ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Consumption Response Heterogeneity and Dynamics with an Inattention Region

Jérémy Boccanfuso

Quaderni - Working Paper DSE N°1172

Consumption Response Heterogeneity and Dynamics with an Inattention Region

Jérémy Boccanfuso[☆]

June 2022

Abstract

A theory in which the timing of consumer expectation adjustments is endogenously state-dependent and stochastic is proposed. These expectation adjustments generate highly heterogenous consumption responses to income windfalls: many households do not respond, those who do over-react, the marginal propensity to consume depends on windfall size and is asymmetric. We document these features in the Bank of England survey of consumers and find that they simultaneously rule out most previous explanations for these effects, including consumption adjustment cost and liquidity constraints. At the aggregate level, consumption is less sensitive to expansionary policies during recessions and its excess smoothness varies significantly over the business cycle with consumers' attention, a feature that we document in US data.

Keywords: Consumption, expectation adjustments, asymmetries, excess smoothness.
JEL Codes: D11, D84, E21.

[☆]Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126, Bologna (BO). E-mail: jeremy.boccanfuso@unibo.it. I am extremely grateful to Philippe Andrade, Olivier Compte, Lena Dräger, Antoine Ferey, Gaetano Gaballo, Xavier Gabaix, Erwan Gautier, Xavier Ragot, Mirko Wiederholt, Basit Zafar anonymous referees and participants at the AEA, CEF and EEA annual meetings, the Second Behavioral Macroeconomics Workshop, seminars at the University of Amsterdam, the Bank of France, Italy and Latvia, the University of Bologna, and the Paris School of Economics for helpful comments and suggestions. An early version of the paper was awarded at the CEF Graduate Student Paper Contest.

Non-Technical Summary

Following an income shock, many individuals do not adjust their consumption, those who do over-react, and the probability to adjust consumption depends on the magnitude and sign of the shock. These adjustment frictions shape the size and dynamics of consumption responses, which are central to the transmission of monetary and fiscal policies. They are, however, largely absent in macroeconomic models.

Motivated by recent evidence from information experiments, we show that the stochasticity and state-dependence of consumer expectation adjustments can account for these behaviors. To this end, we propose a trackable model in which a consumer faces a fixed cost for paying attention to noisy signals about her permanent income and her attention choice can be a function of signal realizations. At the optimum, the consumer faces an inattention region where she remains momentarily inattentive to income shocks, and expectation adjustments are stochastic and state-dependent. As a result, in the model, many individuals do not adjust their consumption following an income shock *because* they remain inattentive to it, those who do over-react *because* they also adjust for the shocks that they previously ignored, and the probability to adjust consumption depends on the magnitude and sign of the shock *because* of the state-dependence of attention.

We further show that the model is consistent with the size effects and asymmetries from both negative and positive income shocks at the extensive, intensive and overall responses of consumption that we document in the Bank of England survey of consumers. In particular, we identify a decreasing size effect from positive shocks which is not specific to credit, liquidity or cash-on-hand constrained consumers. This feature of consumption data rule out most previous explanations for size effects and asymmetries, including consumption adjustment cost and liquidity constraints.

Simulating shocks that are comparable to the 2008 Tax Rebates and the Alaska's Permanent Fund payments, with the latter being 4 times larger than the former, we further illustrate the aggregate implication of our theory. We predict quantitatively different dynamics following these two shocks. This is because a large shock prompts more consumers to become attentive and, thereby, to adjust consumption in response to this shock. Moreover, during bad times, consumers are more likely to be attentive to further destabilizing shocks and aggregate consumption responds more promptly to negative than to positive shocks, while the opposite is true during good times.

Finally, our theory predicts that the excess smoothness of aggregate consumption varies significantly over the business cycle with consumer inattention. We document this feature of aggregate consumption in US data.

1. Introduction

In surveys and quasi-experiments, 30 to 60 percent of households do not adjust nondurable consumption following an income shock (Misra and Surico, 2014; Bunn et al., 2018; Christelis et al., 2019; Fuster et al., 2021; Andreolli and Surico, 2021). These adjustment frictions shape the size and dynamics of consumption responses, which are central to a broad set of macroeconomic mechanisms including the transmission of monetary and fiscal policies. Despite their potential importance, these adjustment frictions are largely absent in consumption models and their consequences unexplored.[1]

Some authors have hypothesized that consumers' cognitive cost to process information may cause these adjustment frictions in consumption (Caballero, 1995; Browning and Collado, 2001; Fuster et al., 2021). Consistent with this view, evidence indicates that consumers do not continuously adjust their expectations, but rather that expectation adjustments are sporadic, with consumers discretely updating their expectations to incorporate continuous information. In particular, the timing of expectation adjustments is stochastic and state-dependent – as illustrated by the strictly positive and U-shape hazard function reported in Khaw et al. (2017) and Henckel et al. (2021).[2]

Our first contribution is to propose a dynamic model of rationally inattentive consumers consistent with these features of expectation adjustments. Our theory builds on sticky expectation models in which consumers face a fixed cost to be attentive, i.e., to observe information (Gabaix and Laibson, 2001; Reis, 2006; Carroll et al., 2020). These models predict that consumers adjust their expectations on a calendar basis or with a constant probability, which is inconsistent with the aforementioned evidence on expectation adjustments. Hence, we extend these models in two directions. First, we introduce an information friction by assuming that a consumer can observe noisy Gaussian signals about her permanent income when attentive, but not the true value. Second, we allow the consumer to condition her attention behavior on signal realizations. Otherwise, the consumption-saving problem coincides with Hall's (1978) random walk model with quadratic utility and Gaussian income shocks.

[1] Exceptions are Caballero (1995) and Fuster et al. (2021) who consider fixed costs to adjust nondurable consumption. The former demonstrates that consumption adjustment frictions can explain the excess smoothness and sensitivity of aggregate consumption. The latter finds that consumption adjustments are important determinants of household consumption response heterogeneity in survey data.

[2] Randomized controlled trials and laboratory experiments also report that consumer expectation adjustments are affected by the provision of free and salient information, imperfect, and both large and small (Armantier et al., 2016; Khaw et al., 2017; Armona et al., 2018; Roth and Wohlfart, 2020; Henckel et al., 2021). Our model is consistent with these features. We discuss these empirical findings and their relation to existing models of inattention in Section 3.

At the optimum, the attention behavior is governed by an inattention region with respect to the expectation wedge between a consumer's current beliefs under inattention and the latent beliefs under continuous attention. Consequently, a consumer is inattentive to new signal realizations most of the time and it is only when (the absolute value of) the expectation wedge is larger than some endogenously defined threshold that she becomes attentive. When attentive, she catches up with the signals she previously ignored to close the expectation wedge.

In the absence of information frictions, the expectation wedge coincides with the expectation error, i.e., the difference between the consumer's true and expected permanent income. It is, therefore, an observable variable and the attention behavior is purely state-dependent with a probability that jumps from 0 to 1. However, in the presence of information frictions in the form of noisy signals, the expectation wedge is unobservable and only partially correlates with the consumer's expectation error (or other observable variables). Hence, consistent with the aforementioned evidence, the timing of expectation adjustments is stochastic and state-dependent, with a strictly positive and U-shape hazard function. We further illustrate that the inattention region can generate attention behaviors that are *simultaneously* stochastic, state- and time-dependent, as well as smooth unimodal and bimodal distributions of expectation revisions.

Our second contribution is to illustrate how these expectation adjustments generate consumption adjustments that are consistent with household consumption data. In the model, many households do not adjust consumption following an income shock because of inattention, the probability to adjust consumption correlates with income shock size because of the state-dependent nature of expectation adjustments, and the marginal propensity to consume (MPC) of adjusting consumers is large because attentive consumers over-react to income shocks by also closing their expectation wedge. These predictions are consistent with evidence on reported spending from hypothetical scenarios (Fuster et al., 2021) and, as we find, from experienced shocks in the Bank of England surveys of consumers. They cannot be attributed to precautionary motives, borrowing and liquidity constraints, non-homothetic preferences, nor sticky expectations (Kaplan and Violante, 2014, 2022; Jappelli and Pistaferri, 2017; Carroll et al., 2020; Andreolli and Surico, 2021).

Arguably, a fixed cost to adjust consumption could also reproduce these features of consumption adjustments (Caballero, 1995; Chetty and Szeidl, 2016; Fuster et al., 2021). However, the difficulty in matching the aforementioned micro evidence on sporadic adjustments to income shocks is to preserve indisputable features of household nondurable consumption dynamics. Consumption adjustment costs imply that household consumption should remain anchored at a constant level (or trend) for many periods and, occasionally, jump by a lot to reach a new anchor.[3] Such jerky behavior has been forcefully criticized in the context of nondurable consumption (Reis, 2006; Carroll et al., 2020). Instead, in our model, consumers behave essentially as if they were systematically attentive and adjusting consumption to idiosyncratic income shocks at a quarterly frequency, despite being largely inattentive to aggregate shocks at this frequency.[4]

Our third contribution is to offer an explanation to seemingly conflicting findings about size effects and asymmetries from positive and negative income shocks reported across the literature. Using consumer surveys, Bunn et al. (2018), Christelis et al. (2019) and Fuster et al. (2021) report that MPC from negative income shocks are larger than MPC from positive shocks on average, a so called *negative* asymmetry. This finding is generally interpreted as a consequence of liquidity constraint and precautionary motive, but Fuster et al. (2021) find that these explanations cannot quantitatively account for all the negative asymmetry found in their data – suggesting that there is missing negative asymmetry. In contrast, Ballantyne (2021) identifies a missing positive asymmetry after controlling for liquidity constraints.[5] Similarly, existing studies report opposite size effects from positive shocks: Christelis et al. (2019) and Fagereng et al. (2021) find an average MPC decreasing with shock size, while Fuster et al. (2021) find an average MPC increasing with size.

Our model provides an explanation for the missing asymmetry and shifts in size effects. The consumption function is linear everywhere in our model. Thus, nonlinearities in consumption responses arise solely from the extensive margin of expectation adjustments. Size effects are due to the state-dependence of consumers' attention and asymmetries are a consequence of consumers' prior localization in the inattention region. Hence, size effects and asymmetries can reverse at the individual level depending on a consumer's expectation wedge. Averaging across consumers, size effects and asymmetries in the cross-sectional MPC are governed by the distribution of expectation wedges and can also reverse as aggregate income shocks shift this distribution.

[3] A similar behavior emerges in the entropy-based model of Tutino (2013).

[4] Carroll et al. (2020) review evidence that consumers appear to be attentive to idiosyncratic income shocks but not aggregate shocks in quarterly data. This feature is assumed in their model, whereas it arises endogenously here.

[5] Ballantyne (2021) reports an absence of asymmetry on average resulting from a positive asymmetry for financially secure households and a negative asymmetry for households who are likely to face financial frictions.

An ideal test of our theory on cross-sectional data would require to elucidate the distribution of expectation wedges. These wedges are, however, unobservable and we cannot infer this distribution. Nevertheless, we show that knowledge of the sign of the first moment of this distribution is generally sufficient to characterize size effects for the probability to adjust consumption (extensive margin), the conditional MPC of consumers who adjust (intensive margin), and unconditional MPC (overall response). Importantly, size effects reverse across margins for a given sign of the mean wedge, and within margins depending on the sign of the mean wedge. For instance, when the mean wedge is negative, our model predicts a probability to adjust consumption (P1) and an unconditonal MPC (P2) increasing with the size of negative income shocks, while the size effect associated to positive shocks is unclear. In opposition, the conditional MPC always displays decreasing size effects from both negative and positive income shocks (P3), and is larger (with a discontinuity at zero) for negative income shocks (P4).[6]

We are not aware of previous work estimating size effects from *both* negative and positive income shocks at the extensive, intensive and overall responses of consumption.[7] We make progress in this direction using the 2012-2014 Bank of England surveys of consumers, previously analyzed in Bunn et al. (2018), in which respondents are asked to report their perceived income surprise over the last year and how they adjusted their spending to it. The timing of the survey coincides with a period of economic downturn in the UK economy, suggesting that the mean expectation wedge was negative during this period. Controlling for households observable characteristics, we find suggestive evidence for all predictions P1-P4. Our conclusions are unaffected when focusing on households who are the less likely to face binding borrowing and liquidity constraints. Furthermore, our data are characterized by a state-dependent extensive margin and decreasing size effect from positive shocks, features of consumption responses that we show to simultaneously rule out most previous explanations for size effects and asymmetries.

Our fourth contribution is to illustrate how the size effects and asymmetries induced by expectation adjustments affect the dynamics of aggregate consumption. We illustrate the dynamic consequences of size effects by simulating shocks that are comparable to

[6]These predictions hold for linear size effects. We asses the strength of these predictions using Monte Carlo simulations. We also extend our analysis to nonlinear size effects.

[7]Fuster et al. (2021) analyze all margins of consumption responses, but estimate size effects only for positive shocks. We explain in Section 4 how their results are a weaker test of, though consistent with, our theory.

the 2008 Tax Rebates (Parker et al., 2013) and the Alaska's Permanent Fund payments (Hsieh, 2003; Kueng, 2018), with the latter shock being 4 times larger than the former. We predict quantitatively different dynamics following these two shocks. This is because a large shock prompts relatively more consumers to become attentive and, thereby, to adjust consumption in response to this shock, resulting in sharper aggregate consumption response from larger shocks. Thus, our theory offers novel insights to understand the heterogenous effects of fiscal stimuli reported across the literature.[8] Our theory also provides novel insights regarding asymmetries to positive and negative aggregate shocks along the business cycle. During bad times, consumers are more likely to be attentive to further destabilizing shocks and aggregate consumption responds more promptly to negative than to positive shocks, while the opposite is true during good times. These asymmetric dynamics have been previously documented in US aggregate consumption data (Caballero, 1995) and could participate in explaining the asymmetry and state-dependence of fiscal multipliers (Barnichon et al., 2022).

Finally, we document and explain significant variations in the persistence of aggregate consumption related to business cycle fluctuations. Consistent with the meta-analysis of Havranek et al. (2017), the inattention region implies that consumption is highly persistent, or 'excessively smooth', at the aggregate level but not at the household level. This feature is not specific to our model and, for instance, also emerges in consumption models with random attention as in Carroll et al. (2020). A distinctive feature of our theory is, however, that aggregate consumption excess smoothness varies over time depending on the share of inattentive consumers. We document these variations in the persistence of US consumption data. The share of inattentive consumers is not directly observable in the data. Therefore, we propose two proxies to identify periods when these shares are the largest. The first relates to the deviation of lagged consumption growth from its mean and the second coincides with recessionary periods.[9] In accordance with our prediction, we find that US consumption persistence drops drastically during these periods. Our conclusion is robust to different set of instruments and weak-instrument inference. It contributes to the prominent literature on consumption excess smoothness initiated by Campbell and

[8]See Jappelli and Pistaferri (2010), Parker et al. (2013) or Carroll et al. (2017) for a review. The dynamic size effect that we document relates to the magnitude hypothesis that postulates a negative relation between excess sensitivity and shock size (Browning and Collado, 2001; Jappelli and Pistaferri, 2010; Scholnick, 2013).

[9]We use these proxies based on the model predictions. The second proxy is also supported by the evidence that information rigidities drop persistently in the aftermath of a recession in surveys of professional forecasters (Dräger and Lamla, 2012; Coibion and Gorodnichenko, 2015).

Deaton (1989) and we are not aware of previous work explaining this feature of aggregate consumption dynamics. It holds important implications beyond the realm of consumption because aggregate consumption persistence is a key driver of sluggish dynamics in DSGE models.

The paper is organized as follows. Section 2 introduces and solves the consumer problem. Section 3 discusses the implications of the inattention region for expectation adjustments and revisions. Section 4 analyses size effects and asymmetries in household consumption. Sections 5 and 6 focus on aggregate consumption dynamics. All proofs are relegated to the Appendix.

2. Consumer problem

This section first introduces the consumer problem using a two-agents analogy. It then formally sets and solves the problem.

2.1. Informal description of the consumer problem: A two-agents analogy

Consider the textbook random walk model of consumption (Hall, 1978) with quadratic utility and Gaussian innovations. We extend this model to account for recent evidence on consumer inattention. To this end, consider that the household is composed of two individuals. The first individual, the 'consumer', is in charge of buying consumption goods. The only information she has comes from the second individual. This second individual, the 'worker', has private information about realizations of the stochastic components of the household's income. The worker's information can be imperfect. Both individuals share the common objective of maximizing the household's intertemporal utility, but they dislike meeting to share information (fixed utility cost).

The structure of the household raises novel questions: When should the two individuals meet? What is the information disclosed during meetings? Can the consumer infer some information between meetings? Does it affect consumption between meetings?

In the rest of this section, we demonstrate the following results. First, the consumer and worker problems are separable. Hence, the certainty equivalence holds and the consumption policy is unaffected. Second, the worker manages the frequency of the meetings. She calls for a meeting whenever her expected permanent income is too different from the one of the consumer. Thus, a symmetric 'no meeting' region emerges. Fourth, the worker discloses her expected permanent income and the consumer revises the consumption plan accordingly during meetings. Last, between meetings, the consumer realizes that the worker's information does not trigger a meeting. Nevertheless, it has no effect on consumption choices between meetings (because the 'no meeting' region is symmetric).

The two-agents analogy discussed here is useful insofar that it helps to clarify the information structure considered in the present paper. Arguably, it could as well illustrate the communication between two separate parts of the brain – one for observing new information and another for acting upon new information – instead of two individuals. More generally, it captures the behavior of a consumer who sometimes remains inattentive to new information. This is the interpretation that we retain in the rest of the paper.

2.2. Formalizing the consumer problem

Utility and income – Consider the problem of a consumer with memory who lives from period 0 to period $T-1$. She consumes c_t each period and her utility is quadratic $u(c_t) = -(c_t - \bar{c})^2$ with $\bar{c} \in \mathbb{R}^+$ a consumption bliss point.[10] This agent discounts future utility by the factor $\beta \in (0,1)$ and can borrow and lend freely at the gross interest rate $1+r$. At each period, she receives an exogenous stochastic income y_t which follows from a multivariate linear state space model with Gaussian white noise innovations. The consumer's budget constraint writes $a_{t+1} = (1+r)a_t - c_t + y_t$ where a_t are asset holdings.

Permanent income – We reformulate the consumer's problem in terms of permanent income. To this end, let $s_t \equiv a_t + \mu_t - \sum_{t=0}^{T-1}(1+r)^{-t}\bar{c}$ be the consumer's permanent income net of a constant consumption stream equal to $\bar{c}$ at each period.[11] $\mu_t \equiv \sum_{k=t}^{T-1} E_t[(1+r)^{(t-k-1)}y_k]$ is the discounted expected present value of current and future incomes. Accordingly, the period budget constraint may be written in terms of permanent income: $s_{t+1} = (1+r)s_t - (c_t - \bar{c}) + \zeta_{t+1}$ where $\zeta_{t+1} \equiv \sum_{k=t+1}^{T-1}(1+r)^{t-k}\Big(E_{t+1} - E_t\Big)[y_k]$ is the innovation to permanent income. The assumption on the income process implies that this innovation is a Gaussian white noise with variance σ_ζ^2.

Terminal condition – In the following, we will express the problem as a linear-quadratic control problem. Therefore, we impose a terminal condition to account for the no-Ponzi game and transversality conditions. To this end, we incorporate the terminal condition as a large utility penalty $q_T s_T^2$ when the consumer dies at time T.[12] In the rest of the paper,

[10] The quadratic utility assumption allows to derive an analytical solution and is a standard framework for the study of rationally inattentive consumers (Sims, 2003; Luo and Young, 2014; Carroll et al., 2020; Miao et al., 2022).

[11] Incorporating the constant consumption stream in the state variable allows to directly take $u_t \equiv c_t - \bar{c}$ as a control variable. Note that since $\bar{c}$ is known to the consumer, it is equivalent to choose u_t or c_t. Consequently, it plays no other role than simplifying the notation when deriving the solution.

[12] Intuitively, we may want to impose the standard condition with finite horizon that $s_T = 0$ in expectation given the consumer information. However, the information structure considered thereafter does

we focus on the infinite horizon limit of the problem so that this penalty has a negligible effect on optimal policies.

Noisy information – Because of information frictions, the consumer never perfectly observes permanent income s_t. Nevertheless, she receives imperfect information about s_t in the form of additive noisy signals $z_t = s_t + \vartheta_t$ with i.i.d. Gaussian white noises the variance of which is σ_ϑ^2. The smaller the variance of the noise σ_ϑ^2 is, the more informative the signal is. We collect these signals in a latent information set, denoted $\mathcal{I}_t$. Using the previously discussed two-agents analogy, it corresponds to the worker's information.

Information at the attention choice – When the consumer is attentive, she can use the information in $\mathcal{I}_t$ for her consumption choices. However, we assume that attention is cognitively costly and we model this cost as a fixed utility cost denoted λ. We allow the attention choices to depend on the information in $\mathcal{I}_t$. Without lack of generality, the information set $\mathcal{I}_t$ also contains past actions of the consumer. These actions are twofolds. They consist in the consumption choices (c_t) and whether the consumer was attentive ($\tau_t = 1$) or not ($\tau_t = 0$). Hence, the information set at the attention choice writes

$$\mathcal{I}_t \equiv \{z_0, \tau_0, c_0, \ldots, z_{t-1}, \tau_{t-1}, c_{t-1}, z_t\} \tag{1}$$

Information at the consumption choice – The information set at the consumption choice, denoted $\overline{\mathcal{I}}_t$, is not necessarily incremented at each period by the signal z_t as the consumer can be inattentive. Instead, let $\bar{z}_t$ be the incremental information at the consumption choice. Thus, we write

$$\overline{\mathcal{I}}_t \equiv \{\bar{z}_0, \tau_0, c_0, \ldots, \bar{z}_{t-1}, \tau_{t-1}, c_{t-1}, \tau_t, \bar{z}_t\} \tag{2}$$

By definition, $\bar{z}_t = \emptyset$ is empty when the consumer is inattentive ($\tau_t = 0$). When the consumer is attentive, the information conveyed in $\bar{z}_t$ is a priori not clear. It could, for instance, be the last signal that was received z_t, a sequence of past signals, or a filtration of these signals. Therefore, we remain agnostic about the form of $\bar{z}_t$. The only restriction we impose is that this information must come from the latent information set $\mathcal{I}_t$. That

not preclude situations where the consumer prefers to remain inattentive for a long period of time before dying. Consequently, a terminal condition $s_T = 0$ is not strong enough to prevent exploding deviations in terms of realized permanent income. Hence, we may want to unsure that the consumer finds it optimal to use all the information she can access at period $T-1$ to satisfy the terminal condition. When $\beta^T q_T$ is arbitrarily large, a rationally inattentive consumer decides to observe all the information she can to adjust her consumption at period $T-1$. It thus prevents the type of deviations that we have just mentioned.

is, we impose that $\sigma(\{\bar{z}_k\}_{k=0}^t) \subseteq \sigma(\{z_k\}_{k=0}^t)$ where $\sigma(.)$ denotes a σ-algebra. Following Molin and Hirche (2010), we refer to this property as the nestedness of the information structure.[13]

Problem formulation – Collecting the consumption and attention parts, the consumer problem writes as follows:

$$\begin{aligned}
\min_{\{c_t,\tau_t\}_{t=0}^{T-1} \in \mathbb{R}^T \times \{0,1\}^T} \quad & E_0\left[\sum_{t=0}^{T-1} \beta^t \Big((c_t - \bar{c})^2 + \lambda \tau_t\Big) + \beta^T q_T s_T^2 \middle| \Big\{\mathcal{I}_t, \overline{\mathcal{I}}_t\Big\}\right] \qquad (3)\\
\text{s.t.} \quad & s_{t+1} = (1+r)s_t - (c_t - \bar{c}) + \zeta_{t+1}\\
& c_t = f_t(\overline{\mathcal{I}}_t); \tau_t = g_t(\mathcal{I}_t)\\
& s_0 | \mathcal{I}_0 \sim \mathcal{N}(s_0, \sigma_{s_0}^2) \quad ; \; s_0 \text{ and } \sigma_{s_0}^2 \text{ given}
\end{aligned}$$

Problem (3) states that the consumer maximizes her intertemporal utility given the aforementioned period budget constraint and information structure. The instantaneous utility is quadratic and the consumer must pay a fixed utility cost λ whenever she decides to be attentive. The presence of the term $\beta^T q_T s_T^2$ follows from the discussion on the terminal condition. The consumption choice depends on the information set $\overline{\mathcal{I}}_t$, while the attention choice depends on $\mathcal{I}_t$. Formally, this implies that the policies $f_t(.)$ and $g_t(.)$, which respectively refer to the consumption and attention choices, are Borel-measurable functions with respect to $\overline{\mathcal{I}}_t$ and $\mathcal{I}_t$. The last condition characterizes the initial uncertainty surrounding the perceived permanent income.

Comparison to some related models – The above formulation encompasses some well-known models of inattentive consumers as limiting cases. When the attention cost λ is nil, the consumer is always attentive to the Gaussian signals. Hence, Problem (3) collapses to a consumption problem with noisy information (Sims, 2003; Luo and Young, 2014). When the signals are noiseless ($z_t = s_t$) and the information at the attention choice is restricted to $\overline{\mathcal{I}}_t$, Problem (3) collapses to a consumption problem with sticky information (Reis, 2006).[14] Problems related to (3) have been studied recently in engineering. The closest papers are Molin and Hirche (2010, 2017) who also study a discrete-time linear-quadratic Gaussian control problem with a similar information structure. The only deviation with

[13] In Section 3.3, we discuss an extension breaking the nestedness of the information structure by assuming that the consumer observes s_t when attentive.

[14] Because of this restriction, Reis (2006) finds that the optimal attention behavior is purely time-dependent. Reis (2006) argues that consumers may, nevertheless, be attentive to extreme events. By allowing $\tau_t = g_t(\mathcal{I}_t)$, Problem (3) is designed to allow consumers to be attentive to the (extreme) events that they deem attention worthy.

respect to their framework is that we introduce discounting in the objective function.

2.3. Consumption and expectations

The detailed solution to Problem (3) is relegated to Appendix A. In the following, we present main results for consumption and attention behaviors. The last subsection concludes with the main proposition.

Consumption policy – The attention and consumption choices are separable. This separation result arises as a consequence of the nestedness of the information structure and the quadratic utility. On the one hand, the nestedness of the information structure implies that the information available at the consumption choice is observable at the attention choice. As a result, any attempt to adjust consumption to affect the attention behavior would be vain. To demonstrate this, we show that attention policies are functions of the random variables realizations and are, thereby, independent of consumption policies. On the other hand, the quadratic structure of the problem implies that *for any* given attention policy, the optimal consumption policy is the certainty equivalent one. Finally, we can always retrieve a new admissible attention policy, coherent with the certainty equivalent policy for consumption, that will dominate the initial one; Hence, the certainty equivalent policy is optimal *irrespectively* of the attention strategy.

Lemma 1 (Certainty equivalence)**.** *The certainty equivalence holds and consumption is*

$$c_t = L_t E[s_t|\overline{\mathcal{I}}_t] + \bar{c} \qquad \forall t \in 0, \dots, T-1 \tag{4}$$

where $L_t \equiv (1+r)\beta p_{t+1}/(1+\beta p_{t+1})$ *and* p_t *follows from iterating on the backward Riccati equation* $p_t = (1+r)^2\beta p_{t+1}/(1+\beta p_{t+1})$ *with terminal condition* $p_T = q_T$*.*

The consumption policy is not affected by the information structure. Moreover, we retrieve the well-known result that the consumption function is linear when the consumer's utility is quadratic. L_t is the marginal propensity to consume with respect to expected permanent income. It depends solely on the discount rate β, the interest rate r, the time horizon T, and the terminal condition q_T.

Expectations at the attention choice – We next characterize expectations at the attention choice. Following common practice in the literature (e.g. Sims (2003), Luo et al. (2017) and Maćkowiak et al. (2018)), we assume that the initial uncertainty surrounding the state variable $\sigma^2_{s_0}$ is at its steady state value. Consequently, the optimal expectation $E[s_t|\mathcal{I}_t]$ is the least-squares estimator. Thanks to the linearity of the state dynamics and the i.i.d. Gaussian noises and innovations, the least-squares estimator is the Kalman filter. The following Lemma characterizes the dynamics of expectations at the attention choice.

Lemma 2 (Kalman filter)**.** *The optimal estimate of s_t at the attention choice is*

$$E[s_t|\mathcal{I}_t] \quad = \quad (1-K)\Big((1+r)E[s_{t-1}|\mathcal{I}_{t-1}] - c_{t-1} + \bar{c}\Big) + Kz_t \tag{5}$$

where K is the steady state Kalman gain defined in Appendix A.2

Expectations at the consumption choice – When the consumer is attentive ($\tau_t = 1$), she can access the information contained in $\mathcal{I}_t$ to update her expectation $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1]$. Therefore, $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1] = E[s_t|\mathcal{I}_t]$ because the latter expectation is optimal given $\mathcal{I}_t$ and the nestedness property of the information structure implies that there is no other source of information.

Characterizing expectations while the consumer remains inattentive ($\tau_t = 0$) is more involved. Indeed, when inattentive, the consumer realizes that she is inattentive. When the choice to become attentive is not random, this realization conveys some information that the consumer can use to revise her expectations while inattentive. This generates a corrective term in the expectation at the consumption choice

$$\underbrace{E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0]}_{\text{expectation when inattentive}} \quad = \quad \underbrace{E[s_t|\overline{\mathcal{I}}_{t-1}]}_{\text{mechanical update}} \quad + \underbrace{E\Big[(1+r)e_{t-1} + K(z_t - E[s_t|\mathcal{I}_{t-1}])\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big]}_{\text{corrective term } \alpha(.)} \tag{6}$$

where $e_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t]$ is the expectation wedge, i.e., the difference between expectations at the attention and consumption choices. Hence, the corrective term, denoted $\alpha(.)$, corresponds to the consumer's expected wedge between the two expectations when she is inattentive. The corrective terms are intrinsically related to the attention policy.[15] Nevertheless, since these corrective terms are measurable from the consumer's information set when she is inattentive, they can vary only with respect to two inputs: the time period t and the last period when the consumer was attentive l_t. We therefore write $\alpha(t, l_t)$. The following Lemma characterizes the consumer's expectation depending on whether she is attentive $\tau_t = 1$ or not.

[15] To illustrate this dependence, suppose that the consumer is willing to be attentive to negative news only. Implicitly, this attention strategy would imply that a news is positive when the consumer remains inattentive to it, as it would have triggered her attention otherwise. The consumer would therefore infer that, on average, her permanent income is higher than what would be implied by the mechanical update in equation (6) at periods when she is inattentive. In this scenario, the corrective terms in equation (6) would be positive.

Lemma 3 (Perceived permanent income)**.** *The optimal estimate of s_t given the information set at the consumption choice is*

$$E[s_t|\overline{\mathcal{I}}_t] = \begin{cases} E[s_t|\mathcal{I}_t] & \text{if } \tau_t = 1 \\ (1+r)E[s_{t-1}|\overline{\mathcal{I}}_{t-1}] - c_{t-1} + \bar{c} + \alpha(t, l_t) & \text{if } \tau_t = 0 \end{cases} \tag{7}$$

with $\alpha(t, l_t) = E\Big[(1+r)e_{t-1} + K(z_t - E[s_t|\mathcal{I}_{t-1}])\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big]$.

2.4. Optimal attention policy

The realization that the corrective terms $\alpha(t, l_t)$ only depend on t and l_t greatly simplifies the attention problem as it implies that they do not vary with the expectation wedge e_t. Hence, the attention problem can be derived for an arbitrary sequence of $\alpha(t, l_t)$ for all t and $l_t < t$, that we denote by $\boldsymbol{\alpha}$. Consequently, the optimal attention policy is the (functional fixed-point) solution to the attention problem given a sequence $\boldsymbol{\alpha}$ coherent with Equation (6) evaluated at the optimal attention policy.

Attention problem – To find the optimal attention policy, we first characterize the attention problem for an arbitrary sequence of corrective terms $\boldsymbol{\alpha}$. When inattentive, the expectation wedge e_t translates into a consumption wedge $L_t e_t$ (from Lemma 1) and, thereby, into a utility wedge. With a quadratic utility, the utility wedge is proportional to the square of the consumption wedge. When the consumer prefers to become attentive, this utility wedge vanishes but she has to pay the fixed utility cost λ.

Moreover, the expectation wedge evolves dynamically while the consumer remains inattentive. Using Lemmas 2 and 3, the dynamics of the expectation wedge e_t are

$$e_{t+1} = (1-\tau_t)(1+r)e_t + K\Big(z_{t+1} - E[s_{t+1}|\mathcal{I}_t]\Big) - \alpha(t, l_t) \tag{8}$$

That is, the expectation wedge is incremented at each period by the new information incorporated at the attention choice minus the corrective term. The wedge grows at the interest rate whilst the consumer remains inattentive. This growth arises because the expectation wedge translates into a consumption wedge and, thereby, also into an asset wedge growing at the interest rate. Importantly, Equation (8) implies that the expectation wedge e_t can be computed from $\mathcal{I}_t$. Therefore, it is observable at the attention choice at time t.

Consequently, the attention policy solves

$$
\begin{aligned}
J_t(e_t, l_t; \boldsymbol{\alpha}) &= \min_{\tau_t \in \{0,1\}} (1-\tau_t)(1+\beta p_{t+1}) L_t^2 e_t^2 + \lambda \tau_t + \beta E[J_{t+1}(e_{t+1}, l_{t+1}; \boldsymbol{\alpha}) | \mathcal{I}_t] \quad (9) \\
\text{s.t.} \quad & e_{t+1} = (1-\tau_t)(1+r)e_t + K\Big(z_{t+1} - E[s_{t+1}|\mathcal{I}_t]\Big) - \alpha(t+1, l_{t+1}) \\
& l_{t+1} = \tau_t t + (1-\tau_t) l_t
\end{aligned}
$$

for a given sequence of corrective terms $\boldsymbol{\alpha}$. This is a standard dynamic problem with two states, the expectation wedge e_t and the last period when the consumer was attentive l_t, which are both observable at the attention choice.

Solution – Assume that the corrective terms are zero for all t and $l_t < t$ (i.e. $\boldsymbol{\alpha} = \mathbf{0}$). Then, the state l_t becomes irrelevant in problem (9) and we find that the optimal attention policy is a symmetric function equal to one when the absolute value of the expectation wedge is larger than a threshold denoted π_t. That is, the consumer is attentive $\tau_t = 1 \iff |e_t| \geq \pi_t$ and inattentive otherwise at the optimum. In other words, a symmetric inattention region emerges with respect to the expectation wedge at the attention and consumption choices.

This solution is locally stable is the sense that the corrective terms, defined in Equation (6), are indeed zero when the attention policy is $\tau_t = 1 \iff |e_t| \geq \pi_t$ and zero otherwise. This follows from the symmetry of this attention policy. The intuition is as follows. Take a consumer who was last attentive one period ago. She knows that the new information incorporated at the attention choice was an innovation drawn from a Gaussian distribution with zero mean. She also knows that this innovation did not trigger her attention. Therefore, the conditional distribution of the innovation follows a truncated Gaussian. The truncation being symmetric, the conditional expectation of this innovation remains zero. Hence, the corrective terms $\alpha(t, t-1) = 0$ for all t. Using this argument recursively, we can show that this is also the case for all $l_t < t$.

Finally, Theorem 1 in Molin and Hirche (2017) allows to conclude that this solution is globally asymptotically stable. That is, starting from any $\boldsymbol{\alpha}_0$, solving the associated optimal attention policy from problem (9) and computing an updated sequence of corrective terms $\boldsymbol{\alpha}_1$ from their definition in equation (6), we asymptotically converge to $\boldsymbol{\alpha}_\infty = \mathbf{0}$. We report these results in the following Lemma.

Lemma 4 (Attention policy)**.** *Let the sequence of corrective terms $\boldsymbol{\alpha} = \mathbf{0}$. The expectation* *wedge law of motion is thus*

$$e_{t+1} = (1-\tau_t)(1+r)e_t + K\Big(z_{t+1} - E[s_{t+1}|\mathcal{I}_t]\Big) \tag{10}$$

Then, the optimal attention policy $\tau_t = g_t(e_t)$ *is symmetric and such that* $g_t(e_t) = 1 \iff |e_t| \geq \pi_t$ *and* 0 *otherwise for all* t. *The thresholds* $\pi_t \in \mathbb{R}^+$ *follow from solving* $\forall t \in \{0, \ldots, T-1\}$

$$\lambda + \beta E[J_{t+1}(e_{t+1}; \mathbf{0})|\mathcal{I}_t, e_t = 0] \;=\; L_t^2(1+\beta p_{t+1})\pi_t^2 + \beta E[J_{t+1}(e_{t+1}, \mathbf{0})|\mathcal{I}_t, e_t = \pi_t] \tag{11}$$

with $J_{t+1}(e_{t+1}; \mathbf{0})$ *defined in (9). This solution is globally asymptotically stable.*

2.5. Stationary policies

Lemmas 1-4 fully characterize the solution to problem (3). In our setup, the time-dependence of the consumption and attention policies results from accounting for the terminal condition $q_T s_T^2$. Therefore, we consider the infinite horizon limit of problem (3) where the impact of this terminal condition is negligible on the optimal policies. When the horizon T tends to infinity, the optimal policies $f_t(.)$ and $g_t(.)$ respectively converge to stationary policies $f(.)$ and $g(.)$. These stationary policies are characterized in the following Proposition.

Proposition 1. *When the horizon tends to infinity, the policy functions converges to stationary policies* $f(.)$ *and* $g(.)$. *Consequently, and assuming it exists, the consumption policy converges to*

$$c_t = \frac{\beta(1+r)^2 - 1}{\beta(1+r)} E[s_t|\overline{\mathcal{I}}_t] + \bar{c} \tag{12}$$

and the consumer updates the information set $\overline{\mathcal{I}}_t$, *that is* $\tau_t = 1$, *whenever* $|e_t| \geq \pi$ *where* $e_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0]$ *and*

$$\pi \;=\; \frac{\sqrt{\beta(1+r)\Big(\lambda + \beta(E[J(e_{t+1}; \mathbf{0})|\mathcal{I}_t, e_t = 0] - E[J(e_{t+1}; \mathbf{0})|\mathcal{I}_t, e_t = \pi]\Big)}}{\beta(1+r)^2 - 1} \tag{13}$$

$J(.; \mathbf{0})$ *is the functional fixed-point solution to the infinite horizon reformulation of the Bellman equation defined in (9).*

The stationary consumption policy (12) is standard and we retrieve the well-known result that the consumption path is constant over time when $\beta^{-1} = (1+r)$. The stationary threshold π can be computed numerically with standard dynamic programing methods. We consider these stationary policies in the rest of the paper.

3. Inattention region, state-dependence and revisions

This section presents key features related to consumers' expectations in the presence of inattention region and discusses them in light of existing empirical evidence. Technical details are relegated to Appendix B.

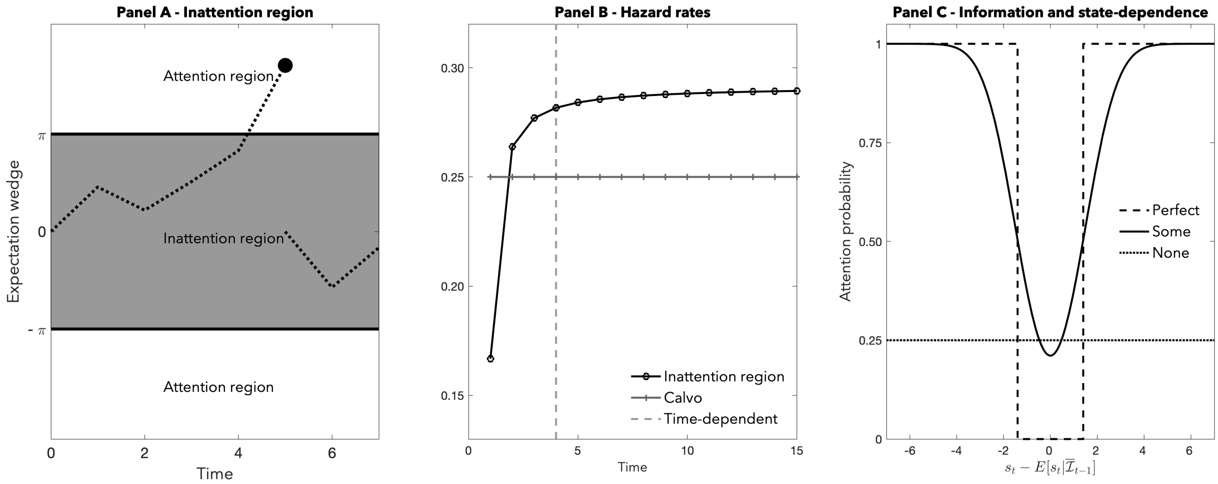
3.1. Inattention region

In the presence of an inattention region, a consumer is inattentive most of the time. Sporadically, however, the expectation wedge $|e_t|$ exceeds the threshold π which triggers her attention. Panel A in Figure 1 illustrates the dynamics behind the attention behavior. Starting from a period 0 when the consumer was attentive, the expectation wedge smoothly incorporates the signals observed at the attention choice. As long as the expectation wedge remains in the inattention region, the consumer remains inattentive to it. However, when it exceeds the lower ($-\pi$) or upper (π) threshold, the consumer becomes attentive and observes e_t. Since the consumer observes the expectation wedge e_t when attentive, she catches up with the information available at the attention choice and the dynamics of the expectation wedge restarts from zero.

Hence, a direct implication of the inattention region is that expectations exhibit discrete adjustments and consumers, sometimes, remain inattentive to salient information. Using randomized controlled trials, Armantier et al. (2016), Armona et al. (2018), and Roth and Wohlfart (2020) investigate how US consumers' expectations are affected by the provision of information about respectively inflation, local home prices, and growth. They all report a significant share of consumers who do not revise their expectations upon the provision of information. Arguably, some consumers may have not adjusted their expectations because they had already observed the information provided during the experiments. Laboratory experiments allow to sidestep this limitation by controlling the information available to each individual. These experiments also evidence that individuals adjust their expectations in discrete jumps and sporadically ignore new information (Khaw et al., 2017; Henckel et al., 2021). Taken together, these studies favor models predicting discrete adjustments in expectations.

The literature already offers models predicting discrete adjustments in expectations. On the one hand, there are the sticky expectation models. As is illustrated in Figure 1 (Panel B), in these models consumers either have a constant probability to be attentive

Figure 1: Inattention dynamics



NOTE: *Panel A* – illustration of the dynamics of the expectation wedge (e_t) and the inattention region. The grey line represents the evolution of the expectation wedge over time. The latter is unobservable to the consumer. The only information available to her corresponds to the black lines (i.e. the threshold π) and the black dot in the upper attention region. *Panel B* – Unconditional hazard rates predicted by different models of sticky expectations. The computation of the hazard rates for the model considered in this paper is presented in Appendix B. The Calvo-like hazards predict a constant probability to update at each period (e.g. Mankiw and Reis (2002), Carroll (2003) and Carroll et al. (2020)). The time-dependent models predict that consumers update on a purely time dependent basis (e.g. Reis (2006)). All model predict an average inattention length of four periods. *Panel C* – illustrations of the predicted state-dependent hazard rates when the information at the attention choice, measured from the precision of the signals $z_t = s_t + \vartheta_t$, is perfect ($\sigma_\vartheta = 0$), imperfect ($0 < \sigma_\vartheta < \infty$), or inexistent ($\sigma_\vartheta \mapsto \infty$). The prior errors $\Delta_t \equiv s_t - E[s_t|\overline{\mathcal{I}}_{t-1}]$ are normalized by the standard deviation of innovations to permanent income.

(Mankiw and Reis, 2002; Carroll, 2003; Carroll et al., 2020)[16] or their attention behavior is purely time-dependent (Gabaix and Laibson, 2001; Reis, 2006). On the other hand, entropy-based models of rational inattention can generate discrete adjustments which neither occur with a constant probability nor are purely time-dependent (Woodford, 2009; Tutino, 2013; Khaw et al., 2017; Jung et al., 2019).

While conceptually different, these models share the prediction that the provision of information does not affect adjustment behaviors.[17] This prediction seems in contradiction with evidence from the aforementioned randomized control trials reporting that consumers are more likely to adjust their expectations after the provision of information (Armantier et al., 2016; Armona et al., 2018; Roth and Wohlfart, 2020). Alternatively, the inattention

[16]We refer to these models as Calvo attention models in analogy to the Calvo price setting model (Calvo, 1983).

[17]It follows from the Calvo or time-dependent adjustments in models of sticky expectations. In entropy-based models of rational inattention, a consumer processes the most relevant information given an entropy constraint. Hence, the provision of new information should have no effect since this information would have been already processed would it have been optimal to use it. Arguably, such interpretation of entropy-based rational inattention may be too literal. However, even assuming that individuals may freely use the information they are provided with during information experiments, then they would be bayesian and systematically adjust their expectations. Either way, these models may not explain why the provision of information leads some, but not all, consumers to adjust their expectation.

region predicts that the provision of free and novel information enters the dynamics of the expectation wedge, as any other signal would, and affects the attention behavior.

Finally, the inattention region also finds support in the psychology, neuroeconomics, and neuroscience literature aiming to explain observed patterns of choice and response times. Of particular importance in this literature is the drift-diffusion model "where the decision maker accumulates evidence until the process hits either an upper or lower stopping boundary and then stops and chooses the alternative that corresponds to that boundary." (Fudenberg et al., 2020, abstract) The inattention region is a manifestation of such drift-diffusion model in the context of continuous consumption choices.

3.2. Stochastic state-dependent attention

While the model predicts that the inattention region takes the form of a strict Ss-threshold policy, the attention behavior is stochastic as the latent information at the attention choice is unobservable. This stochasticity holds from the consumer's perspective, but also from the perspective of a modeler who observes consumption choices, expectations and permanent income.

As is apparent from Panel C in Figure 1, the discrepancy between attention and observable variables emerges because of the imperfect information at the attention choice. More specifically, when the information at the attention choice is perfect, the attention behavior is perfectly predictable from a consumer's prior error $s_t - E[s_t|\overline{\mathcal{I}}_{t-1}]$. The attention behavior is, thereby, purely state-dependent. Alternatively, when there is no information at the attention choice, the attention behavior is not correlated with prior errors. The attention behavior is, thereby, purely stochastic. As a result, in general when the information at the attention choice is imperfect, the attention behavior is stochastic state-dependent. In the presence of an inattention region, the stochastic state-dependence takes the form of smooth U-shape conditional hazard rates with a strictly positive minimum at 0.

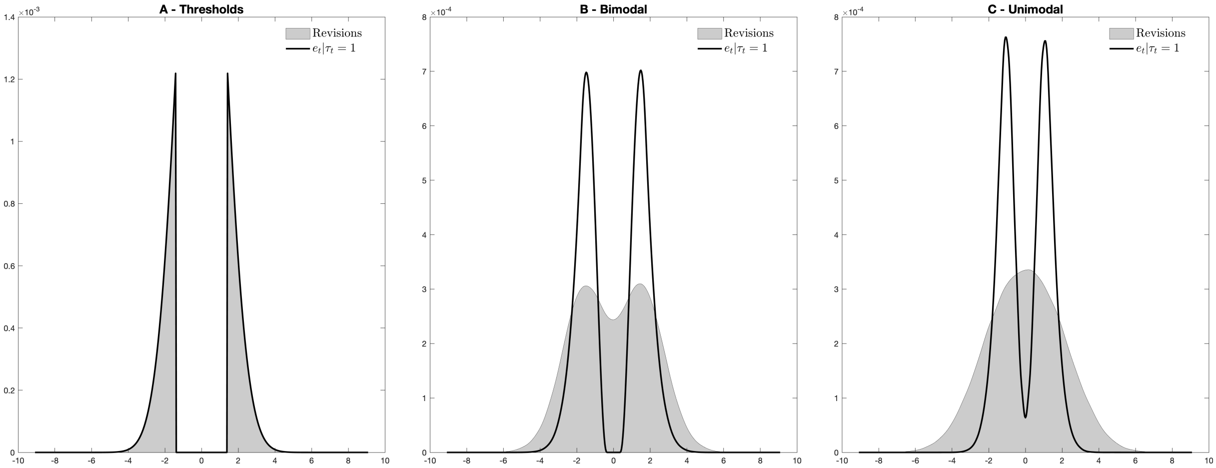
Such U-shape pattern with a positive probability at 0 is characteristic of the attention behaviors observed in Khaw et al. (2017, fig. 8) and Henckel et al. (2021, fig. 2). Khaw et al. (2017) explain at length how this U-shape pattern compares, and is hard to reconcile, with optimizing models of (price) adjustments found in the economic literature.

3.3. Expectation revisions

The aforementioned randomized controlled trials and laboratory experiments also report that individuals do not reach perfect information upon adjusting. Similarly, our model predicts that consumers catch up with the imperfect information at the attention choice when adjusting. Consequently, expectation revisions equal the expectation wedges e_t and, thereby, they almost surely never close the gap with the full information rational expectations.

Nevertheless, the prediction that expectation revisions equal the expectation wedges is, arguably, disputable. Indeed, since the expectation wedge e_t governs the attention dynamics at the inattention region, a sharp prediction of the model is that we should not observe revisions of a magnitude smaller than the threshold π and that they cannot be too far from $\pm\pi$ either. Figure 2 (Panel A) illustrates this prediction.

Figure 2: Distributions of expectation revisions



NOTE: Illustration of the distributions of revisions predicted with an inattention region. The calibrated average inattention length and consumption problem parameters are the same across panels. Values are normalized by the standard deviation of the innovation to permanent income. Panel A – distribution predicted by the model from Section 2 when the information of the consumer coincides with the information at the attention choice when attentive. Panel B – distribution predicted by an extension of the model from Section 2 where the consumer accesses more information (here perfect information) than the one at the attention choice when attentive. Panel C – Same extension than Panel B but with a higher discrepancy between the precision of the information at the attention choice and the one observed when attentive.

The equality of expectation revisions and expectation wedges ensues from the assumption that, when attentive, the consumer can observe neither more nor less than the information at the attention choice. Arguably, this is a special case and it seems as reasonable to suppose, instead, that when attentive the consumer may devote more time and effort to process more information. That is, the consumer could access information that is not in $\mathcal{I}_t$ when attentive, thus generating a discrepancy between expectation revisions and expectation wedges.

To illustrate the potential effects of such discrepancy on the distribution of expectation revisions, Online Appendix H considers an extension of the consumption problem where the consumer perfectly observes her permanent income s_t when attentive.[18] Extending

[18]Considering perfect information when attentive is convenient as it implies that a measure of the

the model in this direction has two effects on the distribution of revisions. First, the consumer also observes the expectation error at the attention choice when attentive. Therefore, the distribution of revisions is now a convolution of a Gaussian (for the errors) and the stationary distribution of expectation wedges. Second, the inattention region is not constant anymore but varies with the inattention length. This dependence generates a form of time-dependence leading to smoother stationary distributions of expectation wedges (as can be seen from the plain lines in Figure 2). Ultimately, and as is reported in Panels B and C, this extension can generate smooth bimodal and unimodal stationary distributions of revisions.[19]

In conclusion, the inattention region can be reconciled with a wide range of empirical distributions of expectation revisions, as long as the consumer's information when attentive is not restricted to coincide with the information at the attention choice. However, relaxing this assumption also implies that the inattention region varies with the consumer's attention history. While this may be more realistic to capture some features of expectation revisions, it also greatly complicates the analyze of consumption dynamics as it requires to track another dimension of heterogeneity. Because this increased complexity comes with little qualitative gains for the analyze of consumption dynamics, we continue to rely on the model from Section 2 in the rest of the paper.[20]

4. Asymmetry and size effects in household consumption

In this section, we derive new predictions on asymmetries and size effects in household consumption responses arising as a consequence of the inattention region. Using household survey data from the Bank of England, we find suggestive evidence for these predictions.

discrepancy between the information at the attention and consumption choices (when attentive) is given by the amount of information rigidities at the attention choice (i.e., the precision of the signal z_t). As we have already mentioned, evidence however shows that consumers do not reach perfect information when adjusting. Fortunately, none of the results from this extension hinge on the assumption of perfect information when attentive. Similar results continue to hold when the information remains imperfect (in the form of Gaussian signals) when the consumer is attentive. Indeed, and as is shown in Online Appendix H, the model predicts that the inattention region varies with the consumer's posterior variance at the attention choice. The assumption of perfect information implies that this posterior variance collapses to zero when the consumer is attentive, so that the inattention length is a sufficient statistic for the attention policy. The dependence of the attention policy on the inattention length, instead of on the posterior variance, is the only feature specific to perfect information.

[19] Unimodal distributions are generally reported in the empirical literature focusing on economic agent expectations. Nevertheless, bimodal distributions are sometimes a characteristics of firm price setting (Costain and Nakov, 2011) and could thus also be of interest in the economic literature.

[20] Consistent with our focus on the extensive margin of expectation adjustments, Caballero and Engel (2007) emphasize the role of the extensive margin of price adjustments to monetary shocks.

4.1. Model predictions

The consumption response depends on whether a consumer becomes attentive after an income shock. Therefore, we distinguish the conditional and unconditional consumption responses. The *conditional* propensity to consume (PC) is the ratio of the expected change in consumption to the perceived shock ω_t for consumers who adjust their consumption.[21] The *unconditional* PC consists in the same ratio computed for all consumers. In both cases, expectations are taken over the cross sectional distribution of consumers.[22]

These measures capture different aspects of consumption responses. As is formally demonstrated in Appendix D.2, the conditional response captures the fact that consumers who are close to the inattention threshold, and become attentive after the shock, have a disproportionate response to this shock because they also adjust for the expectation wedge e_t. On the other hand, the unconditional response essentially measures the contribution of the extensive margin of adjustments (Caballero and Engel, 2007). That is, it captures the additional consumption increase resulting from the rise in the fraction of consumers adjusting upwards after a positive income shock, and the fall in the fraction of consumers adjusting downwards.

Interestingly, the conditional and unconditional responses are affected differently by the size of shocks. At the stationary cross section of consumers, the conditional PC decreases with the size of a shock. This is because a larger shock prompts consumers who are further away from the inattention threshold to adjust their consumption. As a result, the expected over-reaction to the shock decreases, and so does the conditional response. On the other hand, the unconditional PC increases with the size of a shock. This is because a larger positive (resp. negative) shock increases (resp. decreases) the proportion of consumers who revise their consumption upwards and decreases (resp. increases) the proportion of consumers who revise downwards.

The symmetry of the conditional and unconditional consumption responses follows from the symmetry of the stationary cross sectional distribution of consumers. However, in the presence of aggregate income shocks, the cross sectional distribution of consumers is generally asymmetric. Hence, Appendix D.2 analyzes how the conditional and uncon-

[21]We consider the response to a perceived shock ω_t because it is the concept relevant for the empirical analysis below.

[22]We formally introduce the cross sectional distribution of consumers in the next section. It corresponds to the distribution of expectation wedges at period t denoted $a_t(e)$. The stationary distribution is symmetric and we have $E_a[e] = 0$. We say that the distribution is shifted to the left (resp. right) when $E_{a_t}[e] < 0$ (resp. $E_{a_t}[e] > 0$). Left (resp. right) shifts arise as a consequence of recent negative (resp. positive) aggregate income shocks.

ditional PC are affected by the size and sign of a perceived income shock when the cross section of consumers deviates from its stationary distribution. It shows that the size effects are generally asymmetric and not monotonic. Nevertheless, using simulated data, Appendix D.2 derives a set of testable predictions for linear regressions on shock sizes that we summarize in the following. For clarity, we only report the predictions for left shifts (due to negative aggregate shocks) in the distribution of consumers. The predictions for right shifts follow by symmetry.

When the cross sectional distribution of consumers is to the left:

P1 The probability to adjust consumption increases with the size of negative perceived shocks, while the sign of the slope for positive shocks is unclear.

P2 The unconditional PC is increasing for negative shocks, while the sign of the slope for positive shocks is unclear.

P3 The conditional PC is decreasing for negative shocks. The sign of the slope for positive shocks is unclear, though it appears to be negative when shifts are not too drastic.

P4 The intercept of the conditional propensity to consume is larger for negative shocks.

4.2. Evidence from household surveys

Data – We confront predictions P1-P4 to data. To this end, we rely on the 2012-2014 surveys of consumers from the Bank of England (BoE). For the purpose of the present paper, a key advantage of the BoE survey is that respondents are asked to report their perceived income surprise over the last 12 months and by how much they adjusted their spending in response to this surprise. As a result, because some households experienced positive surprises while other experienced negative ones, it allows us to analyze size effects for both positive and negative shocks. However, we observe only one shock and consumption response per respondent. Appendix E.1 provides further information about the survey characteristics and the main variables that we use. Of particular interest, these variables include measures for households' liquid assets net of unsecured debt, cash-on-hand and whether the income surprise was likely to persist.

The average unconditional MPC is 0.41, a value within the 0.2-0.6 range for standard estimates of annual propensities to consume (Carroll et al., 2017). About 50% of the individuals who reported an income surprise did not adjust their consumption, while the average conditional MPC of those who adjusted was 0.82. These averages hide significant variations across time. In 2012, when GDP growth was negative for two quarters in the UK, the unconditional MPC was about 10 percentage points larger than in 2013-2014. As can been seen from Figure E.6, this increase is attributable to a similar increase in the share of consumers who adjusted their consumption. Consistent with a large strand of the literature, the unconditional MPC negatively correlates with liquid asset holdings and cash-on-hand, and is smaller for consumers who are the less likely to be credit constrained. Decomposing the evolution of the unconditional MPC along these dimensions (Figure E.6) again reveals that these patterns of the unconditional MPC are essentially driven by the extensive margin of consumption responses, while variations at the intensive margin are much less pronounced.[23]

Asymmetry and size effects – Figure E.7 plots MPCs across quartile of income surprises. There is a significant asymmetry in responses to positive and negative income surprises. The unconditional MPC increases from 0.14 for positive surprises to 0.63 for negative surprises. This increase is essentially due to a small share of consumers who adjust following a positive surprise (25%) as opposed to the large share of adjustments following a negative surprise (77%). This asymmetry is present in each wave of the survey, persists when controlling for observable household characteristics and is not specific to respondents who are likely liquidity or credit constrained (not reported). It could indicate that the cross section of consumers was shifted to the left, which seems coherent with the economic downturn of the UK economy between 2012 to 2014.

Accordingly, prediction P1 implies that the probability to adjust consumption should increase with the size of negative surprises. Panel A in Table 1 tests this prediction in data. The dependent variable is a binary indicator for wether an individual responded to the income surprise. The independent variables include a set of household and time controls, and linear slopes for the size of negative and positive surprises (in £1,000). Estimating this regression, we find a positive and statistically significant size effect for negative surprises. The size effect for positive surprises is statistically insignificant. These findings are consistent with P1.

Panel B estimates size effects for the unconditional MPC ('All'). The size effect is positive for negative income surprises, and negative for positive surprises. These findings are consistent with prediction P2. Looking at the conditional response ('$\neq 0$'), we in-

[23]The linear consumption function in Lemma 1 can be thought as a first order local approximation of a concave consumption function, with L_t decreasing with liquid assets holding or cash-on-hand. Hence, up to this approximation, the model predicts a wider attention region for individuals with higher financial wealth.

Table 1: Asymmetry and size effects in the BoE survey of consumers

Panel A - Probability to adjust consumption (Logit)

	Full sample				Liquid assets (top 50%)	
Shock size						
Positive	-0.060***	(0.015)	-0.000	(0.018)	-0.037	(0.037)
Negative	0.211***	(0.021)	0.160***	(0.034)	0.173***	(0.056)
Constant	-0.248***	(0.061)	-1.232***	(0.270)	-0.737	(0.479)
Controls	No		Yes		Yes	
Obs. – *pseudo* R^2	2,101	*0.079*	1,706	*0.169*	584	*0.184*

Panel B - Linear shape of the MPC (OLS)

	Full sample				Liquid assets (top 50%)	
MPC	All	$\neq 0$	All	$\neq 0$	All	$\neq 0$
Shock size						
Positive	-0.024***	-0.016***	-0.011***	-0.016***	-0.012***	-0.010
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.008)
Negative	0.022***	-0.013***	0.014***	-0.012***	0.020***	-0.008**
	(0.003)	(0.002)	(0.003)	(0.003)	(0.005)	(0.004)
Intercepts						
Negative		0.152***		0.114***		0.122*
		(0.030)		(0.036)		(0.069)
Constant	0.409***	0.758***	0.191***	0.752***	0.241***	0.710***
	(0.013)	(0.027)	(0.049)	(0.055)	(0.087)	(0.106)
Controls	No	No	Yes	Yes	Yes	Yes
Obs.	2,101	1,049	1,706	844	584	256
R^2	0.085	0.118	0.213	0.154	0.215	0.141

Panel A: Logistic regressions. The dependent variable is a dummy variable for whether an individual MPC is not zero. Panel B: OLS regressions. The dependent variable is the propensity to consume out of an income shock. 'MPC $\neq$ 0' refers to the subsample of non-zero MPC (conditional MPC), while 'All' is the unconditional MPC. Both panels report robust standard errors in parenthesis. *, ** and *** denote significance at the 10%, 5% and 1% levels respectively. Regressions are run separately on all observations (Full sample) and the subsample of individuals with net liquid asset holdings above the median. Control variables are categories for age, employment status, debt concerns, fear about future income drops, whether the household is credit constrained, the type of income shock (temporary or likely to persist), gross and discretionary income quartiles, and survey wave fixed effects. The base is an individual, aged 35-45, who responded in 2012, who is working, has no concern about her debt, does not fear an income drop, has experienced an unexpected income increase, is not credit constrained, has experienced a temporary income shock, has an annual gross income between £25,000 and £49,999 and a monthly discretionary income between £600 and £1,199.

stead find that size effects are negative and statistically significant for both positive and negative surprises, as predicted by P3. Prediction P4 further states that the conditional response should be larger at 0^-. The intercept for negative shocks in Panel B estimates the difference in the conditional response at 0^- and 0^+. The estimate is positive and statistically significant for the conditional response, consistent with prediction P4.[24]

Robustness – The baseline results are consistent with predictions P1-P4. To check the sensitivity of these results, we perform various robustness checks. The persistence of the income surprise experienced by households could likely correlate with its size. In the survey, respondents are asked to report whether they believe the income surprise to be temporary or likely to persist. Table E.4 in the Appendix reports the estimation results from the subsample of temporary income surprises. The results remain consistent with predictions P1-P4, though not statistically significant anymore for the conditional response for which we have fewer observations.

Credit and liquidity constraints, and precautionary savings could cause asymmetries and size effects. Therefore, we estimate the baseline regressions when excluding net liquid asset holdings below the median (Table 1), individuals who fear being credit constraint, or cash-on-hand below the median (Table E.4). These exclusions have little impact on predictions P1-P4, indicating that the size effects and asymmetry found in our data are not attributable to credit and liquidity constraints, nor precautionary savings.

While predictions P1-P4 refer to linear slopes, size effects are generally nonlinear. To asses the shape of these nonlinear effects in the data, we introduce cubic polynomials with respect to size in the baseline regressions. Figure E.8 reports the estimated nonlinear size effects. They echo those reported in Figure D.5 which are obtained from simulating consumption responses in the presence of an inattention region.

Alternative theories – Our analyze reveals two features of consumption responses that cannot be explained simultaneously by alternative consumption theories. First, consumption responses are driven by a state-dependent extensive margin and an over-reaction at the intensive margin. Second, the size effect associated to positive income shocks is decreasing in our data. The extensive margin of nondurable consumption responses is absent from most consumption theories in the literature. Among theories predicting an extensive margin of consumption adjustments, time-dependent and Calvo-like expectation adjustments (Reis, 2006; Carroll et al., 2020) cannot explain the dependence of the probability to adjust consumption on shock size. Salience (Kueng, 2018) can, to some

[24]As predicted by theory, we restrict the unconditional MPC to be continuous at zero.

extend, capture the state-dependence of expectation adjustments, but it is inconsistent with the asymmetric size effects in the probability to adjust consumption and the observed over-reaction at the intensive margin of consumption responses. Models with consumption adjustment costs (Fuster et al., 2021) are consistent with the state-dependence at the extensive margin and over-reaction at the intensive margin. They are, however, inconsistent with the non increasing size effect associated to positive income shocks for the unconditional MPC and probability to adjust.[25]

Reconciling evidence on size effects from positive shocks – Size effects from positive income shocks are ambiguous in the literature. Using survey questions about spending in hypothetical scenarios, Christelis et al. (2019) report quasi-experimental evidence on size effect from positive shocks. Consistent with our results, they find a *decreasing* size effect for the unconditional PC. Using a similar methodology than Christelis et al. (2019), Fuster et al. (2021) reach the opposite conclusion that they are *increasing*. A distinctive feature of our model is that size effects for the unconditional response are likely to reverse with aggregate fluctuations. In particular, predictions P1-P4 are also consistent with the results reported in Fuster et al. (2021).[26]

5. Aggregate consumption dynamics

This section shows that the asymmetry and size effects evidenced at the household level also alter the dynamics of the aggregate consumption response.

[25]The size effect associated to positive shocks is positive in these models. Fuster et al. (2021) show this result for the stationary distribution of consumers. Deviations from the stationary distribution are small, even in the presence of aggregate shocks, and unlikely to reverse the size effect associated to positive income shocks in these models where adjustments are essentially driven by idiosyncratic shocks; Unless the ratio of the variances of aggregate to idiosyncratic income shocks is large (Chetty and Szeidl, 2016). It is, however, well-established that this is not the case and that most of the income uncertainty faced by individuals comes from idiosyncratic shocks. As we shall see more precisely in the next section, significant deviations from the stationary distribution of expectation wedges can arise in our model when consumers use different channels to gather information about idiosyncratic and aggregate shocks.

[26]In Table 3 (p. 1770), they report that 74% of consumers do not respond to a $500 income gain, while this share was lower (42%) for a loss of the same size – an asymmetry that they cannot fully explain with a borrowing constraint. It suggests that the distribution of consumers was to the left. Looking at the unconditional response, they find a larger PC from a $500 loss than an equivalent gain. Looking at the conditional response, they again report a higher PC for looses at $500. These higher conditional and unconditional PC from income losses are also consistent with the distribution being to the left. They find a positive size effect from positive shock in the probability to adjust and the unconditional response (P1 and P2) and a negative size effect in the conditional response (P3). Their experiment does not allow to estimate size effects from negative shocks.

5.1. Calibration

Idiosyncratic and aggregate shocks – There is a unit mass continuum of infinitely-lived consumers, indexed by i, who are ex ante identical and whose problem is given by (3). For simplicity, the discount rate is equal to $\beta = (1 + r)^{-1}$. Innovations to permanent income can be written as the sum of two i.i.d. Gaussian innovations, one for aggregate factors and the other for idiosyncratic ones. Arguably, consumers use separate information channels to be informed about aggregate and idiosyncratic innovations. We therefore allow consumers to be separately attentive to each type of shocks in return of the same attention cost λ. The quadratic structure of problem (3) and the independence of innovations imply that the two attention choices are separable. Similar approaches are taken, for instance, in Maćkowiak and Wiederholt (2009, 2015) and Carroll et al. (2020). Consequently, idiosyncratic income surprises cancel out for aggregate dynamics. Let $\bar{\zeta}_t$ denote an aggregate innovation to permanent income common to all consumers. The upper bar notation emphasizes that we refer to an aggregate shock henceforth.

Income process and attention – In order to provide quantitative illustrations, we follow Wang (2003) and Luo et al. (2017) and assume that the income process follows an AR(1) with persistence $0 < \rho < 1$ and standard deviation of the innovation σ_ϵ. The income calibration is taken from Pischke (1995, Table VII). The time period is a quarter, the observational unit a household and values are expressed in 1982 dollars. Average income is $\bar{y} = \$6,926$ and $\rho = 1 - 0.438$, $\sigma_\epsilon = \$2,470$ and $r = 0.015$. We follow the literature and take as a rule-of-thumb that the variance of idiosyncratic innovations is 100 times the variance of the aggregate innovations (Carroll et al., 2020).

Estimates at the macro-level indicate that consumers update their expectation about macroeconomic variables about once a year on average (Carroll, 2003; Mankiw et al., 2003; Reis, 2006). The attention cost λ is set accordingly. The remaining parameter σ_ϑ stands for the signal informativeness. This parameter determines the Kalman filter gain from Lemma 2 and, thus, the rate at which income shocks are incorporated. We set the Kalman gain to 0.45, thus implying that 32% of the consumption response to a marginal income shock arises on impact at the ergodic distribution. This value is slightly below the unconditional response estimated in Reis (2006) for the US.[27]

Absence of salient consumption jumps – A recurrent objection to non-convex consumption models is that household nondurable consumption does not appear to remain

[27] Online appendix I provides a sensitivity analysis for the model parameters. It also reports the welfare cost of information frictions. Consistent with recurrent findings in the literature, this cost is small.

constant for a long period of time and that we do not observe large sporadic jumps (Reis, 2006; Carroll et al., 2020). We briefly mention why these critics hardly apply to the type of adjustments predicted by the inattention region. First, the type of adjustments predicted by the inattention region do not refer to changes in household consumption levels, but to consumption plans. Consequently, the prediction that consumption should remain constant while the consumer remains inattentive is not a consequence of the inattention region *per se*, but a consequence of the simplifying assumption that there is no trend in consumption (i.e. $\beta^{-1} = 1 + r$).

Second, the jumps in consumption predicted by the model are small and would be hardly observable in consumption data. With our calibration, the predicted jump in consumption at the attention threshold is ± \$11.6 (0.17% of quarterly consumption).[28]

Last, our calibration implies that only 14% of households would wait more than a quarter to incorporate new information about their idiosyncratic shocks. This prediction of close to full attention to idiosyncratic shocks relates to the order of magnitude of the ratio of idiosyncratic to aggregate variances, which is a robust finding in the literature. For instance, this probability remains stable (19%) when halving this ratio. Consequently, frequent unpredictable adjustments are also expected to be a salient feature of household quarterly data in the presence of an inattention region.

5.2. Cross sectional distribution of consumers

While ex ante identical, households differ ex post in terms of their expectation wedges $e_{i,t}$. Therefore, analyzing the aggregate dynamics of consumption requires to track the cross sectional distribution of $e_{i,t}$. To this end, let $a_t(e)$ denote the cross sectional distribution at period t before consumers become attentive. The share of attentive consumers in the economy at period t is therefore $\Pi_t \equiv \int_{e \notin \Xi} a_t(e)de$. Appendix C.1 characterizes the ergodic distribution of consumers and demonstrates that it is symmetric.

Recall that consumers rely on a Kalman filter to slowly incorporate information about the aggregate shock at the attention choice. Consequently, an aggregate shock does not only disturb the cross sectional distribution when it occurs, but does so persistently at a rate that depends on the Kalman gain. More specifically, the translation in the cross sectional distribution at period t is given by a weighted sum of past aggregate income

[28] See Online Appendix I.2 for the computation of these values. As a comparison, Caballero (1995) estimates that in order to capture the stickiness of US aggregate consumption data – that our model matches as we shall see in the next section – the implied jump in a model with consumption adjustment costs are about 6 percent, more than 30 times the jumps that we have here.

shocks

$$S_t = \sum_{s=0}^{t-1} \upsilon_s \bar{\zeta}_{t-s} \tag{14}$$

where $\upsilon_s \equiv K(1+r)^s(1-K)^s$ is the share of the aggregate shock – augmented by its returns – that is internalized on average at period $t+s$ at the attention choice. Accounting for this new state variable, Appendix C characterizes the dynamics of the cross sectional distribution of expectation wedges that we report in the following Proposition

Proposition 2. *The dynamics of the cross sectional distribution in the presence of aggregate shocks is characterized by the following system of dynamic equations*

$$a_{t+1}(e) \propto \underbrace{\int_{\tilde{e}\in\Xi} \phi\left(\frac{e - S_{t+1} - (1+r)\tilde{e}}{\sqrt{\sigma_\omega^2 - \sigma_S^2}}\right) a_t(\tilde{e})d\tilde{e}}_{\textit{Inattentive at } t} + \underbrace{\underbrace{\phi\left(\frac{e - S_{t+1}}{\sqrt{\sigma_\omega^2 - \sigma_S^2}}\right)}_{\textit{resetting at } 0} \underbrace{\int_{\tilde{e}\notin\Xi} a_t(\tilde{e})d\tilde{e}}_{\textit{share attentive}}}_{\textit{Attentive at } t} \tag{15}$$

$$S_{t+1} = (1-K)(1+r)S_t + K\bar{\zeta}_{t+1} \tag{16}$$

along with initial conditions $a_0(e)$ and S_0. $\phi(.)$ is the pdf of the standard normal distribution.

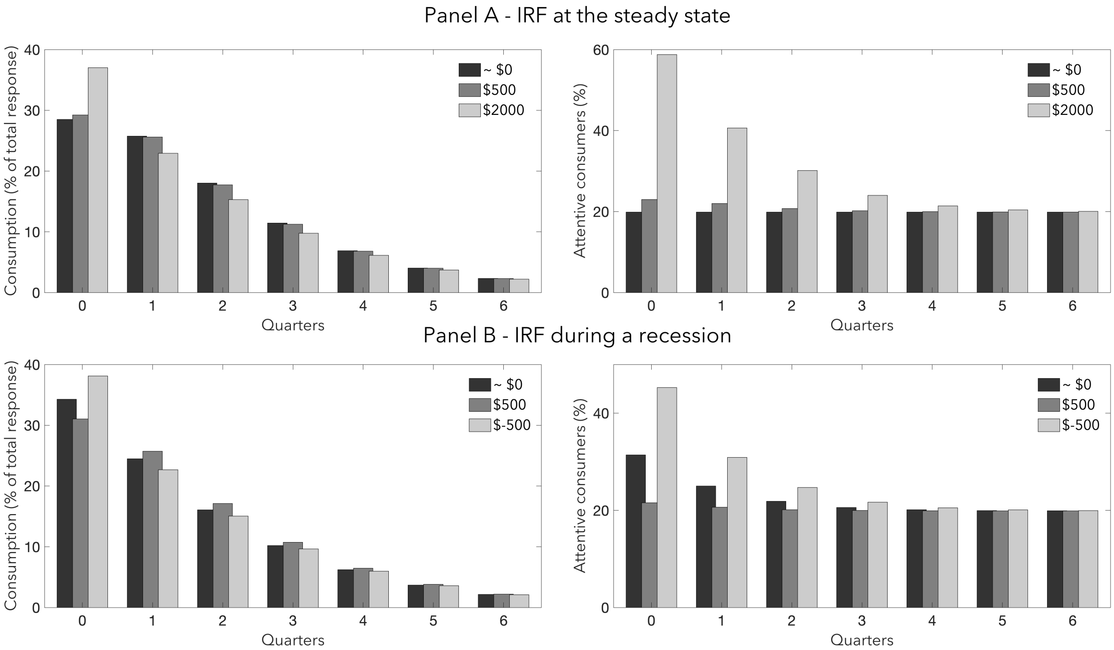
The first term in (15) captures the dependence on past expectation wedges when consumers remain inattentive, while the second term accounts for the fact that attentive consumers adjust for this wedge. Moreover, Proposition 2 indicates that S_t follows an AR(1) process. In general, the dynamics of S_t is stationary since r should be small in comparison to K. Therefore, and importantly, this state variable is related to business cycles in this economy: S_t is positive when recent aggregate shocks were positive, and vice versa.

5.3. Aggregate response dynamics

Appendix C demonstrates how the dynamics of $a_{t+1}(e)$ can be used to derive aggregate consumption growth and impulse response functions. These impulse responses depend on the full history of past aggregate shocks and the sign and size of the current aggregate shock. To provide insights on this dependencce, we consider two initial scenarios: steady state and recession.

In the first scenario, the economy is initially at its steady state. It corresponds to a situation where aggregate shocks were nil for a long period of time. Panel A in Figure 3 reports the impulse response following aggregate shocks to permanent income of increasing

Figure 3: Size and asymmetry in aggregate consumption dynamics



NOTE: Impulse response functions of aggregate consumption (left panels) and shares of attentive consumers (right panels) for aggregate shocks to permanent income of different size at quarter 0. Values are expressed in 1982 dollars. The steady state scenario (panel A) corresponds to a situation where aggregate shocks were initially nil for an infinite number of periods. The recession scenario corresponds to a situation where the initial cross sectional distribution is the steady state distribution centered around S_{rec} such that $P(S_0 < S_{rec}) = 0.025$, that is a one in a ten year recessionary state.

size. The ~ 0 shock illustrates the marginal response. As can be seen, the share of attentive consumers at the steady state is low (20%) and the steady state consumption response is sluggish with only 28.5 percent of the total response arising on impact.

As the size of the shock increases, it prompts more consumers to become attentive, thus resulting in a sharper consumption response. These changes in the consumption response dynamics are particularly salient for the $\$2,000$ shock. As can be seen from Figure 3, this large shock prompts about 60% of consumers to be attentive on impact, leading to an impact response of 37 percent. Interestingly, the share of attentive agents remains significantly higher than its steady state value for about a year and a half. This is due to the slow filtration of the imperfect information at the attention choice, driving the dynamics of S_t. In comparison, recall that a version of the model without imperfect information at the attention choice would induce a perfect state-dependence of the attention behavior. As a result, the cross sectional distribution from Proposition 2 would be a Dirac distribution and an aggregate shock would either prompt all or no consumers to become attentive when it occurs.

To put some perspective on the size of the shocks, the $\$2,000$ shock is similar to the average annual payments from the state of Alaska's Permanent Fund analyzed in Hsieh (2003) and Kueng (2018). The \$500 shock is about the same (resp. twice) size than the 2008 (resp. 2001) Tax Rebates (Parker et al., 2006, 2013). As can be seen from Figure 3, these rebates would not have had much effect on the share of attentive consumers and, thereby, would not have perturbed much consumption dynamics at the steady state. However, our simulations suggest that these tax rebates may have had an adverse effect on consumption dynamics because they were implemented during recessions.

Indeed, panel B in Figure 3 reports impulse response functions during a one-in-a-ten-years recession. The counterfactual with a marginal shock illustrates the dynamics of consumption during this recession. As is apparent from the marginal response, consumers are more attentive on average at the trough and, thereby, more likely to adjust their consumption to account for the state of the economy. However, looking at the $\pm$\$500 shocks, we observe that consumption dynamics are highly asymmetric during recessionary episodes. Indeed, while a further destabilizing shock prompts more consumers to become attentive and to adjust their consumption, a positive income shock has the opposite effect and leads to a more sluggish response. The opposite is true during economic booms. This asymmetry suggests that the state-dependence of consumers inattention may partially mute the short run consumption response to economic policies leaning against the wind.

Interestingly, these asymmetric dynamics have been identified in data. For instance, analyzing US aggregate consumption dynamics, Caballero (1995) reports that "in good times, consumers respond more promptly to positive than to negative wealth shocks, while the opposite is true in bad times." (Caballero, 1995, p. 30) Similarly, Ocal and Osborn (2000) report that consumption response dynamics in the UK also depends on the the state of the economy and the sign and magnitude of a shock.

6. Consumption Persistence

Consumption dynamics is highly persistent in aggregate data, but not in household data. The meta-analysis from Havranek et al. (2017) considers 424 estimates from studies on aggregate data and identifies a median estimate of 0.66. In comparison, they report a median of 0 from 183 estimates on household consumption. In line with Calvo-like model of inattentive consumers (Carroll et al., 2020), the inattention region also implies that consumption is not persistent at the household level (Proposition 3 in the Appendix) but highly persistent at the aggregate level. However, a distinctive feature of the inattention region is that the share of inattentive consumers varies over time. In this section, we demonstrate that it generates time variations in aggregate consumption persistence consistent with US consumption data.

6.1. Inattention and aggregate consumption persistence

To assess the relation between aggregate consumption persistence and inattention, we simulate aggregate consumption dynamics.[29] We then estimate the following equation

$$\Delta C_{t+1} = \alpha_{t+1} + \gamma_{t+1} \Delta C_t + \beta_{t+1} \bar{\zeta}_{t+1} + error_{t+1} \tag{17}$$

that relates the aggregate consumption change to its lag ΔC_t and aggregate income shocks $\bar{\zeta}_{t+1}$. The coefficient γ_{t+1} measures the (potentially time-varying) persistence in aggregate consumption.

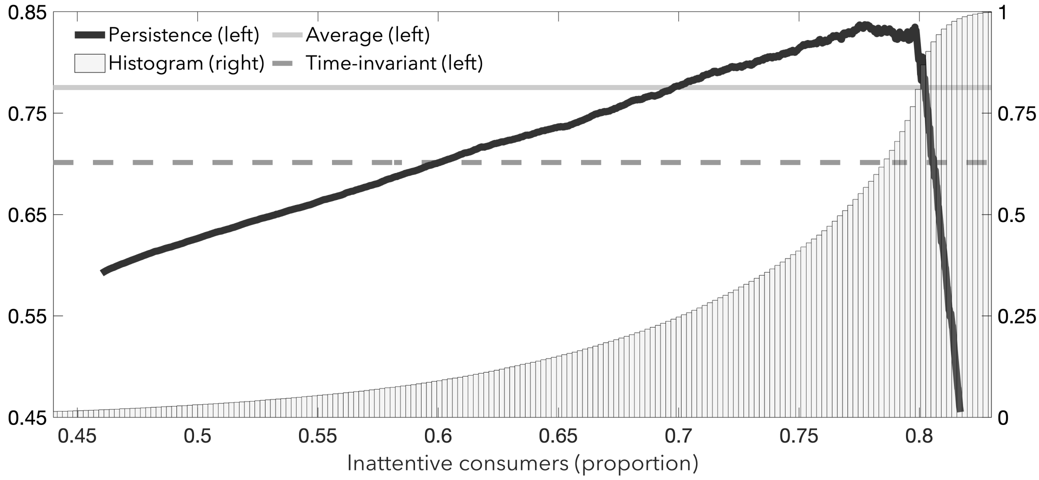
We first estimate a time-invariant regression of Equation (17) as it corresponds to the standard specification used in the literature to estimate the persistence of consumption. The estimated time-invariant persistence is equal to 0.70, consistent with the median estimate reported in Havranek et al. (2017). In models with a constant share of inattentive consumers Π and quadratic utility, we have $\gamma \simeq \Pi$ (Carroll et al., 2020). In the present paper, the share of inattentive consumers Π_t is endogenous and varies over time. Consequently, one may expect that consumption persistence continues to relate to the share of inattentive consumers and, thereby, also varies over time. To assess this hypothesis, we order the simulated data depending on Π_t and estimate a mapping $\gamma_{t+1} = \gamma(\Pi_t)$ using rolling regression with a window of 4,000 observations.

Figure 4 reports the estimated relation between γ_{t+1} and the share of inattentive consumers Π_t. The average persistence (0.77) is larger than the time-invariant estimate. Moreover, consumption persistence increases linearly with the share of inattentive consumers (on most of the domain). The amplitude of this variation is large and holds important implications for consumption dynamics. For instance, the half-life is 1.5 quarters when 50% of consumers are inattentive, while it increases to 3.2 quarters when they are 75% to be inattentive.

Figure 4 displays a sharp decrease in the estimated γ_{t+1} when the share of inattentive consumers is larger than 80%. As is further discussed in Online Appendix J, this

[29] We start the simulation from the steady state cross section of consumers, simulate the model for 4,000 periods and drop the first 500 periods. We repeat these simulations 24 times. The results are thus based on 84,000 simulated data points.

Figure 4: Aggregate consumption persistence and inattention



NOTE: Rolling regression of Equation (17) based on 84,000 simulated data points and a window of 4,000 observations. The black line reports the estimated persistence $\gamma_{t+1} = \gamma(\Pi_t)$. The grey area is a cumulative histogram of the shares of inattentive consumers. The grey plain line reports the sample average of estimated persistence and the grey tilted line the time-invariant estimate obtained from an OLS estimation on the 84,000 observations.

decrease in the estimates is accompanied by a sharp decrease in the model R-squared. This is because an autoregressive process of order 1 is unable to capture the hump-shape consumption dynamics for these specific, and relatively rare, periods. Allowing for a further lag in consumption growth provides a better fit for these periods and confirms that consumption persistence increases monotonically with inattention.

6.2. Persistence in US consumption

To assess whether these variations in consumption persistence are a characteristic of US consumption data, we adopt a standard approach to estimate (nondurables and services) consumption growth (Sommer, 2007; Kiley, 2010; Carroll et al., 2011, 2020). More specifically, we reformulate consumption growth to account for its persistence, as well as potential rule-of-thumb consumers and capital market imperfections. As is discussed in more details in, for instance, Carroll et al. (2020), these considerations lead to the following benchmark specification

$$\Delta \log C_{t+1} = a + \gamma \Delta \log C_t + \eta E_t[\Delta \log Y_{t+1}] + \alpha A_t + \varepsilon_{t+1} \tag{18}$$

where $E_t[\Delta \log Y_{t+1}]$ is the predictable component of income growth and A_t is a normalized measure of asset holding.[30] We follow the literature and interpret η as the share of rule-of-thumb consumers (Campbell and Mankiw, 1989) and γ as consumption growth persistence due to consumer inattention.[31]

It is well-known that because of time aggregation, measurement error and transitory shocks, OLS estimates of Equation (18) are likely to be biased with quarterly data (Sommer, 2007). Therefore, we adopt an instrumental variable approach based on suitable instruments dated at $t-1$ and before. Unfortunately, lagged instruments are also more likely to be weak predictors of the endogenous variables. To limit the risks induced by weak instruments, we adopt two complementary approaches.

First, we report the results for two sets of instruments. The first set of instruments is inspired from Carroll et al. (2011). It consists in two to four lags in $\Delta \log C_{t+1}$, $\Delta \log Y_{t+1}$, A_{t+1}, the unemployment rate, the 3-month Treasury Bill rate, the volatility of personal consumption expenditures (PCE) price deflator[32] and consumer sentiment from the Michigan survey. The second set of instruments is inspired from Kiley (2010) who focuses on the weak instrument problem. It consists in two to four lags in $\Delta \log C_{t+1}$, $\Delta \log Y_{t+1}$, A_{t+1}, the real interest rate from 3-month Treasury Bill rate, working hours annual growth rate and annualized inflation from PCE. We also include the two periods ahead GDP growth Greenbook forecasts made at $t-1$ which has been shown to be a good instrument for predictable income growth (Bhatt et al., 2020). Second, we report weak instrument robust confidence intervals for consumption persistence. These confidence intervals use the two-step method in Andrews (2018) and a projection method. They are implemented using the Stata command presented in Sun (2018) and are robust to HAC errors.

As can be seen from the top panel in Table 2, the results from the benchmark estimation are very similar to previous findings across the literature. Consumption persistence is statistically significant and large, about 0.76. When controlling for $E_t[\Delta \log Y_{t+1}]$ and A_t, the persistence drops to about 0.51, but remains significant. Importantly, the point estimates are barely affected depending on the set of instruments that we consider. The

[30]We follow Carroll et al. (2011) and introduce asset holdings to control for their effects due to either uncertainty and time-variation in interest rates.

[31]γ has generally been interpreted as the consequence of consumption habits in the literature. However, to the extend that this persistence is absent or, at the very least, small in household consumption data, habits are unlikely to generate the high persistence reported in aggregate data. Hence, we follow Carroll et al. (2020) and interpret the serial correlation in consumption growth as a consequence of consumer inattention. Moreover, given specification (18), the serial correlation in consumption growth γ is a sufficient statistic for consumption growth persistence.

[32]See Carroll et al. (2011) for the computation of this volatility.

low Kleibergen-Paap F-statistics, however, raise reasonable concerns about the possibility of weak instruments bias in the 'horse race' regression that pits predictable income growth against lagged consumption growth. Hence, we report the estimation results from the full specification solely for robustness.

In the benchmark specification, consumption persistence is assumed to be constant over-time. However, as we have seen, the persistence is not constant anymore in the presence of an inattention region. More specifically, we have seen that the model predicts a decreasing mapping between aggregate consumption persistence and the share of inattentive consumers. Consequently, one may want to test whether consumption persistence is indeed lower when the share of inattentive consumers is low.

However, we do not observe these shares directly in the data. Nevertheless, the model simulations indicate that the relation between $|\Delta C_t|$ and Π_t is well-approximated by a monotonically decreasing one-to-one mapping. Hence, we proxy periods with a relatively high share of attentive consumers from the top 10% of the distribution of $|\Delta \log C_t - \overline{\Delta \log C_t}|$ where $\overline{\Delta \log C_t}$ is the average consumption growth over the sample size.[33] The middle panel in Table 2 reports the estimates when adding an interaction between this proxy and consumption persistence in Equation (18). As predicted by the model, we observe that consumption persistence drops significantly during periods when consumers were likely more attentive. This conclusion is confirmed by the weak instrument robust confidence interval for the first set of instruments, and also emerges after controlling for $E_t[\Delta \log Y_{t+1}]$ and A_t.

A drawback of this proxy is that it may be affected by measurement errors in quarterly aggregate consumption. Hence, we consider recessionary periods as an alternative proxy for periods when consumers are more likely to be attentive.[34] The bottom panel in Table 2 reports the estimates obtained with this alternative proxy. It shows that consumption persistence drops significantly during these periods.[35] This conclusion is confirmed by the weak instrument robust confidence interval for both sets of instruments, and also emerges when controlling for $E_t[\Delta \log Y_{t+1}]$ and A_t though the difference is statistically significant

[33]We use the deviation of consumption growth from its average because this is the prediction that we obtain from the model when $\beta \neq (1+r)^{-1}$. The 10% threshold is arguably arbitrary. Fortunately, the conclusions are not sensible to this value. For instance, we reach similar conclusions when setting the threshold to 50%, though the difference between the two samples becomes, unsurprisingly, statistically insignificant.

[34]The relation between Π_t and recessionary periods has already been discussed in the previous section. Moreover, Dräger and Lamla (2012) end Coibion and Gorodnichenko (2015) have reported that information rigidities drop persistently in the aftermath of a recession using surveys of professional forecasters.

[35]Kumar and Jia (2019) report similar results using rolling regression.

Table 2: Persistence in US consumption

	OLS		IV 1			IV 2		
	coeff.	(s.e.)	coeff. [weak instr.	(s.e.) 95% CI]	1-st. Shea partial R^2	coeff. [weak instr.	(s.e.) 95% CI]	1-st Shea partial R^2
Benchmark $-\ \mathbf{\Delta}\log\mathbf{C_{t+1}} = a + \gamma\mathbf{\Delta}\log\mathbf{C_t} + \eta\mathbf{\Delta}\log\mathbf{Y_{t+1}} + \alpha\mathbf{A_t} + \varepsilon_{\mathbf{t+1}}$								
	$\Delta\log C_t$ only							
γ	0.426***	(0.064)	0.756***	(0.072)	0.372	0.771***	(0.063)	0.439
			[0.656, 0.957]			[0.702, 0.993]		
KP F-stat, $\overline{R}^2$	–	0.182	*9.89*		0.071	*11.11*		0.155
	$\Delta\log C_t$ and controls							
γ	0.306***	(0.062)	0.524***	(0.104)	0.273	0.505***	(0.113)	0.322
η	0.205***	(0.041)	0.279***	(0.095)	0.114	0.318***	(0.100)	0.120
α	-7.55e-04**	(3.22e-04)	-5.06e-04**	(2.40e-04)	0.949	-4.99e-04*	(2.76e-04)	0.955
KP F-stat, $\overline{R}^2$	–	0.325	*1.55*		0.237	*1.21*		0.207
Top 10% $-\ \mathbf{\Delta}\log\mathbf{C_{t+1}} = a + (\gamma + \delta\mathbb{1}_{\mathbf{10\%}})\mathbf{\Delta}\log\mathbf{C_t} + \eta\mathbf{\Delta}\log\mathbf{Y_{t+1}} + \alpha\mathbf{A_t} + \varepsilon_{\mathbf{t+1}}$								
	$\Delta\log C_t$ only							
γ	0.495***	(0.092)	0.800***	(0.084)	0.390	0.783***	(0.055)	0.434
			[0.693, 1.247]			[0.773, 1.398]		
δ	-0.135	(0.108)	-0.312***	(0.085)	0.495	-0.361***	(0.049)	0.632
			[-0.839,-0.019]			[-1.025, 0.232]		
KP F-stat, $\overline{R}^2$	–	0.185	*17.70*		0.127	*22.78*		0.208
	$\Delta\log C_t$ and controls							
γ	0.359***	(0.085)	0.556***	(0.099)	0.342	0.598***	(0.062)	0.403
δ	-0.104	(0.091)	-0.262***	(0.097)	0.492	-0.305***	(0.061)	0.626
η	0.203***	(0.041)	0.268***	(0.046)	0.174	0.233***	(0.032)	0.224
α	-7.49e-04**	(3.13e-04)	-5.20e-04**	(2.09e-04)	0.956	-4.88e-04**	(2.37e-04)	0.959
KP F-stat, $\overline{R}^2$	–	0.326	*5.18*		0.278	*31.36*		0.304
Recession $-\ \mathbf{\Delta}\log\mathbf{C_{t+1}} = a + (\gamma + \delta\mathbb{1}_{\mathbf{rec}})\mathbf{\Delta}\log\mathbf{C_t} + \eta\mathbf{\Delta}\log\mathbf{Y_{t+1}} + \alpha\mathbf{A_t} + \varepsilon_{\mathbf{t+1}}$								
	$\Delta\log C_t$ only							
γ	0.483***	(0.069)	0.843***	(0.062)	0.449	0.815***	(0.047)	0.512
			[0.822, 1.257]			[0.713, 1.151]		
δ	-0.314*	(0.177)	-0.646***	(0.175)	0.335	-0.793***	(0.179)	0.425
			[-2.231,-0.242]			[-2.368,-0.532]		
KP F-stat, $\overline{R}^2$	–	0.191	*6.10*		0.084	*9.31*		0.140
	$\Delta\log C_t$ and controls							
γ	0.343***	(0.062)	0.519***	(0.081)	0.306	0.577***	(0.074)	0.385
δ	-0.193	(0.168)	-0.239	(0.184)	0.317	-0.335*	(0.171)	0.354
η	0.201***	(0.040)	0.303***	(0.059)	0.176	0.226***	(0.042)	0.208
α	-7.35e-04**	(3.22e-04)	-5.55e-04**	(2.20e-04)	0.950	-4.37e-04**	(2.21e-04)	0.966
KP F-stat, $\overline{R}^2$	–	0.327	*2.01*		0.238	*2.53*		0.287

The $\mathbb{1}_{10\%}$ indicator variable equals one for the top ten percent of $|\Delta\log C_t - \overline{\Delta\log C_t}|$ where $\overline{\Delta\log C_t}$ is the average consumption growth over 1952q4:2019q1. The $\mathbb{1}_{\text{rec}}$ indicator variable equals one for recessionary periods (NBER). The sample runs from 1954q4 to 2019q1 for OLS and IV 1 and from 1967q2 to 2015q4 for IV 2. The IV estimates are estimated using the generalized method of moments. The first set of instrument (IV 1) includes two to four lags in consumption growth, income growth, unemployment, 3-month Treasury Bill rate, personal consumption expenditures price deflator volatility, consumer sentiment (Michigan survey) and net worth as a percentage of disposable personal income (households and non-profit). The second set of instrument (IV 2) includes two to four lags in consumption growth, income growth, real interest rate from 3-month Treasury Bill rate, working hours annual growth rate (in nonfarm business sector), annualized inflation from PCE, net worth as a percentage of disposable personal income (households and non-profit), and two periods ahead GDP growth Greenbook forecasts made at $t-1$. The instruments are interacted whit the indicator variable in the first stage estimation. C_t is measured as consumption of non durables and services, Y_t as personal disposable income per capita and A_t as the ratio of net worth to disposable personal income (households and non-profit). Standard errors in parenthesis are robust (HAC up to 4 lags) and corrected for small sample size. *, ** and *** denote significance at the 10%, 5% and 1% levels respectively. The 95% confidence interval are robust to weak instruments. They are obtained using the two-step method from Andrews (2018) with a 5% threshold and implemented using the Stata command by Sun (2018).

at the 10% level only for the second set of instruments.

Overall, this analysis provides suggestive evidence that US consumption persistence is not constant over time and negatively correlates with proxies for the share of inattentive consumers, as predicted by the inattention region.[36]

7. Conclusions

Much of the macroeconomic literature is trying to make sense of the large heterogeneity in consumption responses to income windfalls, which governs the effectiveness of monetary and fiscal policies. Some of this heterogeneity correlates with observable characteristics such as holdings of liquid wealth or age. However, for the most part, this heterogeneity remains unexplained and largely absent in macroeconomic models. Motivated by recent evidence, the present paper studies a factor that can account for some of this unexplained heterogeneity: the stochasticity and state-dependence of consumer expectation adjustments.

To this end, we develop a model in which a consumer faces a fixed cost for paying attention to noisy signals and her attention choice can be a function of signal realizations. At the optimum, she faces an inattention region, and expectation adjustments are stochastic and state-dependent. Unlike in models with consumption adjustment costs, the emergence of an extensive margin of expectation adjustments does not translate into jerky household consumption dynamics at quarterly frequencies. Then, we illustrate how this model can provide novel qualitative insights about the unexplained heterogeneity in consumption responses to income windfalls by revisiting three pillar topics in the consumption literature: cross-sectional marginal propensities to consume (MPC), impulse response functions to fiscal stimuli and aggregate consumption excess smoothness.

We find that the theory provides an explanation to the state-dependence of the extensive margin of consumption responses, the over-reaction at the intensive margin, and the size effects and asymmetries in households MPC that are not attributable to borrowing, liquidity and cash-on-hand constraints. At the aggregate level, impulse response functions to fiscal stimuli are more sluggish when stimuli are small or implemented at the trough of a recession, and the excess smoothness of aggregate consumption varies significantly over the business cycle with consumer inattention.

[36] Furthermore, the results are robust to the exclusion of the bottom 20% of observations with the smallest $|\Delta \log C_t - \overline{\Delta \log C_t}|$, for which we would expect the shares of inattentive consumers to be the largest.

In future work, enriching the model with cross sectional heterogeneity in the determinants of the inattention region could improve the model quantitative predictions. In particular, analyzing how liquidity constraints interact with the inattention region seems a promising avenue. Indeed, we have found suggestive evidence that most of the negative correlation between MPC and liquid wealth (or cash-on-hand) is driven by the extensive margin of adjustments.

References

Anderson, G., Bunn, P., Pugh, A., Uluc, A., 2016. The bank of england/nmg survey of household finances. Fiscal Studies 37, 131–152.

Andreolli, M., Surico, P., 2021. Less is more: Consumer spending and the size of economic stimulus payments. CEPR Discussion Paper No. DP15918 .

Andrews, I., 2018. Valid two-step identification-robust confidence sets for gmm. Review of Economics and Statistics 100, 337–348.

Armantier, O., Nelson, S., Topa, G., Van der Klaauw, W., Zafar, B., 2016. The price is right: Updating inflation expectations in a randomized price information experiment. Review of Economics and Statistics 98, 503–523.

Armona, L., Fuster, A., Zafar, B., 2018. Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment. The Review of Economic Studies 86, 1371–1410.

Åström, K.J., 2012. Introduction to stochastic control theory. Courier Corporation.

Ballantyne, A., 2021. Household consumption: Mpc asymmetry and financial frictions.

Barnichon, R., Debortoli, D., Matthes, C., 2022. Understanding the size of the government spending multiplier: It's in the sign. The Review of Economic Studies 89, 87–117.

Bhatt, V., Kundan Kishor, N., Marfatia, H., 2020. Estimating excess sensitivity and habit persistence in consumption using greenbook forecasts. Oxford Bulletin of Economics and Statistics 82, 257–284.

Browning, M., Collado, M.D., 2001. The response of expenditures to anticipated income changes: panel data estimates. The American Economic Review 91, 681–692.

Bunn, P., Le Roux, J., Reinold, K., Surico, P., 2018. The consumption response to positive and negative income shocks. Journal of Monetary Economics 96, 1–15.

Caballero, R.J., 1995. Near-rationality, heterogeneity, and aggregate consumption. Journal of Money, Credit and Banking 27, 29–48.

Caballero, R.J., Engel, E.M., 2007. Price stickiness in ss models: New interpretations of old results. Journal of Monetary Economics 54, 100–121.

Calvo, G.A., 1983. Staggered prices in a utility-maximizing framework. Journal of Monetary Economics 12, 383–398.

Campbell, J., Deaton, A., 1989. Why is consumption so smooth? The Review of Economic Studies 56, 357–373.

Campbell, J.Y., Mankiw, N.G., 1989. Consumption, income, and interest rates: Reinterpreting the time series evidence. NBER macroeconomics annual 4, 185–216.

Carroll, C., Slacalek, J., Tokuoka, K., White, M.N., 2017. The distribution of wealth and the marginal propensity to consume. Quantitative Economics 8, 977–1020.

Carroll, C.D., 2003. Macroeconomic expectations of households and professional forecasters. the Quarterly Journal of Economics 118, 269–298.

Carroll, C.D., Crawley, E., Slacalek, J., Tokuoka, K., White, M.N., 2020. Sticky expectations and consumption dynamics. American Economic Journal: Macroeconomics 12, 40–76.

Carroll, C.D., Slacalek, J., Sommer, M., 2011. International evidence on sticky consumption growth. Review of Economics and Statistics 93, 1135–1145.

Chetty, R., Szeidl, A., 2016. Consumption commitments and habit formation. Econometrica 84, 855–890.

Christelis, D., Georgarakos, D., Jappelli, T., Pistaferri, L., Van Rooij, M., 2019. Asymmetric consumption effects of transitory income shocks. The Economic Journal 129, 2322–2341.

Cochrane, J.H., et al., 1989. The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives. American Economic Review 79, 319–337.

Coibion, O., Gorodnichenko, Y., 2015. Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. American Economic Review 105, 2644–78.

Costain, J., Nakov, A., 2011. Distributional dynamics under smoothly state-dependent pricing. Journal of Monetary Economics 58, 646–665.

Dias, D.A., Marques, C.R., 2010. Using mean reversion as a measure of persistence. Economic Modelling 27, 262–273.

Dräger, L., Lamla, M.J., 2012. Updating Inflation Expectations: Evidence from Micro-data. Economics Letters 117, 807–810.

Fagereng, A., Holm, M.B., Natvik, G.J., 2021. Mpc heterogeneity and household balance sheets. American Economic Journal: Macroeconomics 13, 1–54.

Fudenberg, D., Newey, W., Strack, P., Strzalecki, T., 2020. Testing the drift-diffusion model. Proceedings of the National Academy of Sciences 117, 33141–33148.

Fuster, A., Kaplan, G., Zafar, B., 2021. What would you do with 500? spending responses to gains, losses, news, and loans. The Review of Economic Studies 88, 1760–1795.

Gabaix, X., Laibson, D., 2001. The 6d bias and the equity-premium puzzle. NBER macroeconomics annual 16, 257–312.

Hall, R.E., 1978. Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. Journal of Political Economy 86, 971–987.

Havranek, T., Rusnak, M., Sokolova, A., 2017. Habit formation in consumption: A meta-analysis. European Economic Review 95, 142–167.

Henckel, T., Menzies, G.D., Moffatt, P.G., Zizzo, D.J., 2021. Belief adjustment: A double hurdle model and experimental evidence. Experimental Economics , 1–42.

Hsieh, C.T., 2003. Do consumers react to anticipated income changes? evidence from the alaska permanent fund. The American Economic Review 93, 397–405.

Jappelli, T., Pistaferri, L., 2010. The consumption response to income changes. Annual Review of Economics 2, 479–506.

Jappelli, T., Pistaferri, L., 2017. The economics of consumption: theory and evidence. Oxford University Press.

Jaskowski, M., van Dijk, D., 2016. First-passage-time in discrete time and intra-horizon risk measures. Unpublished manuscript .

Jung, J., Kim, J.H., Matějka, F., Sims, C.A., 2019. Discrete actions in information-constrained decision problems. The Review of Economic Studies 86, 2643–2667.

Kaplan, G., Violante, G.L., 2014. A model of the consumption response to fiscal stimulus payments. Econometrica 82, 1199–1239.

Kaplan, G., Violante, G.L., 2022. The Marginal Propensity to Consume in Heterogeneous Agent Models. Technical Report. National Bureau of Economic Research.

Khaw, M.W., Stevens, L., Woodford, M., 2017. Discrete adjustment to a changing environment: Experimental evidence. Journal of Monetary Economics 91, 88–103.

Kiley, M.T., 2010. Habit persistence, nonseparability between consumption and leisure, or rule-of-thumb consumers: Which accounts for the predictability of consumption growth? The Review of Economics and Statistics 92, 679–683.

Kueng, L., 2018. Excess sensitivity of high-income consumers. The Quarterly Journal of Economics 133, 1693–1751.

Kumar, S., Jia, P., 2019. Financial crisis and persistence: evidence from sticky expectations consumption growth model. Applied Economics 51, 1799–1807.

Ljungqvist, L., Sargent, T.J., 2004. Recursive macroeconomic theory (Second edition). Cambridge, Mass: MIT Press.

Luo, Y., Nie, J., Wang, G., Young, E.R., 2017. Rational inattention and the dynamics of consumption and wealth in general equilibrium. Journal of Economic Theory 172, 55–87.

Luo, Y., Young, E.R., 2014. Signal Extraction and Rational Inattention. Economic Inquiry 52, 811–829.

Maćkowiak, B., Matějka, F., Wiederholt, M., 2018. Dynamic rational inattention: Analytical results. Journal of Economic Theory 176, 650–692.

Maćkowiak, B., Wiederholt, M., 2009. Optimal sticky prices under rational inattention. American Economic Review 99, 769–803.

Maćkowiak, B., Wiederholt, M., 2015. Business Cycle Dynamics under Rational Inattention. The Review of Economic Studies 82, 1504–1532.

Mankiw, N.G., Reis, R., 2002. Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. The Quarterly Journal of Economics 117, 1295–1328.

Mankiw, N.G., Reis, R., Wolfers, J., 2003. Disagreement about Inflation Expectations. NBER Working Papers 9796. National Bureau of Economic Research, Inc.

Miao, J., Wu, J., Young, E.R., 2022. Multivariate rational inattention. Econometrica 90, 907–945.

Misra, K., Surico, P., 2014. Consumption, income changes, and heterogeneity: Evidence from two fiscal stimulus programs. American Economic Journal: Macroeconomics 6, 84–106.

Molin, A., Hirche, S., 2010. Structural characterization of optimal event-based controllers for linear stochastic systems, in: Proceedings of the IEEE International Conference on Decision and Control (CDC 2010).

Molin, A., Hirche, S., 2017. Event-triggered state estimation: An iterative algorithm and optimality properties. IEEE Transactions on Automatic Control 62, 5939–5946.

Ocal, N., Osborn, D.R., 2000. Business Cycle Non-Linearities in UK Consumption and Production. Journal of Applied Econometrics 15, 27–43.

Parker, J.A., Souleles, N.S., Johnson, D.S., McClelland, R., 2013. Consumer spending and the economic stimulus payments of 2008. The American Economic Review 103, 2530–2553.

Parker, J.M., Jonathan, A., Souleles, N., 2006. Household expenditure and the income tax rebates of 2001. The American Economic Review 96, 1589–1610.

Pischke, J.S., 1995. Individual income, incomplete information, and aggregate consumption. Econometrica: Journal of the Econometric Society , 805–840.

Reis, R., 2006. Inattentive consumers. Journal of Monetary Economics 53, 1761–1800.

Roth, C., Wohlfart, J., 2020. How do expectations about the macroeconomy affect personal expectations and behavior? Review of Economics and Statistics 102, 731–748.

Scholnick, B., 2013. Consumption smoothing after the final mortgage payment: Testing the magnitude hypothesis. Review of Economics and Statistics 95, 1444–1449.

Shi, D., Shi, L., Chen, T., 2016. Event-based state estimation. Stud. in Syst. Decis. & Control , 47–75.

Sims, C.A., 2003. Implications of Rational Inattention. Journal of Monetary Economics 50, 665–690.

Sommer, M., 2007. Habit formation and aggregate consumption dynamics. The BE journal of macroeconomics 7.

Sun, L., 2018. Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models. The Stata Journal 18, 803–825.

Tutino, A., 2013. Rationally Inattentive Consumption Choices. Review of Economic Dynamics 16, 421–439.

Wang, N., 2003. Caballero meets bewley: The permanent-income hypothesis in general equilibrium. American Economic Review 93, 927–936.

Woodford, M., 2009. Information-Constrained State-Dependent Pricing. Journal of Monetary Economics 56, S100–S124.

Appendix A. Solution to the consumer's problem in Section 2

This first section of the appendix provides proofs for Lemmas 1 to 4. These results, and the following demonstrations, closely follow from a series of paper in engineering by Molin and Hirche. The only deviation from the mathematical framework that they consider consists in the introduction of discounting. It then concludes with the proof for Proposition 1.

Appendix A.1. Proof of Lemma 1: Consumption certainty equivalence

This subsection demonstrates that the optimal consumption choice of the consumer coincides with the certainty equivalent one. The demonstration relies on five steps. First, it computes an alternative formula for the consumer's value function at time 0. Second, it shows that the attention choices depend on the shock realizations and are not directly affected by consumption choices. This result arises because of the nestedness of the information structure. Third, it finds that the perception errors at the consumption choice are also independent of the consumption policy. Fourth, it shows that it is optimal to set consumption to its certainty equivalent for any fixed attention strategy that depends only on the shock realizations. Last, it concludes that the certainty equivalence holds independently of the chosen attention policy.

We can use the quadratic structure of the objective function V_0 to rewrite it in a more convenient form. To this end, define p_t from the following (backward) dynamic Ricatti equation $p_t = (1+r)^2\beta p_{t+1}/(1+\beta p_{t+1})$ with terminal condition $p_T = q_T$ and $L_t \equiv (1+r)\beta p_{t+1}/(1+\beta p_{t+1})$. Then, we can use the following identity (used in, e.g., Åström (2012), Chapter 8 - Proof of Lemma 6.1)

$$\beta^T q_T s_T^2 \quad = \quad p_0 \bar{s}_0^2 + \sum_{t=0}^{T-1} \left(\beta^{t+1} p_{t+1} s_{t+1}^2 - \beta^t p_t s_t^2\right) \tag{A.1}$$

where we have used the terminal condition $p_T = q_T$. Moreover, using the flow budget constraint we can write $p_{t+1}s_{t+1}^2 = ((1+r)s_t - u_t + \zeta_{t+1})^2 p_{t+1}$ where $u_t \equiv c_t - \bar{c}$. Further noticing from the definition of L_t that $p_{t+1}(1+r)s_t u_t = \frac{1+\beta p_{t+1}}{\beta} L_t s_t u_t$ and $p_{t+1}u_t^2 = \frac{1+\beta p_{t+1}}{\beta} u_t^2 - \frac{1}{\beta} u_t^2$, it holds

$$\begin{aligned} E[\beta^{t+1} p_{t+1} s_{t+1}^2 | \mathcal{I}_0] \quad &= \quad E\Big[\beta^t (u_t - L_t s_t)^2 (1+\beta p_{t+1}) + \beta^{t+1} p_{t+1} \zeta_{t+1}^2 \\ &\quad + \beta^{t+1}(1+r)^2 p_{t+1} s_t^2 - \beta^t (1+\beta p_{t+1}) L_t^2 s_t^2 - \beta^t u_t^2 \Big| \mathcal{I}_0\Big] \end{aligned} \tag{A.2}$$

because ζ_{t+1} is independent with respect to u_t and s_t and nil in expectation. Moreover, $\beta^t p_t s_t^2 = \beta^t (1+r) L_t s_t^2$ so that equation (A.1) writes in expectation

$$E[\beta^T q_T s_T^2 | \mathcal{I}_0] \quad = \quad E\Big[p_0 \bar{s}_0^2 + \sum_{t=0}^{T-1} \beta^t (u_t - L_t s_t)^2 (1+\beta p_{t+1}) + \beta^{t+1} p_{t+1} \zeta_{t+1}^2 - \beta^t u_t^2 \Big| \mathcal{I}_0\Big]$$

Consequently, the objective function $V_0 \equiv E\Big[\sum_{t=0}^{T-1} \beta^t \Big(u_t^2 + \lambda \tau_t\Big) + \beta^T q_T s_T^2 \Big| \mathcal{I}_0\Big]$ is

$$V_0 \quad = \quad E\Big[p_0 \bar{s}_0^2 + \sum_{t=0}^{T-1} \beta^t \lambda \tau_t + \beta^{t+1} p_{t+1} \zeta_{t+1}^2 + \beta^t (u_t - L_t s_t)^2 (1+\beta p_{t+1}) \Big| \mathcal{I}_0\Big] \tag{A.3}$$

Given this expression for the value function, we can now prove that a separation result holds for the consumption and attention choices.

The attention policy can be equivalently expressed as a function of the

random variable realizations, i.e., noises, permanent income innovations, and initial conditions. To see that, consider an alternative formulation of problem (3) without consumption choices. This alternative formulation allows to study the attention choice in the absence of consumption choice. It is similar to the initial problem (3) excepted that it imposes that consumption is constant over time and equal to $\bar{c}$. In particular, all shock realizations are the same in both problems. Let $\tilde{z}_t$ be the signal that would be received at period t in the alternative problem. By definition, these signals are independent of consumption choices (because generated from a system without consumption choices). Nevertheless, simple algebra shows that these signals are related to the signals (z_t) received in the initial problem (3) with a consumption choice such that $\tilde{z}_t \equiv z_t + \sum_{k=0}^{t-1}(1+r)^{t-1-k}u_k$ with, again, $u_t \equiv c_t - \bar{c}$. Hence, for a given sequence of consumption choices $\{c_k\}_{k=0}^{t-1}$, $\tilde{z}_t$ is a bijective mapping of z_t for all $t \in \{0, 1, \ldots, T-1\}$. Since past consumption choices are observable from $\mathcal{I}_t$, the attention policy $g_t(\mathcal{I}_t)$ may as well be expressed by another mapping g'_t that depends only on the sequence $\{\tilde{z}_k\}_{k=0}^{t}$. Therefore, we have

$$\tau_t = g_t(\mathcal{I}_t) = g'_t(\{\tilde{z}_k\}_{k=0}^{t}) \quad \forall t \in \{0, 1, \ldots, T-1\} \tag{A.4}$$

Broadly speaking and using the two agents analogy discussed in Section 2 to describe the information structure, this result states that the individual managing consumption cannot distort nor manipulate the behavior of the agent tracking information because the latter is always better informed (not strictly). Therefore, for a given history of past consumption choices made by the former individual, the latter individual only conditions the transmission of information on the realizations of the random variables (and initial state, i.e., the primitives of the model).

For a given attention policy that is a function of the random variable realizations, the perception error from $\overline{\mathcal{I}}_t$ – i.e, at the consumption choice – **is independent of the consumption policy.** We demonstrate this result in the following. Let the attention policy be a function of the primitives, i.e., let the $g'_t(.)$ be fixed. Then, the attention choices $\tau_t = g'_t(.)$ are random variables also independent of the consumption policy $f_t(.)$. Furthermore, define $\epsilon_t \equiv s_t - E[s_t|\overline{\mathcal{I}}_t]$ the perception error from the information set at the consumption choice. Then, consider again the model with fixed consumption equal to $\bar{c}$ at all t. The evolution of permanent income in this model follows from $\tilde{s}_{t+1} = (1+r)\tilde{s}_t + \zeta_{t+1}$. Iterating backward, we can rewrite

$$s_t = (1+r)^t s_0 + \sum_{k=0}^{t-1}(1+r)^{t-1-k}u_k + \sum_{k=0}^{t-1}(1+r)^{t-1-k}\zeta_{k+1} \tag{A.5}$$

and

$$\tilde{s}_t = (1+r)^t s_0 + \sum_{k=0}^{t-1}(1+r)^{t-1-k}\zeta_{k+1} \tag{A.6}$$

In turn, because the consumption choices are observable from $\overline{\mathcal{I}}_t$, the above two equations imply that $\epsilon_t = s_t - E[s_t|\overline{\mathcal{I}}_t] = \tilde{s}_t - E[\tilde{s}_t|\overline{\mathcal{I}}_t]$. That is, given $\overline{\mathcal{I}}_t$, the estimation error is independent of the consumption policy $f_t(.)$ for all t. We now have to show that $E[\tilde{s}_t|\overline{\mathcal{I}}_t]$ is also independent of $f_t(.)$. First, recall that we have seen that when the $g_t'(.)$ are fixed, the τ_t only depend on the primitives so that they are the same in the models with and without consumption choices. Second, realize that there is no restriction on the information acquired by the consumer when attentive ($\tau_t = 1$) as long as the nestedness property of the information structure holds. As a result, $\sigma(\overline{\mathcal{I}}_t) = \sigma(\mathcal{I}_t)$ whenever $\tau_t = 1$. This is equivalent to saying that when attentive, it is 'as if' the consumer was observing all past signals. We have already seen that the signals $\tilde{z}_t$ are a bijective mapping of the signals z_t when the consumption choices (up to $t-1$) are known. That is, if $l \leq t$ is the last period when the consumer was attentive, it must be that

$$\sigma\Big(\Big\{\{\tau_k\}_0^t, \{c_k\}_0^{t-1}, \{\{\tilde{z}_t\}_0^l, \{\emptyset\}_l^t\Big\}\Big) = \sigma(\overline{\mathcal{I}}_t) \tag{A.7}$$

so that $E[\tilde{s}_t|\overline{\mathcal{I}}_t] = E[\tilde{s}_t|\Big\{\{\tau_k\}_0^t, \{\{\tilde{z}_t\}_0^l, \{\emptyset\}_l^t\Big\}]$ because $\tilde{s}_t$ is independent of consumption. Again realizing that for fixed $g_t'(.)$ the τ_t (and therefore l) are functions of primitives, then it is clear that $E[\tilde{s}_t|\overline{\mathcal{I}}_t]$ only depends on primitives and, hence, is independent of the consumption policy. To sum up, we have shown that the estimation error ϵ_t only depend on primitives for given $g_t'(.)$.

For a given attention policy that is a function of the primitives, we have a certainty equivalence for the optimal consumption policy. We now want to characterize the optimal consumption policy for given $g_t'(.)$. Again realizing that this implies that the τ_t are random variables and independent of consumption policy, minimizing V_0 in equation (A.3) with respect to consumption is equivalent to minimizing

$E[\sum_{t=0}^{T-1}\beta^t(u_t - L_t s_t)^2(1+\beta p_{t+1})|\mathcal{I}_0]$ with respect to $u_t \equiv c_t - \bar{c}$. Furthermore, we have

$$\begin{aligned} E\Big[(u_t - L_t s_t)^2\Big|\mathcal{I}_0\Big] &= E\Big[\Big(u_t - L_t(E[s_t|\overline{\mathcal{I}}_t] + \epsilon_t)\Big)^2\Big|\mathcal{I}_0\Big] \\ &= E\Big[(u_t - L_t(E[s_t|\overline{\mathcal{I}}_t])^2 - 2(u_t - L_t(E[s_t|\overline{\mathcal{I}}_t])\epsilon_t + \epsilon_t^2\Big|\mathcal{I}_0\Big] \quad \text{(A.8)} \end{aligned}$$

Where the first equality uses the definition of $\epsilon_t \equiv s_t - E[s_t|\overline{\mathcal{I}}_t]$. By applying the power property of conditional expectations, we obtain

$$\begin{aligned} E\Big[(u_t - L_t(E[s_t|\overline{\mathcal{I}}_t])\epsilon_t\Big|\mathcal{I}_0\Big] &= E\Big\{E\Big[\Big(u_t - L_t(E[s_t|\overline{\mathcal{I}}_t]\Big)\epsilon_t\Big|\overline{\mathcal{I}}_t\Big]\Big|\mathcal{I}_0\Big\} \\ &= E\Big\{\Big(u_t - L_t(E[s_t|\overline{\mathcal{I}}_t]\Big)E\Big[\epsilon_t\Big|\overline{\mathcal{I}}_t\Big]\Big|\mathcal{I}_0\Big\} = 0 \quad \text{(A.9)} \end{aligned}$$

where the second equality uses the fact that $u_t = f_t(\overline{\mathcal{I}}_t)$ is a measurable function with respect to $\overline{\mathcal{I}}_t$ and the last equality relies on $E[\epsilon_t|\overline{\mathcal{I}}_t] = E[s_t|\overline{\mathcal{I}}_t] - E\Big[E[s_t|\overline{\mathcal{I}}_t]\Big|\overline{\mathcal{I}}_t\Big] = 0$. Recalling that the estimation error ϵ_t is independent of the policy function, we see that $u_t = L_t E[s_t|\overline{\mathcal{I}}_t]$ minimizes V_0 for given $g'_t(.)$. This is the certainty equivalent consumption policy.

Certainty equivalence. So far, we have reach a certainty equivalence result conditional on given g'_t. We now demonstrate that this result hold independently of g'_t. The intuition behind this broader result comes from the observation that the optimal consumption policy for given g'_t is independent of g'_t. To formalize this result, let g and f respectively refer to the sequence of attention and consumption policies. For each and every admissible pair (g, f), we have seen that it leads to another admissible pair (g', f), where the g'_t are functions of the primitives, that outputs the same τ_t. We have also seen that for a given g' it is always optimal to set the consumption policy to its certainty equivalent $f^\star$. Because $(g', f^\star)$ are admissible policies, we can retrieve admissible policies $(\hat{g}, f^\star)$ that dominate (g, f) in the sense of minimizing V_0. The existence of $\hat{g}$ follows from equation (A.4) when $f^\star$ is given. This concludes the proof for Lemma 1. As can be seen, this proof rests on the linear-quadratic structure of the problem and the nestedness property of the information structure.

Appendix A.2. Proof of Lemma 2: Kalman filter at the attention choice

It is easy to show that the optimal estimator given the information at the attention choice $\mathcal{I}_t$ is a Kalman filter. Apply the tower property of conditional expectations with respect to $\mathcal{I}_t$ in equation (A.3). Thanks to the nestedness property of the information structure, $E[s_t|\mathcal{I}_t]$ is the least-squares estimator.[37] The linearity of the state dynamics, the Gaussian structure of the initial state, and the (i.i.d.) noises and shocks, the least-squares estimator is the Kalman filter here.

As is stated in the text, we impose that the Kalman filter is initially at its steady state. The steady state prior variance solves the algebraic Riccati equation $p_+ = (1+r)^2\Big(p_+ - p_+^2(p_+ + \sigma_\vartheta^2)^{-1}\Big) + \sigma_\zeta^2$. The Kalman gain is $K = p_+(p_+ + \sigma_\vartheta^2)^{-1}$ and the posterior steady state variance is $p_- = (1-K)p_+$. Normalizing the initial uncertainty to $\sigma_{s_0}^2 = p_-$, we have that $E[(s_t - E[s_t|\mathcal{I}_t])^2|\mathcal{I}_0] = p_-$ remains at its steady state. (Q.E.D. Lemma 2).

In order to characterize the optimal attention strategy, it will be useful to incorporate the result in Lemma 2 into the objective function V_0. Using the optimal consumption policy in Lemma 1, the last term of V_0 in equation (A.3) becomes $\beta^t L_t^2(E[s_t|\overline{\mathcal{I}}_t] - s_t)^2(1+\beta p_{t+1})$ where $s_t - E[s_t|\overline{\mathcal{I}}_t] = s_t - E[s_t|\mathcal{I}_t] + e_t$ and $e_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t]$ is the expectation wedge between the two information sets. Thus,

$$E[(s_t - E[s_t|\overline{\mathcal{I}}_t])^2|\mathcal{I}_0] \quad = \quad E[(s_t - E[s_t|\mathcal{I}_t])^2|\mathcal{I}_0] + E[e_t^2|\mathcal{I}_0] \tag{A.10}$$

as the estimation error from $E[s_t|\mathcal{I}_t]$ is independent from e_t (as e_t is observable from $\mathcal{I}_t$). Hence,

$$V_0 \quad = \quad E\Big[p_0\bar{s}_0^2 + \sum_{t=0}^{T-1}\beta^t\lambda\tau_t + \beta^{t+1}p_{t+1}\zeta_{t+1}^2 + \beta^t L_t^2\Big((s_t - E[s_t|\mathcal{I}_t])^2 + e_t^2\Big)(1+\beta p_{t+1})\Big|\mathcal{I}_0\Big] \tag{A.11}$$

Appendix A.3. Proof of Lemmas 3 and 4: Optimal attention strategy

In this section, we derive the optimal estimator $E[s_t|\overline{\mathcal{I}}_t]$ driving the consumption choice and reported in Lemma 3. Because of the nontrivial interaction between this estimator and the attention policy, we must solve for both simultaneously. This requires us to also prove Lemma 4 in this section.

We first consider the simple case **when the consumer is attentive** at time t ($\tau_t = 1$). Looking at the expression for V_0 in equation (A.11) and using the tower property of conditional expectations (for $E[\bullet|\overline{\mathcal{I}}_t, \tau_t = 1]$), it is clear that in this case $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1]$ must minimize the square of the error $e_t = E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1]$. Because the

[37] More precisely, $E[s_t|\mathcal{I}_t]$ is the *discounted* least-squares estimator. However, since we assume that the Kalman filter is initially at its steady state, this is also the least-squares estimator here.

consumer is able to observe the information $\mathcal{I}_t$ when attentive, then it is optimal to set $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1] = E[s_t|\mathcal{I}_t]$ when the consumer is attentive.

We now focus on the more complex situation **when the consumer is inattentive** ($\tau_t = 0$). The difficulty arises from the fact that when inattentive, the consumer realizes that she is inattentive. Indeed, since the choice to be attentive may be conditioned on certain realizations of the random variables, the consumer may adjust her perceived permanent income when inattentive to account for the fact that these conditions are not met while she remains inattentive. This is a form of negative information, i.e., the consumer knows that she doesn't know, which requires to simultaneously characterize the optimal attention strategy and the evolution of perceptions while the consumer is inattentive.

To this end, we proceed in five steps. First, we derive a necessary optimality condition on the corrective terms arising in consumer's perceptions when she remains inattentive. These corrections are found to be predetermined for a given attention policy, i.e., unrelated to the random variable realizations whilst the consumer is inattentive. Second, we characterize the problem related to the joint determination of the attention policy and the corrections. Third, we characterize the optimal attention policy when the corrective terms are nil. The optimal attention policy is found to be symmetric when the corrective terms are nil. Fourth, we show that the corrective terms implied by the necessary optimality condition from the first step are indeed zero when the attention policy is symmetric. This characterizes a candidate solution for the problem under consideration. Last, we rely on Theorem 1 in Molin and Hirche (2017) to further conclude that this solution is globally asymptotically stable.

Any correction in the consumer's perception must be predetermined while she is inattentive. To demonstrate this, take the attention policies $g'_k(.)$ as functions of the primitives as given $\forall k \in \{0, 1, \ldots, t\}$ and let $l_t \equiv \sup\{k : \tau_k = 1, k \leq t\}$ be the most recent period when the consumer was attentive. Consider a period t when the consumer is inattentive ($l_t < t$). Then, compute $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0] = E[E[s_t|\mathcal{I}_t]|\overline{\mathcal{I}}_t, \tau_t = 0]$ to get

$$\underbrace{E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0]}_{\text{estimate when inattentive}} = \underbrace{(1+r)E[s_{t-1}|\overline{\mathcal{I}}_{t-1}] - u_{t-1}}_{\text{update}} + \underbrace{E\left[(1+r)e_{t-1} + K(z_t - E[s_t|\mathcal{I}_{t-1}])\middle|\overline{\mathcal{I}}_t, \tau_t = 0\right]}_{\text{corrective term accounting for inattention } (\equiv \alpha(t,.))}$$

This is equation (6) in the text. The error terms $z_t - E[s_t|\mathcal{I}_{t-1}]$ correspond to the estimation errors from the Kalman filter in Lemma 3 and are therefore orthogonal to the information in $\mathcal{I}_{t-1}$. The dynamics of the expectation wedge e_t is given by $e_t = (1-\tau_{t-1})(1+r)e_{t-1} + K(z_t - E[s_t|\mathcal{I}_{t-1}]) - \alpha(t,.)$. It therefore follows an AR(1) process with a reseting at 0 when the consumer is attentive. Hence, we can also write $e_t = \sum_{k=l_t+1}^{t}(1+r)^{t-k}\Big[K(z_k - E[s_k|\mathcal{I}_{k-1}]) - \alpha(k,.)\Big]$. Consequently, the corrective term also writes

$$\alpha(t,.) = E\Big[K\sum_{k=l_t+1}^{t}(1+r)^{t-k}(z_k - E[s_k|\mathcal{I}_{k-1}])\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big] - \sum_{k=l_t+1}^{t-1}(1+r)^{t-1-k}\alpha(k,.) \tag{A.12}$$

By definition of the information set $\overline{\mathcal{I}}_t$, we have $\{\overline{\mathcal{I}}_t|\tau_t = 0\} = \Big\{\mathcal{I}_{l_t}, \{c_k\}_{k=l_t}^{t-1}, \{\tau_k = 0\}_{k=l_t}^{t}\Big\}$. Because the estimation errors from the Kalman filter are orthogonal to the information in $\mathcal{I}_{l_t}$ and the consumption choices since l_t, the only information that may potentially turn out to be relevant for the consumer here is that we are at period t and that the last time she was attentive was at period l_t. Thus, $\alpha(t, l_t)$ takes only two arguments: time t and the last period when the consumer was attentive l_t. It is therefore predetermined and unrelated to the shock and noise realizations.

Deriving the problem related to the joint determination of the attention policy and the corrective terms. A difficulty in characterizing the corrective term $\alpha(t, l_t)$ is that it depends on the attention policy $g_t(.)$ and vice versa. The objective to minimize is given by V_0 in equation (A.11). Noting that only the second and fourth terms in (A.11) depend on the attention policy $g_t(.)$ and that we have found $e_t = 0$ when the consumer is attentive ($\tau_t = 1$) leads to the following problem

$$\begin{aligned} \min_{\{\tau_t, \alpha(t,l_t)\}_{(0\le t\le T-1, 0\le l_t<t)}} \quad & E\Big[\sum_{t=0}^{T-1}\beta^t\lambda\tau_t + (1-\tau_t)\beta^t L_t^2(1+\beta p_{t+1})e_t^2\Big|\mathcal{I}_0\Big] \\ \text{s.t.} \quad e_{t+1} &= (1-\tau_t)(1+r)e_t - \alpha(t+1, l_{t+1}) + \omega_{t+1} \\ l_{t+1} &= \tau_t t + (1-\tau_t)l_t \end{aligned} \tag{A.13}$$

where $\omega_{t+1} \equiv K\Big(z_{t+1} - (1+r)E[s_t|\mathcal{I}_t] + c_t - \bar{c}\Big)$ is the innovation from the latent Kalman filter and is an i.i.d Gaussian white noise with variance $\sigma_\omega^2 = K^2(p_+ + \sigma_\vartheta^2)$. The fact that we are restricting the admissible corrective terms to depend only on time t and the date when the consumer was last attentive l_t follows from equation (A.12), which must hold at the optimum of problem (A.13).

Assume that the corrective terms are nil, then the optimal attention policy only depends on e_t and is symmetric around zero. To demonstrate this, impose

$\alpha(t, l_t) = 0$ for all t and $l_t < t$. Then (A.13) is a standard dynamic programming problem with a binary choice and its Bellman form is given by

$$\begin{aligned} J_t(e_t) &= \min_{\tau_t \in \{0,1\}} (1-\tau_t) L_t^2 (1+\beta p_{t+1}) e_t^2 + \tau_t \lambda + \beta E[J_{t+1}(e_{t+1}) | \mathcal{I}_t] \quad \text{(A.14)} \\ \text{s.t.} \quad e_{t+1} &= (1-\tau_t)(1+r)e_t + \omega_{t+1} \\ l_{t+1} &= \tau_t t + (1-\tau_t) l_t \end{aligned}$$

Note that the expectation is taken over $\mathcal{I}_t$ because the choice to be attentive is taken with respect to this information set. The e_t are measurable given $\mathcal{I}_t$. Consequently, (A.13) is a dynamic problem with perfect state observation (e_t, l_t) .

We can solve this problem backward. At the last period T, we have $J_T = q_T e_T^2 = q_T\Big((1-\tau_{T-1})(1+r)e_{T-1}\Big)^2$. The last equality uses $\omega_T = 0$ as there is no new signal observed at the last period. Hence, $J_T(e_T)$ is observable from $\mathcal{I}_{T-1}$ and the consumer will be attentive at $T-1$ whenever

$$\frac{((1+r)\beta q_T)^2}{1+\beta q_T} e_{T-1}^2 + q_T\Big((1+r)e_{T-1})\Big)^2 \geq \lambda$$

As stated in the text, we consider a cost associated to the terminal condition q_T to be arbitrarily large. Rearranging the above equation and taking the limit, it is clear that as $q_T \mapsto \infty$ the above inequality becomes $(1+r)e_{T-1}^2 + \Big((1+r)e_{T-1}\Big)^2 \geq 0$ which holds for any $e_{T-1} \in \mathbb{R}$. That is, when the penalty q_T is large enough, the consumer will almost surely be attentive at time $T-1$. Given this last observation, we have $J_{T-1}(e_{T-1}) = \lambda$ and the attention policy $g_{T-1}(.) = 1$. (Note that both are symmetric around 0 on the real line.)

At the period $T-2$, the consumer is attentive if and only if $L_{T-2}^2(1+\beta p_{T-1})e_{T-2}^2 \geq \lambda$. That is, if and only if $|e_{T-2}| \geq \frac{1}{L_{T-2}}\sqrt{\frac{\lambda}{1+\beta p_{T-1}}}$ $(\equiv \pi_{T-2})$. The resulting value function is therefore $J_{T-2}(e_{T-2}) = \Big(L_{T-2}^2(1+\beta p_{T-1})e_{T-2}^2\Big)^{\mathbb{1}(|e_{T-2}| \leq \pi_{T-2})} \lambda^{(1-\mathbb{1}(|e_{T-2}| \leq \pi_{T-2}))} + \lambda$. Clearly, we find again that the value function and the attention policy are symmetric around zero on the real line.

We now demonstrate by induction that the symmetry property of the value function in fact holds for any $t \in \{0, 1, \ldots, T-1\}$. Let $J_{t+1}(e_{t+1})$ be an even function. The conditional expectation $E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = e]$ preserves the symmetry.[38] The cost function $L_t^2(1+$

[38] We have

$$E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = e] = \int_{\mathbb{R}} J_{t+1}(x)\phi(x; e, \sigma_\omega)dx = \int_{\mathbb{R}} J_{t+1}(-x)\phi(-x; -e, \sigma_\omega)dx = E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = -e]$$

$\beta p_{t+1})e_t^2$ is even. Given that the sum of two even functions is also even, the objective to minimize is an even function. Taking the minimum with respect to τ_t preserves the symmetry so that $J_t(e_t)$ is en even function. Since $J_T(e_T)$ is an even function,[39] we have that $J_t(e_t)$ is also even for all $t \in \{0, 1, \dots, T-1\}$.

To conclude that this leads to an optimal attention policy of the form $g_t(e_t) = \mathbb{1}(|e_t| \geq \pi_t)$ follows from realizing that the objective to minimize is non-decreasing in $\mathbb{R}^+$. Moreover, it is easy to show that the threshold $\pi_t \in \mathbb{R}^+$ solves $\forall t \in \{0, 1, \dots, T-1\}$

$$\lambda + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = 0] \quad = \quad L_t^2(1 + \beta p_{t+1})\pi_t^2 + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = \pi_t] \quad \text{(A.15)}$$

That is, the threshold π_t coincides with the point (on the positive real line) where the consumer is indifferent between being attentive or not at time t.

The corrective terms implied by condition (A.12) are indeed zero when the attention strategy is $g_t(e_t) = \mathbb{1}(|e_t| \geq \pi_t)$ for all t. We demonstrate this result in the following. We have already seen that the optimal corrective terms must satisfy equation (A.12). This condition relies on a given attention policy $g_t(.)$. We now demonstrate that when using the optimal attention policy that we have characterized for $\alpha(t, l_t) = 0$ indeed implies that the necessary optimality condition (A.12) implies that they are zero.

From equation (A.12) we have that the change in the correction terms at any period $l_t < t$ when the consumer is inattentive is given by

$$\alpha(t, l_t) - \sum_{k=l_t+1}^{t-1} (1+r)^{t-1-k}\alpha(k, l_t) = K \ E\Big[\sum_{k=l_t+1}^{t} (1+r)^{t-k}\omega_k \Big| \overline{\mathcal{I}}_t, \tau_t = 0\Big] \qquad \text{(A.16)}$$

Let's characterize the distribution of the random variable inside the expectation. Applying Bayes law, the probability distribution function $f_t(x|j = t - l_t)$ of the random variable $\sum_{k=l_t+1}^{t}(1+r)^{t-k}\omega_k$ given that the consumer was last attentive j periods ago follows from the recursion

$$f_t(e|k) \quad \propto \quad \int_{-\pi_{t-1}}^{\pi_{t-1}} \phi\Big(\frac{(1+r)e - \bar{e}}{(1+r)\sigma_\omega}\Big) f_{t-1}(\bar{e}|k-1) d\bar{e} \qquad \forall k, t \in \mathbb{N}^2, t > k \quad \text{(A.17)}$$

and initial condition $f_t(\bar{e}|0) = \delta(\bar{e})$ $\forall t$ with $\delta(\bar{e})$ the Dirac distribution. We recognize that the inner term of the integral is a convolution. Importantly, we can observe that this

where the second equality uses the symmetry of both $J_{t+1}(x)$ and the Gaussian pdf $\phi(x; e, \sigma_\omega)$ with mean e and standard deviation σ_ω.

[39] Or, as we have also shown, $J_{T-1}(e_{T-1})$ and $J_{T-2}(e_{T-2})$.

recursion preserves the symmetry of $f_t(e|k)$ when the integral is taken with respect to a symmetric set around zero. We can therefore conclude that $E\Big[\sum_{k=l_t+1}^{t}(1+r)^{t-k}\omega_k\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big] = \int_{-\pi_t}^{\pi_t} e f_t(e|k)de = 0$ for all t and $l_t < t$.

Consequently, the right hand side of equation (A.16) equals zero. Further realizing that $\forall t: \ \alpha(t, t-1) = \int_{-\pi_t}^{\pi_t} e f_t(e|1)de = \int_{-\pi_t}^{\pi_t} e\phi(e/\sigma_\omega)de = 0$ leads to the conclusion that $\alpha(t, l_t) = 0$ for all t and $l_t < t$.

The candidate solution that we have identified is globally asymptotically stable. We have shown in this subsection that an attention policy such that the consumer becomes attentive if and only if $|e_t| \geq \pi_t$, where the thresholds π_t can be computed from equation (A.15), and the consumer's expectations are given by $E[s_t|\mathcal{I}_t]$ when the consumer is attentive and a mechanical update without corrective term when inattentive. We have shown that this solution satisfies standard necessary conditions. However, it could be that this optimum is only local. Thanks to Theorem 1 in Molin and Hirche (2017) we can however conclude that this solution is globally asymptotically stable. More specifically, their Theorem shows (by means of Lyapunov stability) that if we start from any fixed sequence of corrective terms $\alpha_0(t, l_t)$ (instead of imposing that they are zero), solve the optimal attention policy from problem (A.13) for this sequence of corrective terms $\alpha_0(t, l_t)$, use equation (A.16) to update the sequence of $\alpha_1(t, l_t)$ and iterate until convergence, then we would asymptotically converge to the solution with $\alpha_\infty(t, l_t) = 0$ for all t and $l_t < t$. The Online Appendix F indicates how to transform the variables from the model considered in the present paper so that it can be directly mapped into their theorem. The key assumptions in our framework that allows to rely on their theorem is that the shocks to permanent income and information noises follow a Gaussian distribution.

Appendix A.4. Proof of Proposition 1

We now demonstrate that the attention and consumption policies converge to stationary policies. As we have seen, the consumption policy is the certainty equivalent policy irrespectively of the attention policy. Therefore, we first characterize the stationary consumption policy. It simply requires to show that p_t converges to a stationary solution. To this end, consider the following infinite horizon deterministic linear quadratic control problem:

$$\begin{aligned} \min_{\{c_t\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t (c_t - \bar{c})^2 \\ \text{s.t.} \quad x_{t+1} \quad &= \quad (1+r)x_t - c_t + \bar{c} \end{aligned} \tag{A.18}$$

where $\beta \in (0,1)$. Assuming it exists,[40] it is well-known that the stationary control law is $\bar{L} = (1+r)\frac{\beta\bar{p}}{1+\beta\bar{p}}$ where $\bar{p}$ is the solution to algebraic Ricatti equation

$$\bar{p} \quad = \quad (1+r)^2 \frac{\beta\bar{p}}{1+\beta\bar{p}} \tag{A.19}$$

that is, $\bar{p} = \frac{\beta(1+r)^2-1}{\beta}$. Thanks to the certainty equivalence in the original problem, the consumption policy (4) admits a stationary solution $L = (1+r)\frac{\beta\bar{p}}{1+\beta\bar{p}} = \frac{\beta(1+r)^2-1}{\beta(1+r)}$.

Consequently, we can show that the attention problem also admits a stationary policy when T maps to infinity. When the consumption policy is stationary, $L_t^2(1+\beta p_{t+1})$ equal the constant $\frac{[\beta(1+r)^2-1]^2}{\beta(1+r)}$. Therefore, the reward function in the Bellman equation (A.14) is stationary. The latent Kalman filter being at its steady state, the distribution of its innovation is also stationary. Hence, problem (A.14) is an infinite horizon discrete time Markov decision problem where the reward, transition, constraint and shock distribution are independent of time. As such, the problem is stationary and the Bellman equation takes the form of a functional fixed-point equation

$$\begin{aligned} J(e_t) \quad &= \quad \min_{\tau_t \in \{0,1\}} \frac{\left(\beta(1+r)^2-1\right)^2}{\beta(1+r)} e_t^2 + \tau_t \lambda + \beta E[J(e_{t+1})|\mathcal{I}_t] \\ \text{s.t.} \quad & \quad\quad e_{t+1} = (1-\tau_t)(1+r)e_t + \omega_{t+1} \end{aligned} \tag{A.20}$$

We thus have $\tau_t = g(e_t)$ where $g(e_t) = 1 \iff |e_t| \geq \pi$ and 0 otherwise. π solves

$$\lambda + \beta E[J(e_{t+1})|\mathcal{I}_t, e_t = 0] \quad = \quad \frac{\left(\beta(1+r)^2-1\right)^2}{\beta(1+r)} \pi^2 + \beta E[J(e_{t+1})|\mathcal{I}_t, e_t = \pi] \tag{A.21}$$

and $J(.)$ is the stationary value function from (A.20).

Appendix B. Inattention lengths and hazard rates

Appendix B.1. Inattention lengths

How long will a consumer remain inattentive? A consumer inattention duration is stochastic. Answering this question therefore requires to derive the distribution of inattention lengths. This is the purpose of this section. It shows that this distribution is the solution to a first passage problem. It then provides a method to characterize this distribution and an easily implementable approximation procedure.

[40] See Ljungqvist and Sargent (2004) section 5.4.1 for a discussion on stability. Section 5.2.2 characterizes the solution to the problem under consideration.

The attention dynamics discussed in Section 3.1 may be apprehended as resulting from a first time passage problem with reseting at 0 when the consumers' attention is triggered. That is, how long will it take for the expectation wedge e_t to reach one of the attention regions? Formally, let $l_t \equiv \sup\{i : \tau_i = 1, i \leq t\}$ be the most recent period when the consumer was attentive and the first passage time be defined as $d \equiv \inf\{i : \tau_{l_{t-1}+i} = 1, i \in \mathbb{N}\}$. The associated probability density function is thus

$$q(k) \equiv P(d = k) = P(\tau_{l_t+k} = 1 | \cap_{i=1}^{k-1} \tau_{l_t+i} = 0) \qquad \forall k \in \mathbb{N} \tag{B.1}$$

where $q(k)$ is the probability that a consumer remains inattentive for k consecutive periods. Following from Jaskowski and van Dijk (2016), a first passage time always exists here as $P(d = \infty) = 0$ at the limit. Similarly, a finite average inattention length $\bar{d}_t \equiv \sum_{i=1}^{T-1-t} i q_{i,t+i}$ exists.

It is well-known that directly computing the probabilities $q(k)$ is difficult. Therefore, I use the relation between these probabilities and the hazard rates, denoted $\Lambda(k)$, which are easier to compute. By definition, we have that

$$\Lambda(k) \equiv 1 - \int_{\Xi} f(e|k) de \qquad \forall k \in \mathbb{N} \tag{B.2}$$

where $\Xi \equiv [-\pi, \pi]$ is the inattention region and $f(e|k)$ is the distribution of the expectation wedge e_t given that the consumer was inattentive for k consecutive periods. Equation (B.2) thus states that the hazard rate $\Lambda(k)$ is equal to the probability that the latent perceived forecast does not belong to the inattention region after k periods of inattention.

In Section 3.1 we have seen that the latent perceived forecast error follows an AR(1) process with a reseting at 0 when the consumer is attentive. Therefore, we have from Bayes law

$$f(e|k) \propto \int_{\Xi} f_{ar}(e|\bar{e}) f(\bar{e}|k-1) d\bar{e} \qquad \forall k \in \mathbb{N} \tag{B.3}$$

where $f_{ar}(e|\bar{e}) = \frac{1}{\sigma_\omega}\phi\Big(\frac{e-(1+r)\bar{e}}{\sigma_\omega}\Big)$ and the initial condition $f(e|0) = \delta(e)$ with $\delta(.)$ the Dirac distribution.

In order to compute the hazard rates in Figure 1 and the expected inattention length, it is necessary to compute the distribution $f(e|k)$ from equation (B.3). The latter distribution is not standard and explicitly iterating on equation (B.3) may lead to large numerical errors (Shi et al., 2016). Therefore, we rely on the approximation procedure presented in Lemma 5 which provides closed-form approximations. This approximation relies on a truncation of histories, a procedure which is well-suited for realistic calibrations of the problem under consideration. Indeed, the average inattentiveness length being generally of a few periods, the share of consumers who will encounter a long duration without being attentive should be small. Therefore, a good approximation method here should be close to exact for small k. As is highlighted in Lemma 5, the proposed method is exact when k is equal to one or two periods.

Lemma 5. *For $k = 1$, we have*

$$f(e|1) \quad = \quad \frac{1}{\sigma_\omega}\phi\Big(\frac{e}{\sigma_\omega}\Big)$$

and for $k = 2$,

$$f(e|2) \quad \propto \quad \phi\Big(\frac{e}{\sqrt{1+(1+r)^2}\sigma_\omega}\Big)\Big[\Phi\Big(\frac{\pi - \frac{(1+r)}{1+(1+r)^2}e}{\frac{\sigma_\omega}{\sqrt{1+(1+r)^2}}}\Big) - \Phi\Big(-\frac{\pi + \frac{(1+r)}{1+(1+r)^2}e}{\frac{\sigma_\omega}{\sqrt{1+(1+r)^2}}}\Big)\Big]$$

For higher $k \in \{3, 4, \ldots, \infty\}$, the distribution $f(e|k)$ is approximated by truncating the histories and we have

$$f^{\text{app}}(e|k) \quad \propto \quad \phi\Big(\frac{e}{\sqrt{z(k)}\sigma_\omega}\Big)\Big[\Phi\Big(\frac{\pi - \frac{(1+r)u(k)e}{z(k)}}{\sqrt{\frac{u(k)}{z(k)}}\sigma_\omega}\Big) - \Phi\Big(-\frac{\pi + \frac{(1+r)u(k)e}{z(k)}}{\sqrt{\frac{u(k)}{z(k)}}\sigma_\omega}\Big)\Big]$$

where $z(k) = \sum_{i=0}^{k-1}(1+r)^{2i}$ and $u(k) = \sum_{i=0}^{k-2}(1+r)^{2i}$.

Proof. See Online Appendix G. □

Lemma 5 also shows that the distribution of inattention lengths is driven by three parameters: the interest rate r which captures the propagation of past errors over time, the triggering threshold π which characterizes the shape of the inattention region and the variance σ_ω^2 of the expectation wedge shocks (noise + income shocks).

Appendix B.2. State-dependent hazard rate

We here compute the state-dependent hazard reported in Section 3. Let $\Delta_t \equiv s_t - E[s_t|\overline{\mathcal{I}}_{t-1}]$ be a consumer's prior error. Then,

$$\Delta_t = s_t - E[s_t|\mathcal{I}_t] + \underbrace{\underbrace{E[s_t|\mathcal{I}_t] - E[s_t|\mathcal{I}_{t-1}]}_{=\omega_t} + \underbrace{E[s_t|\mathcal{I}_{t-1}] - E[s_t|\overline{\mathcal{I}}_{t-1}]}_{=(1+r)e_{t-1}}}_{=e_t}$$

and $s_t - E[s_t|\mathcal{I}_t] \sim \mathcal{N}(0, p_-)$ is the posterior error at the Kalman filter for the attention choice. Therefore, the conditional distribution of e_t given Δ is the Gaussian distribution with mean Δ and variance p_-. Using the definition of the hazard rate, we obtain

$$\Lambda(\Delta) \quad = \quad 1 - \int_{\Xi} \phi\left(\frac{x-\Delta}{\sqrt{p_-}}\right) = 1 - \Phi\left(\frac{\pi-\Delta}{\sqrt{p_-}}\right) + \Phi\left(\frac{-\pi-\Delta}{\sqrt{p_-}}\right) \tag{B.4}$$

Appendix C. Additional proofs for the cross sectional distribution

Appendix C.1. Stationary distribution

This section characterizes the stationary distribution of expectation wedges, denoted $a^\star(e)$, using the tools developed in Appendix B. We consider that income shocks are not correlated across individuals, and relegate the introduction of aggregate shocks to the next section.

Let $\lambda_t(k)$ be the share of consumers who were inattentive for $k \in \{1, 2, \ldots, \infty\}$ periods and $a_t(e)$ the distribution of expectation wedges e at time t. Then,

$$a_t(e) \quad = \quad \sum_{k=1}^{\infty} \lambda_t(k) f_t(e|k) \tag{C.1}$$

Similarly, at the next period it must be that

$$a_{t+1}(e) \quad = \quad \sum_{k=2}^{\infty} \lambda_t(k-1)(1-\Lambda_t(k-1)) f_{t+1}(e|k) + f_{t+1}(e|1)\Big(\sum_{k=1}^{\infty} \lambda_t(k)\Lambda_t(k)\Big) \tag{C.2}$$

In the absence of aggregate shocks, the $f_t(e|k)$ are time invariant. Equalizing the RHS of Equations (C.1) and (C.2), a candidate solution is such that the stationary shares of attentive consumers $\lambda^\star(k)$ solve

$$\lambda^\star(1) \quad = \quad \sum_{k=1}^{\infty} \lambda^\star(k)\Lambda(k) \tag{C.3}$$

$$\lambda^\star(k) \quad = \quad \lambda^\star(k-1)(1-\Lambda(k-1)) \quad \forall k \geq 2 \tag{C.4}$$

$$1 \quad = \quad \sum_{k=1}^{\infty} \lambda^\star(k) \tag{C.5}$$

where $\Lambda(k)$ are the time invariant hazard rates from equation (B.2) and (C.5) ensures that the shares $\lambda^\star(k)$ sum to one. Iterating backward, Equation (C.4) writes

$$\lambda^\star(k) \quad = \quad \lambda^\star(1) \prod_{i=1}^{k-1} (1-\Lambda(i)) \quad \forall k \geq 2 \tag{C.6}$$

Noting that $S^\star(k-1) \equiv \prod_{i=1}^{k-1}(1-\Lambda(i))$ is the time invariant survival function with $S^\star(0)=1$, we may introduce this expression in (C.5) to get

$$\lambda^\star(1) = \frac{1}{\sum_{k=1}^{\infty} S^\star(k-1)} \tag{C.7}$$

Equations (C.7), (C.6) and (B.3) fully characterize the stationary cross sectional distribution

$$a^\star(e) = \sum_{k=1}^{\infty} \lambda^\star(k) f^\star(e|k) \tag{C.8}$$

As a weighted sum of unimodal and symmetric distributions centered around zero, the stationary cross sectional distribution of consumers is itself symmetric, unimodal and centered around zero.

Appendix C.2. Cross section dynamics and IRF

The derivation of the stationary distribution based on the distributions $f^\star(e|k)$ is useful insofar that Lemma 5 provides closed-form approximations for these distributions. However, it is intractable to assess the dynamics of the cross section of consumers in the presence of aggregate shocks as it would require to track $k \mapsto \infty$ distributions at each period. For this reason, we consider an alternative method in the presence of aggregate shocks $\bar{\zeta}_t$.

Starting from the law of motion for the expectation wedge in Equation (10), we have

$$\begin{aligned} e_{t+1} &= (1-\tau_t)(1+r)e_t + K\left(s_{t+1} + \vartheta_{t+1} - (1+r)E[s_t|\mathcal{I}_t] + c_t - \bar{c}\right) \\ &= (1-\tau_t)(1+r)e_t + K\left[\vartheta_{t+1} + \bar{\zeta}_{t+1} + (1+r)\left(s_t - E[s_t|\mathcal{I}_t]\right)\right] \\ &= (1-\tau_t)(1+r)e_t + K(\vartheta_{t+1} + \bar{\zeta}_{t+1}) + K(1+r)\left((1-K)\bar{\zeta}_t - K\vartheta_t\right) \\ &\quad + K(1-K)(1+r)^2\left(s_{t-1} - E[s_{t-1}|\mathcal{I}_{t-1}]\right) \end{aligned} \tag{C.9}$$

where the second equality uses the budget constraint and the third equality Lemma 2. Iterating backward and taking the limit when $t \mapsto \infty$, we have

$$e_{t+1} = (1-\tau_t)(1+r)e_t + K\left(\sum_{i=0}^{\infty}(1-K)^i(1+r)^i\bar{\zeta}_{t+1-i} + \vartheta_{t+1} - K(1+r)\sum_{i=0}^{\infty}(1-K)^i(1+r)^i\vartheta_{t-i}\right)$$

where we recognize the definition of $S_{t+1} \equiv K\sum_{i=0}^{\infty}(1-K)^i(1+r)^i\bar{\zeta}_{t+1-i}$ presented in the text. From this expression, it is clear that the conditional random variable $e_{t+1}|e_t = e, \{\bar{\zeta}_{t+1-i}\}_{i=0}^{\infty}$ follows a Gaussian distribution with mean $(1+r)(1-g(e))e + S_{t+1}$ and variance

$$\sigma_\omega^2 - \sigma_S^2 = K^2 \left[1 + \frac{(1+r)^2}{\left(1-(1-K)(1+r)\right)^2}\right] \sigma_\vartheta^2 \tag{C.10}$$

under the usual assumption that $(1+r)(1-K) < 1$. Using the law of total probability, it implies

$$\begin{aligned} a_{t+1}(e) &= \frac{1}{\sqrt{\sigma_\omega^2 - \sigma_S^2}} \int_{\mathbb{R}} \phi\left(\frac{e-(1+r)(1-g(\tilde{e}))\tilde{e} - S_{t+1}}{\sqrt{\sigma_\omega^2 - \sigma_S^2}}\right) a_t(\tilde{e}) d\tilde{e} \qquad \text{(C.11)} \\ &= \frac{1}{\sqrt{\sigma_\omega^2 - \sigma_S^2}} \Bigg[\underbrace{\int_{\tilde{e}\in\varXi} \phi\left(\frac{e-(1+r)\tilde{e} - S_{t+1}}{\sqrt{\sigma_\omega^2 - \sigma_S^2}}\right) a_t(\tilde{e}) d\tilde{e}}_{\text{Inattentive at } t} + \underbrace{\phi\left(\frac{e - S_{t+1}}{\sqrt{\sigma_\omega^2 - \sigma_S^2}}\right) \int_{\tilde{e}\notin\varXi} a_t(\tilde{e}) d\tilde{e}}_{\text{Attentive at } t} \Bigg] \end{aligned}$$

Note that the time subscript t implicitly refers to the knowledge of the full sequence of aggregate shocks $\{\bar{\zeta}_{t-i}\}_{i=0}^{\infty}$.

Aggregating household consumption changes in Equations (D.1) and (D.2) when $\beta = (1+r)^{-1}$, it then follows that the change in aggregate consumption is

$$\Delta C_{t+1} \quad = \quad L \int_{e\notin\varXi} a_{t+1}(e) de \tag{C.12}$$

Consequently, the consumption change at time t depends on the full history of aggregate shocks. We define the impulse response function (IRF) to an aggregate shock as the difference between the change in aggregate consumption induced by this change and the consumption change that would have been observed without it. Formally, the IRF s periods after the shock is

$$\Delta C_{t+s}|_{\bar{\zeta}_t=\chi} - \Delta C_{t+s}|_{\bar{\zeta}_t=0} = L \int_{e\notin\varXi} e \, \left(a_{t+s}(e)|_{\bar{\zeta}_t=\chi} - a_{t+s}(e)|_{\bar{\zeta}_t=0}\right) de \tag{C.13}$$

where we consider that there are no further destabilizing shocks $\{\bar{\zeta}_{t+i}\}_{i=1}^{\infty} = \{0\}_{i=1}^{\infty}$. The IRF are then computed from iterating on Equation (C.11) for $a_{t+s}(e)|_{\bar{\zeta}_t=\chi}$ and the counterfactual distribution $a_{t+s}(e)|_{\bar{\zeta}_t=0}$.

Appendix C.3. Simulations

In the simulations, the (evenly discretized) stationary distribution is first computed using Lemma 5 and Equation (C.8) as a first guess. We then iterate on

$$a_t(e) \propto \frac{1}{\sigma_\omega}\Bigg[\underbrace{\int_{\tilde{e}\in\Xi}\phi\left(\frac{e-(1+r)\tilde{e}}{\sigma_\omega}\right)a_{t-1}(\tilde{e})d\tilde{e}}_{\text{Inattentive at } t-1}+\underbrace{\phi\left(\frac{e}{\sigma_\omega}\right)\int_{\tilde{e}\notin\Xi}a_{t-1}(\tilde{e})d\tilde{e}}_{\text{Attentive at } t-1}\Bigg] \tag{C.14}$$

to achieve convergence. All integrals are approximated with Reiman sums. We refer to this distribution as the ergodic distribution.

The steady state distribution instead assumes that the history of aggregate shocks is $\{\bar{\zeta}_i\}_{i=0}^t = \{0\}_{i=0}^t$. Its computation is similar to the one for the ergodic distribution excepted that the distributions are conditional on the history of aggregate shocks. That is, we replace σ_ω by $\sqrt{\sigma_\omega^2 - \sigma_S^2}$.

Appendix D. Additional proofs on household consumption

This appendix analyses the consumption change following a shock to permanent income at the household level. It shows that while consumption changes are partially predictable, they are not serially correlated. Moreover, the expected marginal propensity to consume out of an income shock depends on the perceived forecast error and the permanent income shock. It formalizes the results related to the absence of persistence in household consumption and the magnitude hypothesis stated in sections 4.

Appendix D.1. Absence of persistence

At the household level, consumption changes are conditional on the updating behavior. When inattentive, the consumer follows a committed consumption path. As such, the consumption change solely reflects the trend in this consumption path and consumption growth is constant. More specifically, we have from Lemmas 1 and 3 that

$$\Delta c_{t+1}|(\tau_{t+1}=0) = (r-L)[c_t - \bar{c}] \tag{D.1}$$

As a consequence, consumption growth at non-updating periods is predetermined and orthogonal to permanent income shocks and information noises. However, at updating periods the consumer updates her information set and the consumption change

$$\Delta c_{t+1}|(\tau_{t+1}=1) = Le_{t+1} + \Delta c_{t+1}|(\tau_{t+1}=0) \tag{D.2}$$

is a Borel-measurable function with associated σ-statistics $\mathcal{I}_{t+1}$. As such, the change in consumption at updating periods depends on the complete history $\{\{\zeta_i\}_{i=1}^{t+1}, \{\vartheta_i\}_{i=1}^{t+1}\}$ and is therefore partially forecastable using past information about income shocks. In comparison, the sticky expectation models of Carroll (2003) and Reis (2006) predict that, in an otherwise similar setup, a household consumption growth would be unpredictable using information prior to her last update at time l_t. Furthermore, the following proposition about serial correlation holds

Proposition 3. *Consumption growth is not serially correlated at the household level.*

Proof. Equations (8), (D.1) and (D.2) together imply that Δc_t is orthogonal to e_t when $\tau_t = 0$ and that e_{t+1} is orthogonal to e_t when $\tau_t = 1$ so that e_{t+1} is also independent from Δc_t in that case. □

Appendix D.2. Asymmetry and size effects

Household's i change in consumption following a positive perceived shock $\omega > 0$ is

$$c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0} = \begin{cases} L\omega & \text{if} \quad -\infty < e \leq -\pi - \omega \\ L|e| & \text{if} \quad -\pi - \omega < e \leq -\pi \\ 0 & \text{if} \quad -\pi < e < \pi - \omega \\ L(\omega + e) & \text{if} \quad \pi - \omega \leq e < \pi \\ L\omega & \text{if} \quad \pi \leq e < \infty \end{cases} \tag{D.3}$$

where the first condition refers to consumers with negative e who continue to revise despite the positive shock, the second to consumers with negative e who no longer revise because of the shock, the third to consumers who do not revise with and without the shock, the fourth to consumers who revise because of the shock and the last to consumers who would have also revised without the shock.

Conditional average propensity to consume – Let us first focus on the average response of consumers who revise their consumption. Again, we consider a positive perceived shock $\omega > 0$. Given that the response is conditional on $\tau_{t|\omega_{i,t}=\omega} = 1$, we can focus on the first and last two groups in (D.3). The conditional expected change in consumption is thus for $\omega > 0$

$$\begin{aligned} E\Big[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\Big|\tau_t = 1\Big] &= L\left(\omega + \frac{\int_{\pi-\omega}^{\pi} e a_t(e)de}{1 - \int_{-\pi-\omega}^{\pi-\omega} a_t(e)de}\right) \\ \frac{E\Big[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\Big|\tau_t = 1\Big]}{L \times \omega} &= 1 + \frac{1}{\omega}\frac{\int_{\pi-\omega}^{\pi} e a_t(e)de}{1 - A_t(\pi - \omega) + A_t(-\pi - \omega)} \end{aligned} \tag{D.4}$$

where $A_t(.)$ is the cdf associated to $a_t(.)$. The second equation normalizes the response by ω and L. The normalization by ω gives the average conditional propensity to consumer. While the normalization by L allows to directly compare this value to what we would obtain in a full-information rational expectation framework as L would be the average propensity to consume. Hence, any deviation from 1 is a deviation from the response we would obtain in the absence of information frictions.[41]

Equation (D.4) shows that the conditional average response to the perceived shock is the standard response and an over-reaction. The over-reaction captures the fact that consumers close to the attention threshold have a disproportionate response to small income shocks as they also adjust for $e_{i,t}$. Moreover, we can compute the limits with respect to the size of the shock

$$\lim_{\omega \mapsto 0^+} \frac{E\Big[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\Big|\tau_t = 1\Big]}{L \times \omega} = 1 + \frac{a_t(\pi)\ \pi}{1 - A_t(\pi) + A_t(-\pi)} \tag{D.5}$$

$$\lim_{\omega \mapsto \infty} \frac{E\Big[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\Big|\tau_t = 1\Big]}{L \times \omega} = 1 \tag{D.6}$$

The first equation uses l'Hopital's rule and Leibniz integral rule for differentiation. These limits show that the over-reaction is relatively large for small shocks. However, it becomes negligible for large shocks that prompt most consumers to revise their consumption choices. In general, there is no reason to expect that the average propensity to consume decreases monotonically between these two limits. Indeed, when the cross sectional distribution is to the left or/and the shock sufficiently large, then the over-reaction can become negative. Simulations confirm this intuition, though we find that it happens for quite radical parameterizations and, even when it does arise, it does not drastically alter the overall decreasing pattern of the conditional average propensity to consume.

Doing similar computations for negative shocks $\omega < 0$, we obtain

$$\frac{E\Big[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\Big|\tau_t = 1\Big]}{L \times \omega} = 1 + \frac{1}{\omega}\ \frac{\int_{-\pi}^{-\pi-\omega} a_t(e)ede}{1 - A_t(\pi - \omega) + A_t(-\pi - \omega)} \tag{D.7}$$

$$\lim_{\omega \mapsto 0^-} \frac{E\Big[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\Big|\tau_t = 1\Big]}{L \times \omega} = 1 + \frac{a_t(-\pi)\pi}{1 - A_t(\pi) + A_t(-\pi)} \tag{D.8}$$

$$\lim_{\omega \mapsto -\infty} \frac{E\Big[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\Big|\tau_t = 1\Big]}{L \times \omega} = 1 \tag{D.9}$$

[41] Note that because the over-reaction refers to the extensive margin of adjustments, we also know from Caballero and Engel (2007) that this term would be equal to zero in models of Sticky expectations à la Carroll et al. (2020) and Reis (2006).

which leads to a similar pattern with respect to the size of the shock. Interestingly, we observe that in general the conditional average propensity is discontinuous at zero. More specifically, since we know from Equation (C.8) that the the stationary cross sectional distribution is symmetric around zero, the two limits coincide at the steady state. However, the limit from the left is larger when the cross sectional distribution is more concentrated on the left, and vice versa.

Unconditional average propensity to consume – We now turn to the unconditional average response. Following from Equation (D.3), we have for positive shocks $\omega > 0$

$$E\left[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\right] = L\left(\left(1 - \int_{-\pi-\omega}^{\pi-\omega} a_t(e)de\right)\omega + \int_{\pi-\omega}^{\pi} ea_t(e)de - \int_{-\pi-\omega}^{-\pi} ea_t(e)de\right)$$

$$\frac{E\left[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\right]}{L\times\omega} = 1 - (A_t(\pi-\omega) - A_t(-\pi-\omega)) + \frac{1}{\omega}\left(\int_{\pi-\omega}^{\pi} ea_t(e)de - \int_{-\pi-\omega}^{-\pi} ea_t(e)de\right) \quad \text{(D.10)}$$

$$\lim_{\omega\mapsto 0^+}\frac{E\left[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\right]}{L\times\omega} = 1 - (A_t(\pi) - A_t(-\pi)) + (a_t(\pi) + a_t(-\pi))\,\pi \quad \text{(D.11)}$$

$$\lim_{\omega\mapsto\infty}\frac{E\left[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\right]}{L\times\omega} = 1 \quad \text{(D.12)}$$

Equation (D.11) is well-known in the pricing literature and extensively discussed in Caballero and Engel (2007). The first three terms refer to the probability to update at 0 and the last term to the extensive margin. Turning to negative perceived shocks $\omega < 0$. we have

$$E\left[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\right] = L\left(\left(1 - \int_{-\pi-\omega}^{\pi-\omega} a_t(e)de\right)\omega + \int_{-\pi}^{-\pi-\omega} ea_t(e)de - \int_{\pi}^{\pi-\omega} ea_t(e)de\right) \quad \text{(D.13)}$$

$$\lim_{\omega\mapsto 0^-}\frac{E\left[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\right]}{L\times\omega} = 1 - (A_t(\pi) - A_t(-\pi)) + (a_t(\pi) + a_t(-\pi))\,\pi \quad \text{(D.14)}$$

$$\lim_{\omega\mapsto\infty}\frac{E\left[c_{i,t|\omega_{i,t}=\omega} - c_{i,t|\omega_{i,t}=0}\right]}{L\times\omega} = 1 \quad \text{(D.15)}$$

Therefore, the unconditional average propensity to consume is continuous at 0 and tends to 1 at $\pm\infty$. Similarly to what we have done for the conditional response, we now aim to understand whether the limit at 0 is generally smaller or higher than the one at $\pm\infty$. To gauge the size of (D.11), Caballero and Engel (2007) show that in price adjustment models with increasing hazard rates, a good approximation is given by three times the probability to update. However, their approximation does not directly apply to strict Ss-thresholds (they provide another for strict Ss-thresholds and uniform stationary distribution, which also substantially differ from the framework under study here). In the following, we aim to provide a similar rule-of-thumb that applies to our model. Our objective is to illustrate that situations such that the limit at 0 of the unconditional average propensity to consume is smaller than one are rather common in our setup. This is important since it implies that the overall pattern of the unconditional average propensity to consume can be increasing.

Lemma 5 provides a candidate distributions to approximate the stationary cross sectional distribution at $\pm\pi$. Indeed, we expect that

$$a^\star(\pm\pi) \approx \phi(\pm\pi; 0, \sqrt{1+(1+r)^2}\sigma_\omega)$$

where $\phi(x;\mu,\sigma)$ is the pdf of a normally distributed variable with mean μ and standard deviation σ. Given our calibration, $1-(A^\star(\pi)-A^\star(-\pi)) = 0.25$ and $\pi \simeq 1.4\sigma_\omega$. An approximate value for (D.11) at the stationary cross sectional distribution is thus

$$\frac{1}{\sqrt{1+(1+r)^2}\sqrt{2\times 3.14}} \exp\left(-0.5\times\left(\frac{1.4}{\sqrt{1+(1+r)^2}}\right)^2\right)\times 1.4 + 0.25 = 0.73$$

Using our simulations, we find that Equation (D.11) is equal to 0.72, thus confirming our approximation. This value is also very close to Caballero and Engel's rule-of-thumb which gives a value of 0.75 here. These illustrative computations suggest that the overall pattern of the unconditional average propensity to consume out of a positive income shock increases with the size of the shock when the cross section is at its steady state and the steady state proportion of attentive agents sufficiently small.

Shape of the average propensities to consume – The behavior of the conditional and unconditional average propensities to consume is generally not monotonic, making it harder to draw sharp predictions that can be tested in the data. In the following, we therefore report the simulated average propensities to consume for three different scenarios: stationary distribution, small shift of the distribution to the left ($E(e) = -0.5\sigma_\omega$) and a larger shift ($E(e) = -\sigma_\omega$). The results for shifts of the cross sectional distribution to the right follow by symmetry.

Because all shocks are not equally likely to be observed in the data, we then draw 500 shocks ω_t from a Gaussian with mean 0 and standard deviation σ_ω. Fitting a regression line, we can then get predictions about the sign of the slopes that can be confronted to the data. Following our discussion on the limits at 0, we allow for different intercepts only for the conditional average propensity. Based on this exercise, we derive the predictions P1-P4 regarding the behavior of linear/logistic fits reported in the main text.

These predictions are obtained from translating the stationary distribution of consumers. However, the dynamics of the cross sectional distribution may generate more complex distributions. To assess whether these predictions continue to hold for alternative cross sections of consumers, we simulate 500 cross sectional distributions (starting from the ergodic distribution and iterating 150 times on the dynamic equations from Proposition 2) and again draw 500 income shocks. We then test wether P1-P4 hold. Doing so, we reject at least one of the four predictions in 3.4 percent of simulations (type I error). The type II error is 3.0 percent.

Appendix E. Evidence from the BoE survey of consumers

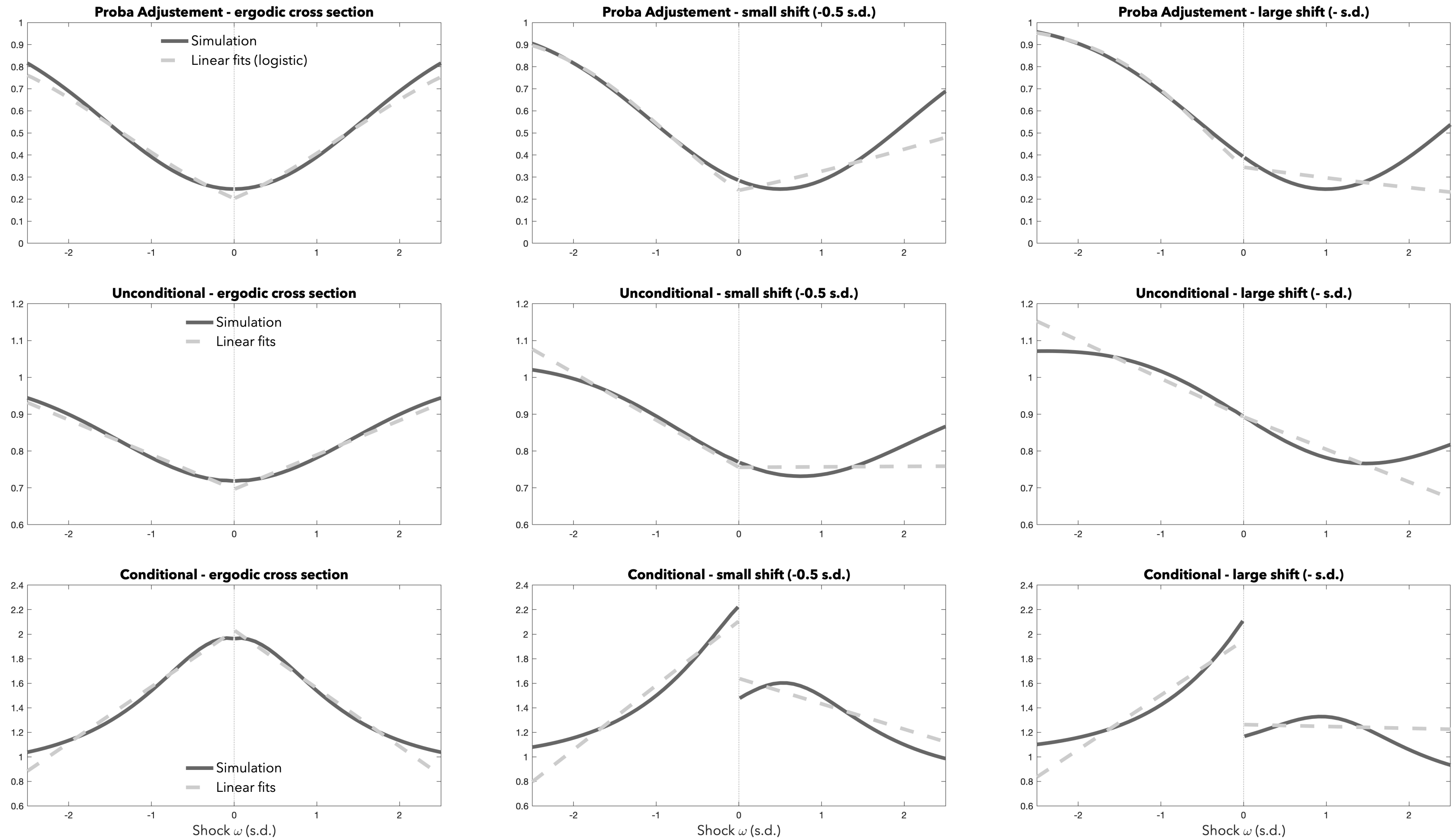
Appendix E.1. Description of the main variables

Anderson et al. (2016) and Bunn et al. (2018) provide an extensive discussion of the survey characteristics and its representativeness. To avoid unnecessary duplication, we refer readers to these papers for more information about the survey characteristics. Moreover, Bunn et al. (2018) verify that the measure of MPC that we use is consistent with an alternative income shock identification strategy based on hypothetical shocks and a more qualitative indicator of the spending response. In the following, we describe the construction of the main variables used in our analysis.

The question about the income surprise is phrased as follows: "*Compared to what you expected this time last year, how much more [less] money did your household receive over the last 12 months? Please consider your income after income tax and National Insurance but before any housing costs or bills are paid. Please include any unexpected pay increases or decreases, bonuses, lottery winnings, unexpected tax bills or repayments, PPI claims and inheritance, lifestyle changes, etc.*" The following question about the persistence of the surprise is then asked "*Are you treating this unexpected increase [decrease] in money received by your household as: A temporary increase [decrease] / An increase [decrease] that is likely to persist.* We exclude income surprises that exceed $\pm£25,000$.

The question about the consumption change is "*You indicated earlier in the survey that*

Figure D.5: Shape of the conditional and unconditional average propensities to consume



Note: Probability to adjust consumption (top panels), unconditional (middle panels) and conditional (bottom panels) average propensity to consume out of positive and negative shocks ω_t as functions of the shock size. The first column considers the stationary cross section, while the second and last columns respectively consider a small $(-0.5\sigma_\omega)$ and a large $(-\sigma_\omega)$ shift of the stationary distributions to the left. The results for shifts to the right then follow by symmetry. The plain lines report the results from numerically computing equations (D.4) and (D.7) in the bottom panels and equations (D.10) and (D.13) in the middle panels. The linear fits are then obtain by generating 500 observations and fitting regression lines. In the bottom panels, we account for the discontinuity at 0 by introducing different intercepts for positive and negative perceived shocks. The average propensities to consume are normalized by the propensity to consume that would be obtained absent information frictions.

your household received £[reported amount] more [less] over the last 12 months than you had expected a year ago. By how much did you increase/decrease your annual spending in response to this?" We compute a respondent's marginal propensity to consume as the ratio of her reported consumption change to her income surprise. Following Bunn et al. (2018), we exclude observations with a negative or too high (> 1.5) propensity to consume. The results do not hinge on these restrictions.

The question about credit constraint is phrased as follows: "*Have you been put off spending because you are concerned that you will not be able to get further credit when you need it, say because you are close to your credit limit or you think your loan application would be turned down? Yes/No.*" Liquid assets are elucidated from the following question "*How much do you (and all other members of your household) currently have in total, saved up in savings accounts? Please include bank/building society accounts or bonds, cash ISAs, NS&I account/bonds, and other investments such as stocks, shares and unit trusts.*" Net liquid assets are net of unsecured debt (excluding credit card balances which the household intends to pay in full over the month). Cash-on-hand is net liquid assets plus quarterly disposable income. Disposable income is elucidated from "*How much of your monthly income would you say your household has left after paying tax, National Insurance, housing costs (e.g. rent, mortgage repayments, council tax), loan repayments (e.g. personal loans, credit cards) and bills (e.g. electricity)?*"

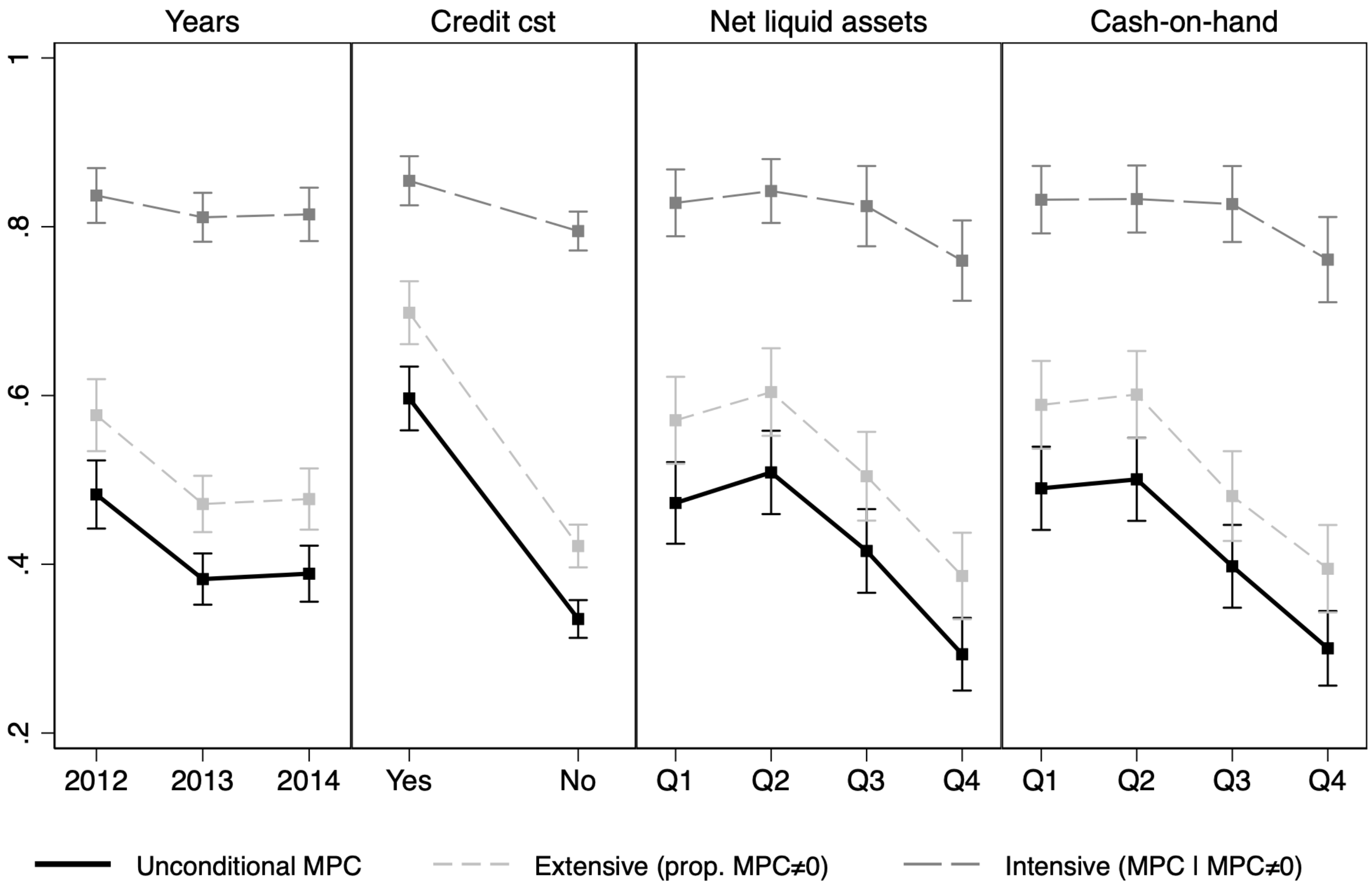
Table E.3 provides summary statistics for the main variables used in the empirical analysis.

Table E.3: Summary statistics

	Mean	Median	s.d.	P25	P75	N
MPC	0.41	0	0.46	0	1	2,101
Income Surprise	-238	-120	5,855	-2,000	2,000	2,101
Net liquid assets	16,675	-250	128,369	-7,000	10,700	1,389
Cash-on-hand	21,138	1,500	130,928	-4,950	16,150	1,370
Monthly disposable income	1,288	550	4,752	250	1,100	2,069
MPC $\neq 0$	50%					2,101
Working	63%					2,101
Temporary surprise	27%					1,966
Credit constrained	29%					2,044
Years (N)	2012	(515)	2013	(859)	2014	(727)

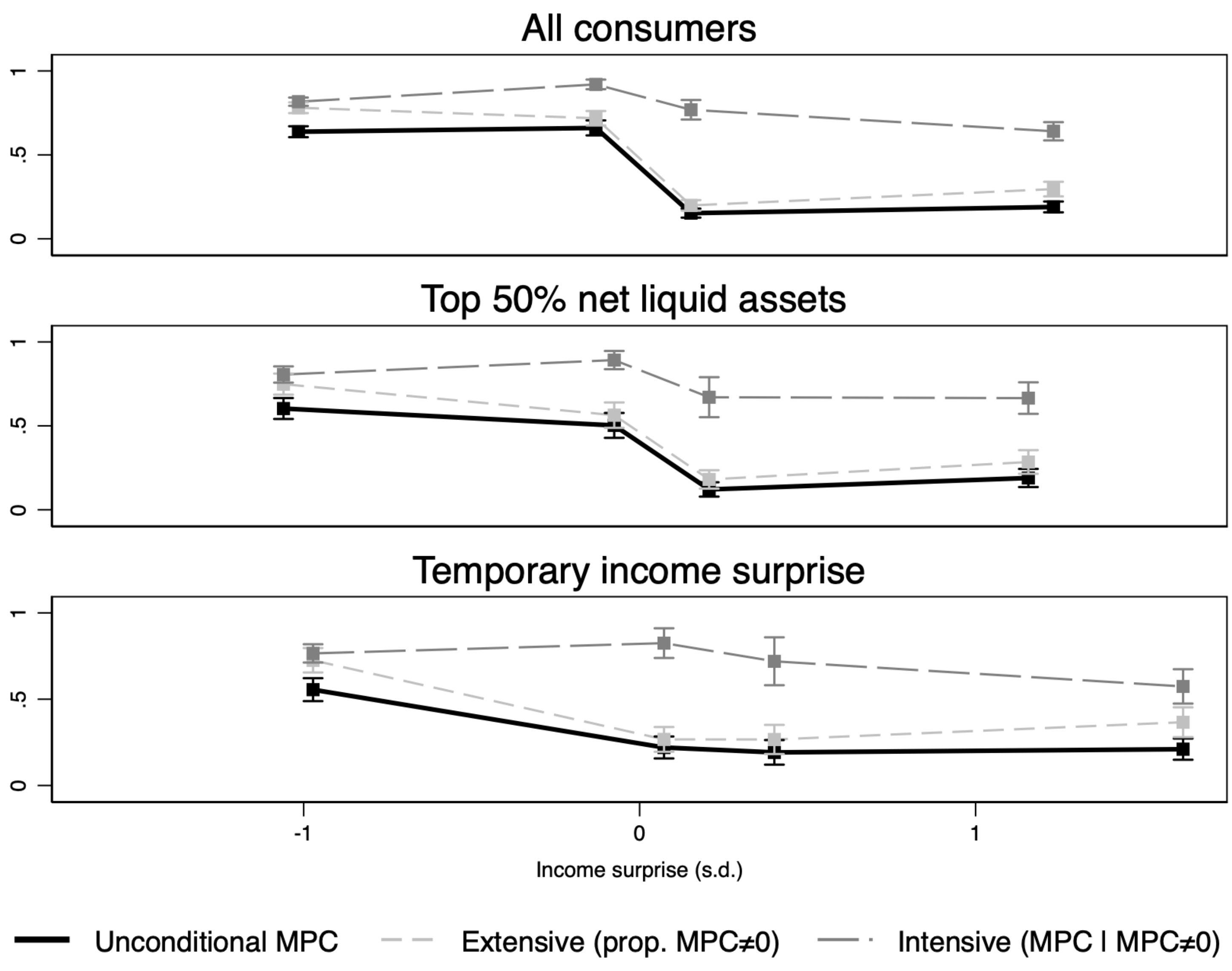
Appendix E.2. Complementary Tables and Figures

Figure E.6: Decomposition of the unconditional MPC



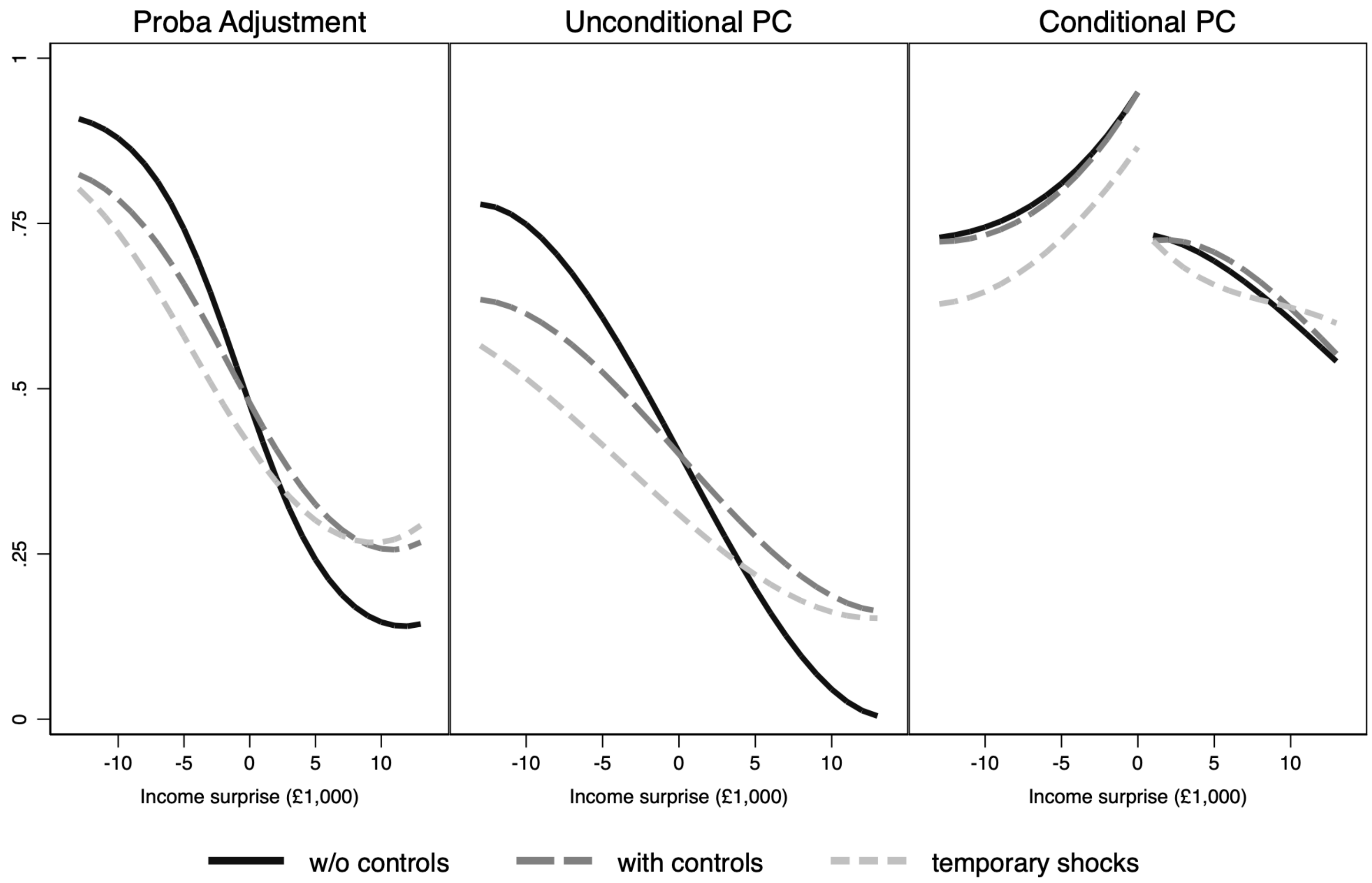
NOTE: Mean MPCs with 95% confidence intervals.

Figure E.7: Unconditional MPC and income surprises



NOTE: Mean MPCs are reported at the income surprise mean within each quartile. 95% confidence intervals.

Figure E.8: Shape of the conditional and unconditional MPC in the BoE survey



NOTE: Probability to adjust consumption (left panel), unconditional (middle panel) and conditional (right panel) average propensity to consume as functions of perceived income shocks ω_t. The results are obtained by fitting the regressions from Table 1 allowing for cubic polynomials.

Table E.4: Asymmetry and size effects: credit constraint, cash-on-hand and temporary surprise

Panel A - Probability to adjust consumption (Logit)

	Not credit constrained		Cash-on-hand (top 50%)		Temporary shocks	
Shock size						
Positive	0.010	(0.020)	-0.033	(0.033)	0.042*	(0.025)
Negative	0.167***	(0.038)	0.168***	(0.051)	0.215***	(0.066)
Constant	-1.092***	(0.316)	-0.862*	(0.480)	-0.807	(0.505)
Controls	Yes		Yes		Yes	
Obs. − *pseudo* R^2	1,222	*0.140*	576	*0.179*	454	*0.128*

Panel B - Linear shape of the MPC (OLS)

	Not credit constrained		Cash-on-hand (top 50%)		Temporary shocks	
MPC	All	$\neq 0$	All	$\neq 0$	All	$\neq 0$
Shock size						
Positive	-0.010***	-0.020***	-0.012***	-0.013*	-0.004	-0.016***
	(0.003)	(0.004)	(0.004)	(0.007)	(0.004)	(0.005)
Negative	0.016***	-0.014***	0.019***	-0.011***	0.021***	-0.007
	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)
Intercepts						
Negative		0.101**		0.128*		0.031
		(0.045)		(0.067)		(0.066)
Constant	0.205***	0.747***	0.218**	0.674***	0.271***	0.742***
	(0.056)	(0.068)	(0.086)	(0.106)	(0.092)	(0.110)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1,222	509	576	247	454	195
R^2	0.176	0.217	0.201	0.137	0.130	0.173

The results are obtained by fitting the regressions from Table 1 on the subsets of individuals who reported not being credit constrained, whose cash-on-hand is above the median and who reported that the income surprise was expected to be temporary. Robust standard errors in parenthesis. *, ** and *** denote significance at the 10%, 5% and 1% levels respectively. Control variables are categories for age, employment status, debt concerns, fear about future income drops, whether the household is credit constrained, the type of income shock (temporary or likely to persist), gross and discretionary income quartiles, and survey wave fixed effects. The base is an individual, aged 35-45, who responded in 2012, who is working, has no concern about her debt, does not fear an income drop, has experienced an unexpected income increase, is not credit constrained, has experienced a temporary income shock, has an annual gross income between £25,000 and £49,999 and a monthly discretionary income between £600 and £1,199.

Online Appendix

Not for publication

This online appendix is composed of three sections. The first section shows that Theorem 1 in Molin and Hirche (2017) – stating that the solution that we have identified for the consumer's problem is globally asymptotically stable – applies. The second section provides the proof for Lemma 5 used in Appendix B. The third section solves the extension with perfect information upon being attentive discussed in Section 3.3.

Appendix F. The solution to the consumer's problem is globally asymptotically stable

In this appendix, we show that Theorem 1 in Molin and Hirche (2017) applies to the problem considered in Appendix A.3. The theorem is meant to derive the optimal design of event-triggered estimation for first-order linear stochastic systems with an identical information structure. The general requirements for the theorem are that the distributions of the initial state e_0 and $\{w_t\}$ are symmetric and unimodal. This is the case in our setup since these distributions are Gaussian. The difference from their problem is with regard to the objective function. They consider the sum of square errors, whereas we are interested in a weighted and discounted sum of these errors here.

In the following, we recast the problem in Appendix A.3 using the notation used in their proof. Let

$$\hat{e}_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0] + \alpha(t, l_t) \tag{F.1}$$

Accordingly, problem (A.13) can be written as

$$\begin{aligned} \min_{g.(.),\alpha(.)} \quad & E\Big[\sum_{t=0}^{T-1} \beta^t \Big((1-\tau_t)\Gamma_t(\hat{e}_t - \alpha(t, l_t))^2 + \lambda\tau_t\Big)\Big|\mathcal{I}_0\Big] \\ \text{s.t.} \quad & \hat{e}_{t+1} = (1-\tau_t)(1+r)\hat{e}_t + \omega_{t+1} \end{aligned} \tag{F.2}$$

where $\Gamma_t \equiv L_t(1 + \beta p_{t+1})$. Note that from (A.12) it is clear that the inattention length is also a sufficient statistic for the corrective $\alpha(t, l_t)$. Moreover, let

$$
\begin{aligned}
\hat{y}_t &\equiv \frac{\hat{e}_t}{R^t}, \quad t = 0, \ldots, N-1 \\
\varrho_{t,l_t} &\equiv \frac{\alpha(t, l_t)}{R^t}, \quad t = 0, \ldots, N-1, l_t = 0, \ldots, t
\end{aligned}
$$

where $R \equiv (1 + r)$. Given this transformation, the running cost is

$$
\hat{c}_t^{\varrho_t}(\hat{y}_t, l_t, \tau_t) = \beta^t \Big((1 - \tau_t) R^{2t} \Gamma_t (\hat{y}_t - \varrho_{t,l_t})^2 + \lambda \tau_t \Big) \tag{F.3}
$$

The optimization problem for Molin and Hirche (2017) iterative procedure is thus given by

$$
\min_{\hat{g}, \varrho} \hat{J} \tag{F.4}
$$

with

$$
\hat{J}(\hat{g}, \varrho) = E_{\hat{g}} \left[\sum_{t=0}^{N-1} \hat{c}_t^{\varrho_t}(\hat{y}_t, l_t, \tau_t) \right] \tag{F.5}
$$

where the subscript $\hat{g}$ emphases that the expectation is taken with respect to the triggering/attention policy. The proof in Molin and Hirche (2017) requires that, for a fixed vector ϱ^i of all ϱ_{t,l_t}, the following symmetry and monotonicity properties hold for the running cost:

$$
\begin{aligned}
\hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i + \Delta, l_t, \tau) &= \hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i - \Delta, l_t, \tau) \\
&\quad \forall \Delta \in \mathbb{R}, l_t \in \{0, \ldots, t-1\}, \tau \in \{0, 1\}
\end{aligned} \tag{F.6}
$$

and

$$
\begin{aligned}
0 \leq \Delta_1 \leq \Delta_2 \Longrightarrow \quad \hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i + \Delta_1, l_t, \tau) &\leq \hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i + \Delta_2, l_t, \tau) \\
&\quad \forall l_t \in \{0, \ldots, t-1\}, \tau \in \{0, 1\}
\end{aligned} \tag{F.7}
$$

It is straightforward to see that these properties continue to hold for our problem given (F.3). Consequently, the subsequent results in the proof in Molin and Hirche (2017) are valid and their Theorem 1 applies.

Appendix G. Approximating the distribution $f_t(e|k, e_{t-k})$

In this appendix, we provide the computations required to derive the approximation in Lemma 5 for the probability distribution functions of the expectation wedge conditional on remaining inattentive for k given an initial wedge $e_{t-k} = a$. The case reported in Lemma 5 corresponds to the case $a = 0$.

Recall that $\sigma_\omega^2 = K^2(\bar{p}_+ + \sigma_\eta^2)$ is the variance of the innovation from the steady state Kalman filter at the attention choice (see Equation A.13). Define $f_t(e|0, a) = \delta(e - a)$. Then from iterating on (B.3), we have for k=1:

$$f_t(e|1, a) \quad = \quad \frac{1}{\sigma_\omega}\phi\left(\frac{e-(1+r)a}{\sigma_\omega}\right) \tag{G.1}$$

When $k = 2$,

$$\begin{aligned} f_t(e|2, a) \quad &\propto \quad \int_{\Xi_{t-1}} \frac{1}{\sigma_\omega^2}\phi(\frac{e-(1+r)\bar{e}}{\sigma_\omega})\phi(\frac{\bar{e}-(1+r)a}{\sigma_\omega})d\bar{e} \qquad (G.2)\\ &\propto \quad \int_{\Xi_{t-1}} \frac{1}{2\pi\sigma_\omega^2}\exp\left\{-\frac{e^2-2(1+r)\bar{e}e+(1+r)^2\bar{e}^2+\bar{e}^2-2(1+r)\bar{e}a+(1+r)^2a^2}{2\sigma_\omega^2}\right\}d\bar{e} \end{aligned}$$

Focusing on the numerator in the exponential and using the shortcut notation $R = 1+r$

$$\begin{aligned} &(1+R^2)\left[\bar{e}^2 - \frac{2R}{1+R^2}\bar{e}(e+a) + \frac{R^2}{1+R^2}a^2 + \frac{1}{1+R^2}e^2\right] \\ = \quad &(1+R^2)\left[\bar{e}^2 - \frac{2R}{1+R^2}\bar{e}(e+a) + \left(\frac{R}{1+R^2}\right)^2(e+a)^2 + \frac{R^2}{1+R^2}a^2 + \frac{1}{1+R^2}e^2 - \left(\frac{R}{1+R^2}\right)^2(e+a)^2\right] \\ = \quad &(1+R^2)\left(\bar{e} - \frac{R}{1+R^2}(e+a)\right)^2 + (1+R^2)\left[\frac{R^2}{1+R^2}a^2 + \frac{1}{1+R^2}e^2 - \left(\frac{R}{1+R^2}\right)^2(e+a)^2\right] \end{aligned}$$

Where

$$\begin{aligned} &\frac{R^2}{1+R^2}a^2 + \frac{1}{1+R^2}e^2 - \left(\frac{R}{1+R^2}\right)^2(e+a)^2 \\ = \quad &\frac{R^2}{1+R^2}a^2 + \frac{1}{1+R^2}e^2 - \left(\frac{R}{1+R^2}\right)^2(e^2+a^2+2ea) \\ = \quad &\frac{1}{(1+R^2)^2}e^2 - 2\left(\frac{R}{1+R^2}\right)^2 ea + \left(\frac{R^2}{1+R^2}\right)^2 a^2 + \left[\frac{R^2}{1+R^2} - \left(\frac{R}{1+R^2}\right)^2 - \left(\frac{R^2}{1+R^2}\right)^2\right]a^2 \\ = \quad &\left(\frac{1}{1+R^2}e - \frac{R^2}{1+R^2}a\right)^2 \end{aligned}$$

Therefore, (G.2) writes

$$\begin{aligned}
f_t(e|2,a) &\propto \int_{\Xi_{t-1}} \frac{1}{2\pi\sigma_\omega^2} \exp\left\{-\frac{(1+R^2)}{2\sigma_\omega^2}\left[\left(\bar{e}-\frac{R}{1+R^2}(e+a)\right)^2+\left(\frac{1}{1+R^2}e-\frac{R^2}{1+R^2}a\right)^2\right]\right\}d\bar{e} \\
&\propto \int_{\Xi_{t-1}} \frac{\sqrt{1+R^2}}{\sqrt{2\pi}\sigma_\omega}\exp\left\{-\frac{(\bar{e}-\frac{R}{1+R^2}(e+a))^2}{2\frac{\sigma_\omega^2}{1+R^2}}\right\}\frac{1}{\sqrt{2\pi}\sqrt{1+R^2}\sigma_\omega}\exp\left\{-\frac{(e-R^2a)^2}{2(1+R^2)\sigma_\omega^2}\right\}d\bar{e} \\
&\propto \int_{\Xi_{t-1}} \frac{\sqrt{1+R^2}}{\sigma_\omega}\phi\left(\frac{\bar{e}-\frac{R}{1+R^2}(e+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\right)\frac{1}{\sqrt{1+R^2}\sigma_\omega}\phi\left(\frac{e-R^2a}{\sqrt{1+(1+r)^2}\sigma_\omega}\right)d\bar{e} \\
&\propto \frac{1}{\sqrt{1+R^2}\sigma_\omega}\left[\Phi\left(\frac{\pi_{t-1}-\frac{R}{1+R^2}(e+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\right)-\Phi\left(-\frac{\pi_{t-1}+\frac{R}{1+R^2}(e+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\right)\right]\phi\left(\frac{e-R^2a}{\sqrt{1+R^2}\sigma_\omega}\right) \quad (G.3)
\end{aligned}$$

When $k=3$,

$$\begin{aligned}
f_t(e|3,a) &\propto \int_{\Xi_{t-1}} \frac{1}{\sigma_\omega}\phi\left(\frac{e-R\bar{e}}{\sigma_\omega}\right)f_{t-1}(\bar{e}|2,a)d\bar{e} \\
&\propto \int_{\Xi_{t-1}} \frac{1}{\sqrt{1+R^2}\sigma_\omega^2}\phi\left(\frac{e-R\bar{e}}{\sigma_\omega}\right)\phi\left(\frac{\bar{e}-R^2a}{\sqrt{1+R^2}\sigma_\omega}\right)\left[\Phi\left(\frac{\pi_{t-2}-\frac{R}{1+R^2}(\bar{e}+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\right)-\Phi\left(-\frac{\pi_{t-2}+\frac{R}{1+R^2}(\bar{e}+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\right)\right]d
\end{aligned}$$

Again, I develop and reduce the product of the two gaussian pdfs. To do so, I first focus on the numerator within the exponential.

$$\begin{aligned}
&\frac{1+R^2}{1+R^2+R^4}\left[\left(e-R\bar{e}\right)^2+\left(\frac{\bar{e}-R^2a}{\sqrt{1+R^2}}\right)^2\right] \\
=\ & \bar{e}^2-2\frac{(1+R^2)R}{1+R^2+R^4}\bar{e}e+\frac{1+R^2}{1+R^2+R^4}e^2-2\frac{R^2}{1+R^2+R^4}a\bar{e}+\frac{R^4}{1+R^2+R^4}a^2 \\
=\ & \left(\bar{e}-\frac{(1+R^2)Re+R^2a}{1+R^2+R^4}\right)^2+\frac{1+R^2}{1+R^2+R^4}e^2+\frac{R^4}{1+R^2+R^4}a^2-\left(\frac{(1+R^2)Re+R^2a}{1+R^2+R^4}\right)^2
\end{aligned}$$

Where

$$\begin{aligned}
&\frac{1+R^2}{1+R^2+R^4}e^2-\left(\frac{(1+R^2)Re+R^2a}{1+R^2+R^4}\right)^2 \\
=\ & \frac{(1+R^2)(1+R^2+R^4)-(1+R)^2R^2}{(1+R^2+R^4)^2}e^2-2\frac{(1+R^2)R^3}{(1+R^2+R^4)^2}ea-\left(\frac{R^2}{1+R^2+R^4}\right)^2a^2 \\
=\ & \left(\frac{\sqrt{1+R^2}}{1+R^2+R^4}\right)^2e^2-2\frac{(1+R^2)R^3}{(1+R^2+R^4)^2}ea+\left(\frac{\sqrt{1+R^2}R^3}{1+R^2+R^4}\right)^2a^2-\left(\frac{R^2}{1+R^2+R^4}\right)^2a^2-\left(\frac{\sqrt{1+R^2}R^3}{1+R^2+R^4}\right)^2a^2 \\
=\ & \frac{1+R^2}{(1+R^2+R^4)^2}(e^2-R^3a)^2-\frac{R^4(1+R^2+R^4)}{(1+R^2+R^4)^2}a^2
\end{aligned}$$

so that

$$\left(e-R\bar{e}\right)^2+\left(\frac{\bar{e}-R^2a}{\sqrt{1+R^2}}\right)^2 = \frac{1+R^2+R^4}{1+R^2}\left(\bar{e}-\frac{(1+R^2)Re+R^2a}{1+R^2+R^4}\right)^2+\frac{(e^2-R^3a)^2}{1+R^2+R^4}$$

Introducing back this expression in (G.4), I obtain

$$f_t(e|3,a) \propto \frac{1}{\sqrt{1+R^2\sigma_\omega^2}}\phi\Big(\frac{e-R^3a}{\sqrt{1+R^2+R^4\sigma_\omega}}\Big)\times \int_{\Xi_{t-1}}\phi\Big(\frac{\bar{e}-\frac{R(1+R^2)e+R^2a}{1+R^2+R^4}}{\sqrt{\frac{1+R^2}{1+R^2+R^4}}\sigma_\omega}\Big)\Big[\Phi\Big(\frac{\pi_{t-2}-\frac{R}{1+R^2}(\bar{e}+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\Big)-\Phi\Big(-\frac{\pi_{t-2}+\frac{R}{1+R^2}(\bar{e}+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\Big)\Big]d\bar{e} \tag{G.4}$$

The above expression may not be expressed in other terms to simplify the computation for $k=4$. As a consequence, the computational cost from conditioning on past histories grows exponentially and will likely generate large approximation error when k increases. Therefore, I approximate the above expression by not accounting for the impact of histories before $t-1$. The approximated distribution is thus

$$\begin{aligned} f_t^{\text{app}}(e|3) &\propto \frac{1}{\sqrt{1+R^2\sigma_\omega^2}}\phi\Big(\frac{e-R^3a}{\sqrt{1+R^2+R^4\sigma_\omega}}\Big)\int_{\Xi_{t-1}}\phi\Big(\frac{\bar{e}-\frac{R(1+R^2)e+R^2a}{1+R^2+R^4}}{\sqrt{\frac{1+R^2}{1+R^2+R^4}}\sigma_\omega}\Big)d\bar{e} \\ &\propto \frac{1}{\sqrt{1+R^2+R^4\sigma_\omega}}\phi\Big(\frac{e-R^3a}{\sqrt{1+R^2+R^4\sigma_\omega}}\Big)\Big[\Phi\Big(\frac{\pi_{t-1}-\frac{R(1+R^2)e+R^2a}{1+R^2+R^4}}{\sqrt{\frac{1+R^2}{1+R^2+R^4}}\sigma_\omega}\Big)-\Phi\Big(-\frac{\pi_{t-1}+\frac{R(1+R^2)e+R^2a}{1+R^2+R^4}}{\sqrt{\frac{1+R^2}{1+R^2+R^4}}\sigma_\omega}\Big)\Big] \end{aligned}$$

For $k=4$,

$$\begin{aligned} f_t^{\text{app}}(e|4) &\propto \int_{\Xi_{t-1}}\frac{1}{\sqrt{1+R^2+R^4\sigma_\omega^2}}\phi\Big(\frac{e-R\bar{e}}{\sigma_\omega}\Big)\phi\Big(\frac{\bar{e}-R^3a}{\sqrt{1+R^2+R^4\sigma_\omega}}\Big)d\bar{e} \\ &\propto \frac{1}{\sqrt{1+R^2+R^4\sigma_\omega^2}}\phi\Big(\frac{e-R^4a}{\sqrt{1+R^2+R^4+R^6\sigma_\omega}}\Big)\int_{\Xi_{t-1}}\phi\Big(\frac{\bar{e}-\frac{R(1+R^2+R^4)e+R^3a}{1+R^2+R^4+R^6}}{\sqrt{\frac{1+R^2+R^4}{1+R^2+R^4+R^6}}\sigma_\omega}\Big)d\bar{e} \\ &\propto \frac{1}{\sqrt{1+R^2+R^4+R^6\sigma_\omega}}\phi\Big(\frac{e-R^4a}{\sqrt{1+R^2+R^4+R^6\sigma_\omega}}\Big) \\ &\qquad \Big[\Phi\Big(\frac{\pi_{t-1}-\frac{R(1+R^2+R^4)e+R^3a}{1+R^2+R^4+R^6}}{\sqrt{\frac{1+R^2+R^4}{1+R^2+R^4+R^6}}\sigma_\omega}\Big)-\Phi\Big(-\frac{\pi_{t-1}+\frac{R(1+R^2+R^4)e+R^3a}{1+R^2+R^4+R^6}}{\sqrt{\frac{1+R^2+R^4}{1+R^2+R^4+R^6}}\sigma_\omega}\Big)\Big] \end{aligned}$$

Using forward iteration, it holds

$$f_t^{\text{app}}(e|k) \propto \frac{1}{\sqrt{z(k)}\sigma_\omega}\phi\Big(\frac{e-R^ka}{\sqrt{z(k)}\sigma_\omega}\Big)\Big[\Phi\Big(\frac{\pi_{t-1}-\frac{Ru(k)e+R^{k-1}a}{z(k)}}{\sqrt{\frac{u(k)}{z(k)}}\sigma_\omega}\Big)-\Phi\Big(-\frac{\pi_{t-1}+\frac{Ru(k)e+R^{k-1}a}{z(k)}}{\sqrt{\frac{u(k)}{z(k)}}\sigma_\omega}\Big)\Big] \quad \forall k\in\{3,\dots,T-t\}$$

where $z(k)=\sum_{i=0}^{k-1}(1+r)^{2i}$ and $u(k)=\sum_{i=0}^{k-2}(1+r)^{2i}$.

A comparison from this approximation method to Monte Carlo simulations shows that the main drawback of this procedure is to potentially over estimate the hazard rates for large k by an order of magnitude of about one to two percentage points. The impact on the survival function and distribution of attention length is however negligible as the proportion of agents who encounters a large k is small. We, therefore, conclude that this approximation is well-suited for the calibration used in the paper since it leads to an average inattention length of 4 periods. More importantly, this approximation is only used for the computation of the hazard rates in Figure 1, to report an average inattention length, and as a first guess when iterating for the numerical computation of the stationary distribution $a^\star(e)$.

Appendix H. Extension with perfect information when attentive

This appendix presents and solves the extension with perfect information upon being attentive used to illustrate the potential distribution of expectation revisions in Section 3.3.

Appendix H.1. Problem statement and solution

Problem extension: Consider the consumer problem given in (3) but now assume that the consumer can perfectly observe the true permanent income s_t when attentive. The information structure at the attention choice is unaffected.

The consumption policy is unaffected. The nestedness property of the information structure is unaffected as the consumption choices are observable at the attention choice so that the true state s_t can be retrieved from the latent information set at the beginning of period $t+1$ (i.e. before making a new attention choice). Equation (A.3) is also unaffected. The rest of the demonstration of Lemma 1 in Appendix A.1 then continues to hold. Thus, the optimal consumption policy is still the certainty equivalent one. In the following, we consider the infinite horizon limit and use the fact that the consumption policy $f_t(.)$ converges to a stationary solution (Proposition 1).

Kalman filter at the attention choice. Applying the tower property of conditional expectations, the estimator at the attention choice must minimize a discounted variance (as before). We can, however, not focus on situations where the the Kalman filter is always at its steady state variance as the perfect observation of s_t systematically disturbs the system. This would require to account for the potential distortion implied by the discounting term for the optimal estimation problem. For simplicity, and to remain as close as possible to the solution of the main problem, we consider that the expectation at the attention choice remains the linear least-squares estimator, that is, the Kalman filter.

Given an initial posterior variance p^-_{t-1}, the prior variance at time t is $p^+_t = (1+r)^2 p^-_{t-1} + \sigma^2_\zeta$, the Kalman gain $K(p^+_t) = p^+_t(\sigma^2_\vartheta + p^+_t)^{-1}$. The posterior variance at the end of period t however depends on whether or not the consumer was attentive. If she wasn't, then it follows from the standard Kalman dynamics and we have $p_t^- = (1 - K(p_t^+))p_t^+$. On the other end, she perfectly observed the state s_t if she was attentive. Because at time $t+1$ the state s_t is measurable from the information at the attention choice $\mathcal{I}_{t+1}$ (because c_t is observable), then the uncertainty conveyed through the posterior variance vanishes. Consequently, the dynamics of the prior variance follows from

$$p_{t+1}^+ = (1 - \tau_t)(1+r)^2\Big((1 - p_t^+(\sigma_\vartheta^2 + p_t^+)^{-1}))p_t^+\Big) + \sigma_\zeta^2 \tag{H.1}$$

Note that the inattention length $t - l_t$ is a sufficient statistic for the evolution of this state variable. Moreover, the expression for the value function in equation (A.11) is unaffected.

Consumer's expectation. We now characterize the optimal attention strategy and the consumer's expectations. When attentive the consumer fully observes permanent income. Thus, we have $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1] = s_t$.

When inattentive the consumer's expectation is given by

$$\underbrace{E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0]}_{\text{estimate when inattentive}} = \underbrace{(1+r)E[s_{t-1}|\overline{\mathcal{I}}_{t-1}] - u_{t-1}}_{\text{update}} + \underbrace{E\Big[(1+r)e_{t-1} + K(p_t^+)(z_t - E[s_t|\mathcal{I}_{t-1}])\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big]}_{\text{corrective term accounting for inattention } (\equiv \alpha(t,.))}$$

and the dynamics of the expectation wedge are given by $e_t = (1 - \tau_{t-1})(1+r)e_{t-1} + K(p_t^+)(z_t - E[s_t|\mathcal{I}_{t-1}])$. It therefore follows an AR(1) process with a reseting at 0 when the consumer is attentive. Therefore, the only difference with respect to the main model is that the innovations have a time-varying variance due to the time-varying Kalman gain.

Consequently, the corrective terms write

$$\alpha(t,.) = E\Big[K(p_t^+)\sum_{k=l_t+1}^{t}(1+r)^{t-k}(z_k - E[s_k|\mathcal{I}_{k-1}])\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big] - \sum_{k=l_t+1}^{t-1}(1+r)^{t-1-k}\alpha(k,.) \tag{H.2}$$

Importantly, we remark that l_t is a sufficient statistic for p_t^+ (see the discussion on the Kalman filter at the attention choice). Hence, the corrective terms may again be expressed as a function of time t and the inattention length l_t.

The attention problem is slightly affected in comparison to the main model. The reason is that now being attentive also lead to a persistent reduction in the posterior error variance at the attention choice. Using equation (A.12) and realizing that p_t^+ is a state variable that enters the attention problem, this problem writes

$$\begin{aligned}
\min_{\{\tau_t,\alpha(t,l_t)\}_{(0\leq t\leq T-1, 0\leq l_t<t)}} \quad & E\Big[\sum_{t=0}^{T-1}\beta^t\lambda\tau_t + (1-\tau_t)\beta^t\frac{(\beta(1+r)-1)^2}{\beta(1+r)}(p_t^- + e_t^2)\Big|\mathcal{I}_0\Big] \quad \text{(H.3)}\\
\text{s.t.} \quad e_{t+1} &= (1-\tau_t)(1+r)e_t - \alpha(t+1, l_{t+1}) + K(p_t^+)(z_{t+1} - E[s_{t+1}|\mathcal{I}_t])\\
p_{t+1}^+ &= (1-\tau_t)(1+r)^2\Big((1 - p_t^+(\sigma_\vartheta^2 + p_t^+)^{-1}))p_t^+\Big) + \sigma_\zeta^2\\
l_{t+1} &= \tau_t t + (1-\tau_t)l_t
\end{aligned}$$

This is again a standard dynamic problem with perfect state observation at the attention choice.

Optimal attention policy. We again assume that the corrective terms are zero. Furthermore, we again see that if the attention policy takes the form a of symmetric inattention region with respect to e_t (for a given inattention length), then equation (H.2) again implies that the corrective terms are indeed zero.

A notable feature of the state p_{t+1}^+ dynamics is that it converges to a steady state posterior error variance as the inattention length $t - l_t$ increases. Because this state is the only term in the objective function that depends on the inattention length, it implies that the attention policy also converges as the inattention length increases. Also recalling that the inattention length $t - l_t$ is a sufficient statistic for p_{t+1}^+, the attention policy may only depend on two inputs: the expectation wedge e_t and the inattention length.

Consequently, denoting by d_t the inattention length at time t and $p^+(d_t)$ the posterior error variance as a function of d, the attention problem writes in its Bellman form

$$\begin{aligned}
J(e_t, d_t) = \min_{\tau_t\in\{0,1\}} \quad & (1-\tau_t)\frac{(\beta(1+r)-1)^2}{\beta(1+r)}\Big(p^-(d_t) + e_t^2\Big) + \tau_t\lambda + \beta E[J(e_{t+1}, d_{t+1})|\mathcal{I}_t] \quad \text{(H.4)}\\
\text{s.t.} \quad & e_{t+1} = (1-\tau_t)(1+r)e_t + \sigma(d_{t+1})\omega_{t+1}\\
& d_{t+1} = (1-\tau_t)d_t + 1
\end{aligned}$$

where $\omega_{t+1} \sim \mathcal{N}(0,1)$ and $\sigma(d_{t+1})^2 = K(p^+(d_{t+1}))(\sigma_\vartheta^2 + p^+(d_{t+1}))$.

This problem is solved numerically using a value iteration algorithm (see below for a discussion). Nevertheless, two characteristics of the solution that are already apparent are worth being emphasized. First, the dependence on the inattention length d_t vanishes as the inattention length increases. This is because when the consumer remains inattentive for a long time, then the latent Kalman filter converges to its steady state. Therefore, for large d_t, the behavior of the solution resembles the one from the main problem without perfect state observation upon being attentive. We can use this observation to show that the solution again takes the form of a symmetric inattention region with respect to the expectation wedges e_t and thresholds $\pm\pi(d_t)$. These thresholds converge to some fixed value as d_t increases.

Second, when the information at the attention choice is low enough (i.e. the precision of the signals small enough), then it can be optimal to be attentive after a given (finite) inattention length. Such maximum inattention length exists whenever $p^{-}(\infty) > \Big[\lambda + \beta E[J(e_{t+1}, 1)|\mathcal{I}_t]\Big]\frac{\beta(1+r)}{(\beta(1+r)-1)^2}$ where $p^{-}(\infty)$ denotes the steady state posterior variance of the latent Kalman filter.

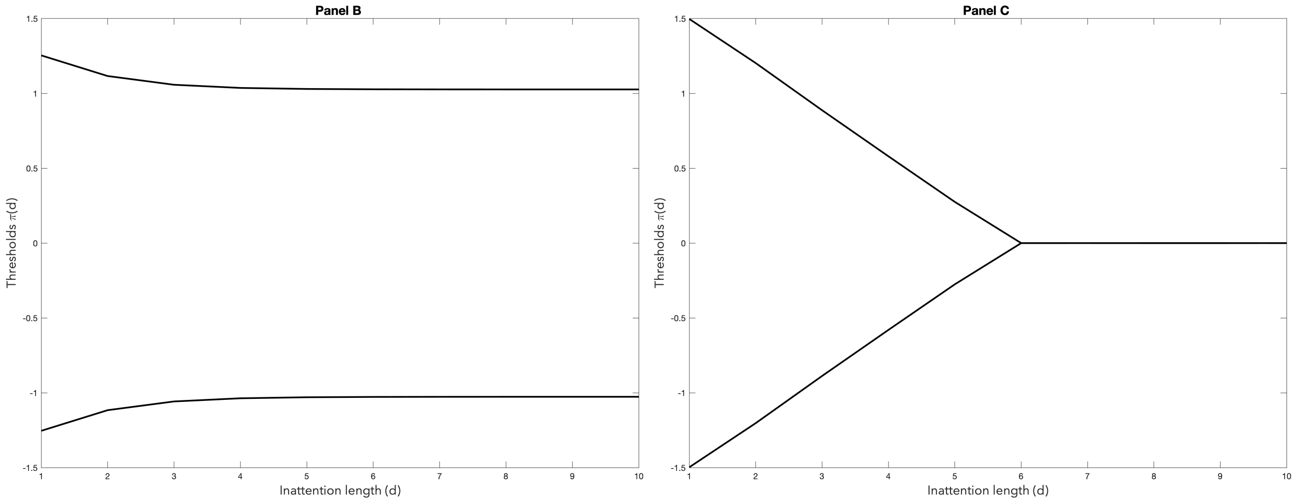
Appendix H.2. Numerical simulations

Value function iteration. The solution is computed as follows. Gaussian integrations rely on Gauss-Hermite quadrature and we use a fixed grid for e_t with 700 equally spaced points. We first guess a maximum inattention length d^0_{max} and compute the predetermined variables related to the latent Kalman filter. We then make a guess on the value function $J^{0,0}(e_t, d_t)$ and deduce the expected value function of being attentive $Ju^{0,0} = \lambda + \beta E[J^{0,0}(e_{t+1}, 1)|\mathcal{I}_t]$. We then iterate backward from d^0_{max} to $d = 1$ to find the optimal thresholds $\pi(d)$ and derive $J^{1,0}(e_t, d_t)$ and $Ju^{1,0}$. We iterate until $||J^{n+1,0}(e_t, d_t) - J^{n,0}(e_t, d_t)||_\infty$ is smaller than a given tolerance. We then increase $d^1_{max} = d^0_{max} + 10$ and restart the iteration over $J^{0,1}(e_t, d_t)$ using $J^{n+1,0}(e_t, d_t)$ as a first guess. We iterate on d^j_{max} until $||Ju^{n+1,j+1} - Ju^{n'+1,j}||_\infty$ is smaller than a given tolerance.

Inattention regions. Figure H.9 reports the inattention regions associated to the distributions of revisions reported in Panels B and C of Figure 2 in the main text. We use the benchmark calibration for the income process and consumption parameters. In both cases, the cost λ is set to match an average inattention length equal to 4 periods. In Panel B the precision of the signals is the same as in the benchmark calibration. In Panel C, the precision is much smaller.

Distributions of revisions. The distributions of revisions reported in Panels B and C of Figure 2 in the main text are then computed from Monte Carlo simulations. We simulate 400 times the dynamics of e_t during 1,500 periods for the optimal $\pi(d)$ found before. These distributions are thus computed on 150,000 points (on average). We rely on kernel distribution estimations (Gaussian for the revisions, Epanechnikov for the e_t).

Figure H.9: Inattention regions for the extension with full-information when attentive



NOTE: Inattention regions (between the two lines) for two calibrations of the extended model with full-information upon being attentive. The calibrated average inattention length and consumption problem parameters are the same across panels. Values are normalized by the standard deviation of the innovation to permanent income. Panel B – predicted inattention region for the distribution of revisions reported in Panel B of Figure 2 in the main text. Panel C – Same extension than Panel B but with a decreased information at the attention choice.

Appendix I. Parameters sensitivity and welfare cost of inattention

As is explained in the main text, consumers are essentially attentive to idiosyncratic shocks on a quarterly basis. Therefore, we only report results for inattention to aggregate innovations in the following.

Appendix I.1. Optimal attention: sensitivity to the model parameters

Table I.5 reports the threshold π normalized by the permanent income standard deviation. At the benchmark calibration, households update whenever their expectation wedge e_t is larger than 1.40 σ_ζ.

Table I.5: Optimal inattentiveness

	Benchmark	Impact of a 5% decrease					
		r	β	ρ	σ_ϵ	λ	σ_ϑ
$\bar{\pi}$	1.40	1.52	0.68	1.47	1.46	1.38	1.41
$\bar{d}$	4.00	4.41	1.92	4.21	4.18	3.91	4.01

NOTE: Optimal normalized threshold $\bar{\pi} = \pi/\sigma_\zeta$ and implied average duration between updates $\bar{d}$ in quarters. The first column is for the benchmark calibration. Subsequent columns evaluate the impact of decreasing one of the parameters by 5% while keeping others constant.

In order to assess how optimal inattentiveness is affected by the model parameters, Table I.5 displays the change in the normalized threshold and average inattention length when one parameter decreases by 5%, keeping the others at the benchmark calibration. Table I.5 thus provides information on the attention threshold and average inattention length elasticities. The attention threshold and average length are decreasing in the persistence of innovations ρ and their standard deviation σ_ϵ. These results are not surprising as an increase in any of these parameters ultimately rises the ex ante standard deviation of permanent income, thus making the consumer willing to be relatively more attentive to changes in permanent income. The effect of the interest rate is unclear a priori. On the one hand, an increase in r decreases the standard deviation of permanent income, thus reducing the ex ante uncertainty faced by the consumer. On the other hand, an increase in r rises the cost of being inattentive today. Simulations indicate that this second effect dominates. The updating threshold and average duration increase with the discount rate β. This is because an individual smoothes consumption more when she values tomorrow more. Therefore, when β increases, the share of permanent income that she consumes (i.e. the variable L in Lemma 1) decreases and so does the instantaneous cost of misoptimization. Consequently, she is willing to wait more between two updates on average. Trivially, the threshold and average inattention length increase when the attention cost λ increases. Finally, the attention behavior is relatively unaffected by the signal noisiness σ_ϑ. This is because, as we saw, the consumer relies on a Kalman filter to smoothly incorporate new information from noisy signals. Consequently, when these signals are noisier, the consumer's optimal strategy is essentially to adjust her estimator with respect to $\mathcal{I}_t$ from Lemma 2. Hence, she compensates for the decreased precision of the signals by optimally adjusting her Kalman gain and anchors even more her latent estimate of s_t on her latent prior beliefs. Since we consider a relatively small change in the signal noisiness in Table I.5, the optimal adjustment in the Kalman gain (almost fully) offsets the increased uncertainty due to the information loss. The latent posterior is barely affected, and so is the optimal attention strategy of the consumer.[42]

Appendix I.2. Welfare decomposition

Taking the infinite horizon version of equation (A.11), the value function at period 0 writes

$$V_0 = E\Big[p\bar{s}_0^2 + \frac{p}{1-\beta}\sigma_\zeta^2 + \frac{L^2(1+\beta p)}{1-\beta}p_- + \sum_{t=0}^{\infty}\beta^t\Big(L^2(1+\beta p)e_t^2 + \lambda\tau_t\Big)\Big|\mathcal{I}_0\Big] \quad \text{(I.1)}$$

[42]The relative independence of the attention strategy with respect to the signal precision holds locally given our calibration. This property is not global.

The above expression provides a straightforward decomposition of the welfare costs from imperfect information. The first term stands for the expected value function of a consumer facing a deterministic linear quadratic control problem. Taking the first two terms leads to the value function of a consumer facing a stochastic control problem with perfect state observation – that is the standard permanent income model with full-information rational expectation. The third term measures the welfare cost from the noisy state observation. Finally, the remaining sum stands for the cost of the information-dependent attention strategy.

Equation (I.1) is conditional on $\mathcal{I}_0$ and therefore imposes that period 0 is an updating period. To avoid such restriction and consider an initial period that does not rely on the specifics of the consumer behavior, I instead compute the expected value function unconditionally on the updating behavior at period 0. Let $E[V_0(e_0)]$ be this expected value function. Furthermore, assume that the consumer has initially already lived for a long time. Accordingly, the pdf associated to e_0 is given by the cross-sectional stationary distribution $a^\star(.)$ from Proposition **??**. Now, realize that $e_t = e_0$ if $e_0 \in \Xi$ and zero otherwise. Consequently, the relevant distribution for e_t is the transformation of $a^\star(.)$ which accounts for the resetting at zero when e is outside the boundaries. Given that $\int ea^\star(e)de = 0$ and denoting $\sigma_a^2 = \int_\Xi e^2 a^\star(e)de$, we find that $E[e_t^2] = \sigma_a^2$ is time invariante. Furthermore, let $\bar{\lambda}^\star \equiv \int_{\notin \Xi} a^\star(e)de$ be the share of updates at the stationary distribution. Then, $\lambda \sum_{t=1}^{\infty} \beta^t E_{a^\star(.)}[\tau_t] = \frac{\lambda\bar{\lambda}^\star}{1-\beta}$. Therefore,

$$E_{a^\star(.)}[V_0(e_0)] \quad = \quad p(\bar{s}_0^2 + p_-) + \frac{p}{1-\beta}\sigma_\zeta^2 + \frac{L^2(1+\beta p)}{1-\beta}(p_- + \sigma_a^2) + \frac{\lambda\bar{\lambda}^\star}{1-\beta} \qquad \text{(I.2)}$$

Following Cochrane et al. (1989), I use a money metric to measure the welfare cost of deviating from the full information rational expectation solution. Dividing the expected welfare loss marginal utility of consumption and converting it to quarterly rates, we get a welfare cost converted in dollars per period:

$$\text{WC} \quad = \quad \frac{r(1-\beta)^{-1}}{2(\bar{c}-\bar{y})(1+r)}\Big[\frac{[\beta(1+r)^2-1]^2}{\beta(1+r)}(p_- + \sigma_a^2) + \lambda\bar{\lambda}^\star\Big] \qquad \text{(I.3)}$$

The overall welfare cost from costly information processing may therefore be apprehended as the sum of three terms: the utility cost from paying λ at each update (updating cost), the misoptimization cost from being inattentive to signals (latent information cost), and the misoptimization cost from observing noisy signals instead of perfect information (noisy information cost).

Table I.6: Welfare cost of inattention to aggregate income shocks

Coeff. of relative risk aversion	1	2	4	10
Bliss point $\bar{c}$	\$13,858	\$10,393	\$8,661	\$7,622
Welfare decomposition (¢/quarter)	1.18	2.37	4.74	11.84
Updating cost	*0.34*	*0.68*	*1.37*	*3.42*
Latent information cost	*0.63*	*1.27*	*2.54*	*6.34*
Noisy information cost	*0.21*	*0.41*	*0.83*	*2.07*
Consumption change to update	0.170%	0.169%	0.169%	0.168%

NOTE: The welfare cost refers to misoptimization cost induced by costly information processing. The coefficient of relative risk aversion is equal to $\bar{y}/(\bar{c}-\bar{y})$. The consumption change to update measures the threshold change in perceived consumption that will prompt the consumer to internalize new information at period 0. These results were obtained under the benchmark calibration for the infinite horizon limit.

Table I.6 reports these welfare costs for different values of the coefficient of relative risk aversion (CRRA) under the benchmark calibration. The costs induced by costly information processing are small, even when one considers extremely risk averse consumers. Decomposing the welfare cost, we find that more than 80% of it is attributable to the extensive margin of expectation adjustments. When the CRRA is equal to one (resp. 10), the monetary equivalent from paying λ when being attentive is ¢0.34 (resp. 3.42) per quarter on average. Given that a consumer updates once a year on average, the monetary equivalent for λ is ¢1.36 (resp. ¢13.68).

Appendix J. AR(2) for the persistence of consumption

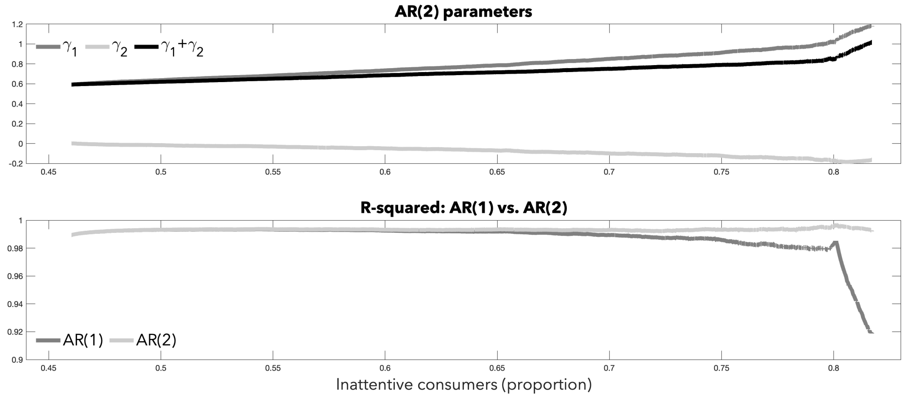
In this section, we show that the dynamics of aggregate consumption from the simulated data in Section 6.1 is better approximated by an AR(2) process when the share of inattentive consumers is large. Overall, accounting for this more complex dynamics allows to extend our conclusion that consumption persistence increases with the share of inattentive consumers to situations where more than 80% of consumers are inattentive.

We amend the Equation (17) to include a further lag

$$\Delta C_{t+1} = \alpha_{t+1} + \gamma_{1,t+1}\Delta C_t + \gamma_{2,t+1}\Delta C_{t-1} + \beta_{t+1}\bar{\zeta}_{t+1} + error_{t+1} \tag{J.1}$$

Looking at the R-squared from the AR(1) and AR(2) specifications, it is clear that the AR(2) performs equivalently for relatively small shares of inattention, while it outperforms the AR(1) specification when Π_t increases. The reason is clear from the top panel in Figure J.10: as Π_t increases, the second lag in consumption growth matters more. A standard

Figure J.10: Aggregate consumption persistence and inattention (extension with an AR(2))



NOTE: Rolling regression of Equation (J.1) based on 84,000 simulated data points and a window of 4,000 observations.

measure of persistence for AR(p) processes is the sum of the autoregressive coefficients.[43] As can be seen from the figure, this measure increases monotonically with inattention.

[43]See the discussion in Dias and Marques (2010) for alternative measures.

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it