirect effect due to labor income. While our estimation implies an even more diminished role for the direct effect, one may still wonder whether an important role for investment is an immediate consequence of a small direct effect (and large indirect effects) on consumption.

This is not obvious at all. In fact, it crucially depends on the relative strengths of the direct effects on consumption vs investment. For instance, it could be that the direct effect on consumption, despite being small, is larger than the direct effect on investment; in that case, shutting down the latter would not affect the output response by much. Thus, it is not at all obvious that the HANK model would give the same result as our estimated HA model.

Another popular model to study monetary policy is a two-agent New-Keynesian model, or “TANK”. This model is has a household sector that consists of some share \( 1 - \mu \) of agents that are like the representative agent in the habit-RA model, and a share \( \mu \) of hand-to-mouth agents. This model can generate high (on-impact) MPCs when \( \mu \) is large, but does not match intertemporal MPCs (Auclert et al. 2018). Does this matter? Appendix ?? estimates this model to match the impulse responses with \( \mu = 0.20 \), to mirror the on-impact MPC in our estimated HA model. While in the TANK model, investment accounts for a larger fraction of output than in the habit-RA model, 27% vs. 23%, this is still far lower than the 88% in the estimated HA model. Similarly, the direct response of consumption (the \( r_C^t \) effect) is still responsible for most of the general equilibrium output response, 81%. This highlights that the TANK model gets the direction of the effect right, but misses the quantitative magnitude.

6 Revisiting other topics

Our estimated HA model is the first heterogeneous-agent model to match both micro evidence on high (i)MPCs as well as macro evidence on hump-shaped impulse responses. In this section, we let the model speak to recent debates in the literature.

6.1 The role of fiscal rules

The seminal work by Kaplan et al. (2018) has demonstrated that the type of fiscal rule operative during a monetary policy change can matter significantly for the response of output and consumption. Figure 12 shows that this is still the case.

Replacing long-term bonds (as in our baseline model) with short-term ones increases the interest savings the government makes in response to the monetary easing. The money that would
Figure 12: Response of monetary policy under different fiscal rules

Note. This figure shows the output and consumption responses in our baseline estimated HA model (green, solid) as we change the fiscal rule. Thick dashed green is an economy with short rather than long-term bonds. Thin dashed green is an economy with a fiscal rule that adjusts lump-sum taxes rather than proportional labor taxes. The thick blue dashed represents a fiscal rule that reduces deficits more than our baseline economy. The thin blue dashed line is an economy that is closer to balanced budget.

have gone to bond holders now is passed through to households in the form of lower taxes. This increases spending somewhat (thick dashed green line).

A fiscal rule that is built on adjusting lump-sum taxes, instead of proportional labor taxes, also leads to a greater response of consumption, and thus output (thin dashed green line). This is because labor income taxes transfer income to households in proportion to their productivity, rather than lump-sum, which lowers the (weighted) average MPC.

Finally, making the fiscal rule more (or less) deficit-reducing—corresponding to a smaller (or larger) $\psi$—can also decreases (or increases) the output response (blue dashed lines). This is in line with the idea that the high MPCs in this model cause a significant violation of Ricardian Equivalence.

6.2 Forward guidance

In figure 13 we run the prototypical “forward guidance” experiment: what happens to the economy when we cut the real interest rate for one period at some horizon $t = h$? In a recent set of papers, Del Negro et al. (2013) and McKay et al. (2016) documented that a textbook representative-agent model with unit intertemporal elasticity exhibits an unrealistic one-for-one response of consumption to forward guidance at any horizon. Werning (2015) shows that a similar one-for-one response is possible in a typical HA model without investment.

In the left panel, we simulate a string of forward guidance experiments in our HA economy with investment, but without household or firm inattention. Output jumps immediately by more than one for one, more so if the horizon $h$ is large, and the output response is always largest at date $t = 0$. In other words, a standard HA model with investment but without inattention makes
the forward guidance puzzle worse! Why is this? As we showed in section 5, the presence of investment significantly increases the potency of monetary policy in HA models. In the context of forward guidance, this greater potency causes a worsening of the puzzle.

Contrast this with the right panel, where we simulate the same experiments in our economy with inattention. Output reacts very sluggishly and peaks just before the real interest rate cut—a significant improvement over the no-inattention model. However, it still suggests that later cuts in the real interest rate lift up the output response at all dates, even if not by much in some periods. This feature (or rather bug) is not robust when the form of inattention is changed. We illustrate this in appendix ?? for the case of cognitive discounting as in Gabaix (2016).

7 Conclusion

The goal of this paper is to bring our current vintage of heterogeneous-agent models in line with the often sluggish behavior of macroeconomic aggregates. Our approach introduces insights from two different literatures into heterogeneous-agent modeling: a voluminous literature on estimation of macroeconomic models by matching impulse responses to identified shocks; and the recent literature on bounded rationality in macroeconomics. Our estimated inattentive heterogeneous-agent model is able to match impulses responses to monetary policy shocks well, thanks, in part, to significant levels of estimated household and firm inattention.

One aspect of household inattention is that it mostly dampens the direct effect of consumption. We show that this matters crucially for understanding the transmission of monetary policy shocks. While in a similarly estimated, standard representative-agent model with habits investment only accounts for 23% of the output response, this number is 88% in our estimated heterogeneous-agent model. We argue that the difference can be traced to the fact that in our model, indirect effects on consumption are much more potent than in a standard representative-agent model.

The outsized role of investment is important for policymakers as it has immediate practical im-
plications. Most importantly, it implies that the effects of monetary policy are significantly state-dependent: when fighting a recession such as the Great Recession that emanated from a housing bust, investment cannot be counted on responding. While according to standard representative-agent models this would not be an issue as most of the monetary transmission operates through direct effects on consumption, it is a major issue in our estimated HA model, which hinges on an operational investment channel. Thus, our model rationalizes why monetary policy appeared to be “pushing on a string” during the Great Recession, and suggests monetary and fiscal policymakers should, if possible, dare to be more aggressive when investment is less likely to respond.

References


In this section we derive the consumption function for an inattentive permanent-income agent that faces a constant interest rate \((1 + r) = \beta^{-1}\) and a known stream of income \(\{y_t\}\). The Bellman equation of the attentive agent is

\[
V_t^A(a_{t-1}) = \max_{c_t, a_t} u(c_t) + \beta V_{t+1}^A(a_t)
\]

\[
c_t + a_t \leq (1 + r)a_{t-1} + y_t
\]
implying the familiar first-order condition \( c_t = c_{t+1} \). The level of consumption at date \( t \) must then satisfy, by the intertemporal budget constraint implies

\[
c^I_t (a_{t-1}) = \frac{r}{1+r} \left( \sum_{k=0}^{\infty} \left( \frac{1}{1+r} \right)^k y_{t+k} + (1+r) a_{t-1} \right)
\]

(14)

in particular, in steady state

\[c^{ss} (a) = y_{ss} + ra_{ss} = c_{ss}\]

and therefore

\[V^{ss} (a) = \frac{1}{1-\beta} u (y_{ss} + ra)\]

(15)

Given 15, the Bellman equation for the inattentive agent is

\[
V^I_t (a_{t-1}) = \max_{c_t, a_t} \left[ u(c_t) + \frac{\beta}{1-\beta} u (y + ra_t) \right] \quad c_t + a_t = (1+r)a_{t-1} + y_t
\]

which implies

\[u'(c_t) = u'(y_{ss} + ra_t)\]

and hence a consumption function

\[c_t = y_{ss} + ra_t\]

This implies that the inattentive agent accumulates assets with the policy

\[a^I_t (a_{t-1}) = a_{t-1} + \frac{y_t - y_{ss}}{1+r}\]

and therefore that, starting from date 0,

\[c^I_t = c_{ss} + \frac{r}{1+r} \sum_{k=0}^{t} (y_k - y_{ss})\]

(16)

Using these results we can solve for the change in consumption of an agent whose stream of agent changes from steady state by \( \{dy_k\} \), in case he becomes informed at date \( u \):

\[
dc_A^{I,u} = \begin{cases} 
\frac{r}{1+r} \sum_{k=0}^{t} dy_k & t < u \\
\frac{r}{1+r} \left( \sum_{k=0}^{\infty} \left( \frac{1}{1+r} \right)^k \min\{k-u,0\} \right) dy_k & t \geq u 
\end{cases}
\]
Aggregate consumption at time $t$ is the sum of consumption of all agents not yet informed, plus the consumption of the

$$\begin{align*}
dc_t &= \theta^t dc_t^I + (1 - \theta) \sum_{u=0}^{t-1} \theta^u dc_t^{1,u} \\
    &= \frac{r}{1 + r} \sum_{k=0}^{\infty} \left( \theta^t 1_{\{k \leq t\}} + (1 - \theta) \sum_{u=0}^{t-1} \theta^u \left( \frac{1}{1 + r} \right)^{\min\{k-u,0\}} \right) dy_k
\end{align*}$$

Which implies that the iMPC at date $t$ for income at date $s$ is

$$\frac{\partial c_t}{\partial y_s} = \frac{r}{1 + r} \left( \theta^t 1_{\{s \leq t\}} + (1 - \theta) \sum_{u=0}^{t-1} \theta^u \left( \frac{1}{1 + r} \right)^{\min\{s-u,0\}} \right) \tag{17}$$

### A.2 Inattentive firms

Let $\eta_t$ be the multiplier on the production constraint (8). We start with firms that have become attentive and consider in turn their first-order conditions

#### Labor choice.

Taking for FOC for $n_t$, we find

$$\begin{align*}
(1 - \eta_t) \frac{p_t}{P_t} F_n(k_{t-1}, n_t) + \frac{\eta_t}{\mu} \left( \frac{F(k_{t-1}, n_t)}{Y_t} \right)^{\frac{1}{\mu-1}} F_n(k_{t-1}, n_t) &= \frac{W_t}{P_t}
\end{align*}$$

which rewrites

$$\frac{p_t}{P_t} \left( 1 - \eta_t \left( 1 - \frac{1}{\mu} \right) \right) = \frac{W_t / P_t}{F_n(k_{t-1}, n_t)} \equiv mc_t$$

where $mc_t$ stands for average real marginal cost. This implies that

$$1 - \eta_t = \frac{\mu mc_t / p_t / P_t - 1}{\mu - 1} \tag{18}$$

#### Price choice.

Taking the FOC for $p_{t+1}$, we find

$$\begin{align*}
\frac{p_{t+1} - p_t}{p_t} \cdot \frac{p_{t+1}}{P_t} &= \frac{\kappa^P (\mu - 1)}{1 + r} \left( 1 - \eta_{t+1} \right) \left( \frac{p_{t+1}}{P_{t+1}} \right)^{\frac{1}{\mu-1}} + \frac{1}{1 + r} \left( \frac{p_{t+2} - p_{t+1}}{p_{t+1}} \right) \left( \frac{p_{t+2}}{p_{t+1}} \right)
\end{align*}$$

using (18) we arrive at the pricing equation

$$\begin{align*}
\frac{p_{t+1} - p_t}{p_t} \cdot \frac{p_{t+1}}{P_t} &= \frac{\kappa^P \left( \mu mc_{t+1} - 1 \right)}{1 + r} \left( \frac{p_{t+1}}{P_{t+1}} \right)^{\frac{1}{\mu-1}} + \frac{1}{1 + r} \left( \frac{p_{t+2} - p_{t+1}}{p_{t+1}} \right) \left( \frac{p_{t+2}}{p_{t+1}} \right) \tag{19}
\end{align*}$$
Note that when all firms are attentive, this delivers a standard Rotemberg Phillips curve

$$\pi_{t+1} \cdot (1 + \pi_{t+1}) = \frac{\kappa^p}{1 + r} (\mu mc_{t+1} - 1) + \frac{1}{1 + r} \pi_{t+2} \cdot (1 + \pi_{t+2})$$

**Investment choice.** The FOC for \( k_t \) is

$$\zeta' \left( \frac{k_t}{k_{t-1}} \right) = \frac{1}{1 + r_{t+1}} \frac{\partial J_{t+1}}{\partial k_t} = \frac{q_{t+1}}{1 + r_{t+1}}$$

where we denote \( q_{t+1} \equiv \frac{\partial J_{t+1}}{\partial k_t} \), while the dynamics of \( q_t \) follow from the Envelope condition

$$q_t = \frac{\partial J_{t+1}}{\partial k_t} = mc_{t+1} F_k(k_t, n_{t+1}) - \zeta \left( \frac{k_{t+1}}{k_t} \right) + \frac{k_{t+1}}{k_t} q_{t+2} \frac{1}{1 + r_{t+2}}$$

**Inattentive firms.** Inattentive firms believe take all date-\( t + k \) variables as constant at their steady state values \( P_{ss}, W_{ss}, Y_{ss} \), and \( r_{ss} \) for \( k \geq 0 \). It follows that the price-setting equation is

$$\frac{p_{t+1} - p_t}{p_t} = \frac{\kappa^p}{1 + r} \left( \mu mc_{ss} \frac{P_{t+1}^*}{P_{ss}^*} - 1 \right) \left( \frac{p_{t+1}}{P_{ss}^*} \right)^{-\frac{r_{t+1}}{1 + r}} + \frac{1}{1 + r_{ss}} \left( \frac{p_{t+2} - p_{t+1}}{p_{t+1}} \right) \left( \frac{p_{t+2}}{p_{t+1}} \right)$$

Starting from the initial state price \( p_0 = P_{ss} \), this equation has as its unique solution \( p_{t+1} = P_{ss} \) for all \( t \). Hence, inattentive firms do not change their price until they become attentive. Similarly, for capital, the set of equations of the firm is

$$(1 + r_{ss}) \zeta' \left( \frac{k_t}{k_{t-1}} \right) = mc_{ss} F_k(k_t, n_{ss}) - \zeta \left( \frac{k_{t+1}}{k_t} \right) + \frac{k_{t+1}}{k_t} \zeta' \left( \frac{k_{t+1}}{k_t} \right)$$

Starting from \( k_{-1} = k_{ss} \), this equation also has a unique solution \( k_t = k_{ss} \).

**Aggregate inflation and investment.** At time \( t \), \( (\theta^f)^t \) firms are still inattentive and have their price constant at the steady state, while the remainder of firms are attentive and setting the price optimally according to 19. It follows that the aggregate price index is

$$p_t^{1-\epsilon} = \left( \theta^f \right)^t p_{ss}^{1-\epsilon} + \left( 1 - \left( \theta^f \right)^t \right) (P_t^*)^{1-\epsilon}$$

which therefore inherits sluggishness. Similarly, since firms that are inattentive are not doing any net investment, aggregate investment is sluggish.
A.3 Inattentive unions

We start by describing attentive unions. At any time $t$, union $k$ sets its wage for next period $W_{kt}$ to maximize, on behalf of all the workers it employs,

$$\sum_{\tau \geq 0} \beta^{t+\tau} \left( \int \{ u(c_{it+\tau}) - v(n_{it+\tau}) \} d\Psi_{it+\tau} - \frac{\psi}{2} \left( \frac{W_{k,t+\tau}}{W_{k,t+\tau-1}} - 1 \right)^2 \right)$$

taking as given the initial distribution of households over idiosyncratic states $\Psi_{it}$ as well as the demand curve for tasks emanating from the labor packers, which is

$$N_{kt} = \left( \frac{W_{kt}}{W_t} \right)^{-\epsilon} N_t$$

where $W_t = \left( \int W_{kt}^{-\epsilon} d\epsilon \right)^{-\frac{1}{\epsilon}}$ is the price index for aggregate employment services.

Each union is infinitesimal and therefore only takes into account its marginal effect on every household’s consumption and labor supply. Household total real earnings are

$$z_{it} = (1 - \tau_t) \frac{W_t}{P_t} N_t e_{it}$$

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

$$\frac{\partial z_{it}}{\partial W_{kt}} = (1 - \tau_t) \frac{e_{it}}{P_t} \left\{ N_{kt} - W_{kt} \epsilon \left( \frac{1}{W_t} \right)^{-\epsilon} N_t W_{kt}^{-\epsilon-1} \right\} = (1 - \tau_t) \frac{e_{it}}{P_t} N_{kt} (1 - \epsilon)$$

On the other hand, household $i$’s total hours worked are

$$n_{it} = \int_0^1 \left( \frac{W_{kt}}{W_t} \right)^{-\epsilon} N_t dk$$

which falls when $W_{kt}$ increases according to

$$\frac{\partial n_{it}}{\partial W_{kt}} = -\epsilon \frac{N_{kt}}{W_{kt}}$$

The first-order condition of the union with respect to $W_{kt}$ is therefore, enforcing constant $N_t$,

$$\frac{\epsilon}{\psi} \left\{ N_t v'(N_t) - \frac{\epsilon - 1}{\epsilon} (1 - \tau_t) \frac{W_t}{P_t} N_t u'(C_t^*) \right\} - \psi \left( \frac{W_{kt}}{W_{kt-1}} - 1 \right) \frac{1}{W_{kt-1}} + \beta \psi \left( \frac{W_{k,t+1}}{W_{k,t}} - 1 \right) \frac{1}{W_{k,t}} = 0 \quad (20)$$
where we have defined the virtual consumption aggregate $C^*_t$ such that
\[ u'(C^*_t) = \int_i e_{it} u'(c_{it}) d\Psi_{it} \]

Inattentive unions do the same but assumes that $N_{t+k}$, $\tau_{t+k}$, and $C^*_t$ will remain constant at their steady-state level starting at $k \geq 1$. Following the same argument as in section A.2, inattentive unions do not change their wage until they become attentive, while the wage path of wage adjusters satisfies, to first order:
\[
\pi^w_{t,\text{adjust}} = \kappa^w \left\{ \frac{1}{\phi} \frac{dN_t}{N} + \frac{1}{\nu} \frac{dC^*_t}{C^*} - \left( \frac{d\tau_t}{\tau^*} + \frac{dW_t/P_t}{W_t/P_t} \right) \right\} + \beta \pi^w_{t+1}
\]

where $\kappa^w = \frac{\epsilon}{\phi N' v(N)}$, $\phi$ is the Frisch elasticity of labor supply, and $\nu$ the elasticity of intertemporal substitution in consumption.

B Alternative models

B.1 Habit RA model

We consider the problem of section 2.1:
\[
\max \mathbb{E} \left[ \sum \beta^t u (c_t - \gamma c_{t-1}) \right]
\]
\[ c_t + a_t = y_t + (1 + r_{t-1}) a_{t-1} \]

where $0 \leq \gamma < 1$.

The consumer’s Euler equation here is:
\[ u' (c_t - \gamma c_{t-1}) - \beta bu' (c_{t+1} - \gamma c_t) = \beta (1 + r_t) \left( u' (c_{t+1} - \gamma c_t) - \beta \gamma u' (c_{t+2} - \gamma c_{t+1}) \right) \]

Consider a small perturbation around a steady state with constant consumption $c$ and $\beta^{-1} = (1 + r) = R$. Then, for any $dz, dr$,
\[
-\gamma dc_{t-1} + (1 + \beta \gamma (\gamma + R)) dc_t - (\beta (\gamma + R) + \beta^2 \gamma^2 R) dc_{t+1} + \beta^2 \gamma R dc_{t+2} = (\beta^2 R - \beta (\gamma + R)) (1 - \gamma) cdr_t
\]

eliminating $R = \beta^{-1}$ we obtain
\[
-\gamma dc_{t-1} + (1 + \gamma + \beta \gamma^2) dc_t - (1 + \beta \gamma + \beta \gamma^2) dc_{t+1} + \beta \gamma d c_{t+2} = - (1 - \beta + \beta \gamma) (1 - \gamma) cdr_t
\]

We can rewrite (21) as
\[
P(L) (1 - L) dc_t = -\kappa dr_t
\]
where \( \kappa \equiv (1 - \beta + \beta \gamma)(1 - \gamma) \), where

\[
P(X) = \frac{\gamma X^2 - (1 + \beta \gamma^2)X + \beta \gamma}{X^2} \]

\[
= \frac{b(X - \beta \gamma)(X - \frac{1}{\gamma})}{X^2}
\]

since \( \gamma < 1 \), this gives two roots, one greater and one smaller than 1. The solution to (22) is then

\[
(1 - \gamma L)(dc_{t+1} - dc_t) = \kappa \sum (\beta \gamma)^k dr_{t+k}
\]

Hence the change in consumption follows an AR(1) with root \( \gamma \),

\[
dc_{t+1} - dc_t = \gamma (dc_t - dc_{t-1}) + \kappa \sum (\beta \gamma)^k dr_{t+k}
\]

To obtain the level of consumption, use the budget constraint

\[
dc_t = dy_t + \frac{1}{\beta} da_{t-1} - da_t + adr_{t-1}
\]

which implies directly

\[
\sum_{t=0}^{\infty} \beta^t dc_t = \sum_{t=0}^{\infty} \beta^t (dy_t + adr_{t-1})
\]

**Intertemporal marginal propensities to consume.** If \( dr_t = 0 \) for all \( t \), then we can solve (23) to obtain

\[
dc_t = \frac{1 - \gamma^{t+1}}{1 - \gamma} dc_0
\]

We then plug in (26)

\[
\frac{1}{1 - b} \left( \sum_{t=0}^{\infty} \beta^t \left( 1 - \gamma^{t+1} \right) \right) dc_0 = dW
\]

leading to the simple expression the the MPC out of income at any date

\[
\frac{dc_0}{dW} = (1 - \beta)(1 - \beta \gamma)
\]

hence we obtain all iMPCs in closed form as

\[
\frac{\partial c_t}{\partial z_s} = (1 - \beta)(1 - \beta \gamma) \beta^s \frac{1 - \gamma^{t+1}}{1 - \gamma}
\]

from the results in the main text follow.
B.2 Heterogeneous-agent habit model

The heterogeneous-agent habit problem can be formulated as follows

\[
V(a, c_-, s) = \max_{c, a'} u(c - \gamma c_-) + \beta E[V(a', c, s') | s] \\
\text{subject to } c + a' = y(s) + Ra
\]

The first-order conditions for \(c\) and \(a'\) are

\[
\lambda = u'(c - \gamma c_-) + \beta E[V_c(a', c, s') | s]
\]

(28)

\[
\lambda = \beta E[V_a(a', c, s') | s]
\]

(29)

while the envelope for \(a\) and \(c_-\) imply

\[
V_a(a, c_-, s) = \lambda R
\]

(30)

\[
V_c(a, c_-, s) = -\gamma u'(c - \gamma c_-)
\]

(31)

We solve this model on a grid for \((a, c_-)\) to obtain the impulse response in figure 3.