*École des Hautes Études en Sciences Sociales (EHESS)*
*École doctorale Économie Panthéon-Sorbonne*
*Paris-Jourdan Sciences Économiques*
*École d'Économie de Paris (PSE)*

THÈSE DE DOCTORAT

Pour l'obtention du grade de docteur en Sciences Économiques de l'École des Hautes Études en Sciences Sociales.
*Discipline : Sciences Économiques - Spécialité : Analyse et Politique Économiques*

Jérémy Boccanfuso

# Trois essais sur les implications de l'attention limitée en économie

Thèse dirigée par Xavier Ragot.
Présentée et soutenue à l'École d'Économie de Paris le 3 décembre 2020.

**Composition du Jury**

| | | |
|---|---|---|
| Rapporteurs | Erwan Gautier | Banque de France et Université de Nantes |
| | Mirko Wiederholt | Sciences Po |
| Président | Philippe Andrade | Federal Reserve Bank of Boston |
| Examinateurs | Olivier Compte | PSE et ENPC |
| | Xavier Gabaix | Harvard University |
| Directeur | Xavier Ragot | Sciences Po, OFCE et CNRS |

*École des Hautes Études en Sciences Sociales (EHESS)*
*École doctorale Économie Panthéon-Sorbonne*
*Paris-Jourdan Sciences Économiques*
*École d'Économie de Paris (PSE)*

Ph.D. Thesis

Submitted to the École des Hautes Études en Sciences Sociales for the degree of Doctor of Philosophy in Economics.
*Field: Economics - Specialty: Analysis and Policies in Economics*

Jérémy Boccanfuso

# Three Essays on the Implications of Limited Attention in Economics

Thesis Advisor: Xavier Ragot.
Presented and defended at the Paris School of Economics on December 3, 2020.

## Committee Members

| | | |
|---|---|---|
| Reviewers | Erwan Gautier<br>Mirko Wiederholt | Banque de France & Université de Nantes<br>Sciences Po |
| President | Philippe Andrade | Federal Reserve Bank of Boston |
| Examiners | Olivier Compte<br>Xavier Gabaix | PSE and ENPC<br>Harvard University |
| Supervisor | Xavier Ragot | Sciences Po, OFCE and CNRS |

# Remerciements - Acknowledgements

On dit qu'une thèse est une entreprise individuelle, parfois solitaire. Cette thèse doit pourtant à tellement de personnes : les enseignants et chercheurs qui ont su me transmettre une partie de leur savoir, passion et rigueur ; les étudiants et doctorants avec qui j'ai eu la joie de partager mes études et mes premiers pas à l'IPP ; mes amis et parents qui m'ont encouragé et soutenu lors de toutes les étapes de cette thèse. C'est du fond du coeur que je remercie chacun d'entre vous.

Cette thèse est l'aboutissement de longues années de recherche au cours desquelles j'ai bénéficié du soutien constant de mon directeur de thèse, Xavier Ragot. Le contenu de cette thèse a largement bénéficié de ses conseils et encouragements. J'ai énormément appris de nos discussions à l'OFCE, Sciences Po et PSE. Par ailleurs, je suis également extrêmement reconnaissant pour la liberté de recherche et la confiance qu'il m'a accordé lors de ces années. J'ai ainsi eu le loisir d'explorer (et de me perdre dans) mes propres sujets de recherche, tout en bénéficiant de la sécurité offerte par l'ombrelle de sa supervision.

Je remercie également l'ensemble des membres du jury pour avoir accepté d'y participer. Philippe Andrade et Olivier Compte ont suivi l'évolution de mon travail depuis le début de ma thèse. Je les remercie sincèrement pour l'aide qu'ils m'ont apportée durant ces années, leurs innombrables conseils et leur soutien lors de mon Job Market. Je remercie également Xavier Gabaix pour avoir accepté de participer à ce jury, ainsi que les occasions où nous avons pu discuter du contenu de cette thèse. I would also like to thank the two reviewers, Erwan Gautier and Mirko Wiederholt, for their time. Your numerous comments helped me to significantly improve the content of this dissertation. I would also like to thank Mirko for his constant support along my Job Market.

Cette thèse n'aurait pas pu voir le jour sans le financement du Labex OSE offert par PSE et les financements d'ATER à l'université Paris-Dauphine et Sciences Po. Je remercie sincèrement ces institutions pour leurs contributions et la confiance dont elles m'ont fait part. The second and third chapters of this dissertation were respectively awarded at the CEF Graduate Student Paper and ITAX Best PhD Paper contests. I am grateful to these organisations for their financial support. Le formidable cadre de travail offert par PjSE et PSE m'ont également permis de participer à de nombreuses conférences internationales, séminaires internes et, plus généralement, de bénéficier d'un environnement propice à l'élaboration de cette thèse. J'en profite également pour remercier l'ensemble des chercheurs et responsables administratifs présents dans ces institutions pour avoir fait de mon master et doctorat à PSE une période stimulante où tout était mis en œuvre pour que mon seul travail soit de m'épanouir intellectuellement.

Lors de mes années d'études entre Paris 1, la MSE, Cà Foscari, l'Ensae, l'IPP et PSE, j'ai eu l'occasion de côtoyer de formidables personnes dont j'ai énormément appris. J'ai une pensée pour chacune d'entre elles. Parmi ces nombreuses rencontres stimulantes et enrichissantes, c'est également dans ce cadre que j'ai rencontré Antoine Ferey avec qui j'ai co-écrit le troisième chapitre de cette thèse. Je me souviens encore de notre première rencontre. Nous étions dans la cour de l'Ensea où nous venions d'être introduits par des amis communs. Je ne me souviens plus exactement ce qui nous a amené à cela, mais nous nous sommes rapidement retrouvés à discuter de la 'rationalité' des agents économiques. Je suis fier de constater où cela nous a mené quelques années plus tard !

Finalement, certaines personnes ont contribué à cette thèse d'une façon plus indirecte

– mais non moins essentielle – en veillant à mon bien-être quotidien. Je pense en particulier à mes amis de toujours, à mes parents, et à celles et ceux qui m'ont rejoint au cours de cette aventure pour en partager les moindres péripéties. Je n'ai que trois choses à vous dire : merci pour tout, je vous aime et, oui, ça y est, cette thèse est enfin finie ! Soyez néanmoins rassurés car je compte bien continuer de vous parler de mes futurs graphiques trop beaux et équations trop compliquées. Je sais maintenant que cela vous passionne autant que moi (même si vous refusez de me l'avouer pour une raison qui m'échappe ;) ).

À mes parents, merci pour votre soutien et votre amour inconditionnel. J'ai conscience des miracles que vous avez accomplis pour que je puisse en arriver là. Cette thèse vous est dédiée.

Le 5 novembre 2020.

Bienvenue à la petite Mahault qui nous a rejoint il y a quelques heures.

Jérémy Boccanfuso

*À mes parents.*

# Summary

This dissertation is a collection of three essays on the economic implications of limited attention. It is supplemented with a general introduction (Chapter 1).

The first chapter introduces an ongoing paradigm shift in the macroeconomic literature from full-information rational expectations to rationally inattentive economic agents. It then presents some characteristics of this new class of models and current challenges in the literature that motivates the work in this dissertation.

The second chapter is a contribution to consumption theory. It studies the consumption-saving problem of a consumer who faces a fixed cost for paying attention to noisy information and whose attention strategy, i.e., whether or not she pays attention, can be a function of the underlying information. At the optimum, consumers chose to be attentive when evidence accumulates far from their prior beliefs. The model provides an explanation for four puzzling empirical findings on consumption and expectations. First, consumers' attention depends on the information content. Second, aggregate information rigidities vary over the business cycle. Third, consumers only react to large anticipated shocks and neglect the impact of small ones. Fourth, aggregate consumption dynamics vary over the business cycle.

The third chapter is a theoretical contribution to the literature in behavioral public economics. It studies how information frictions in agents' tax perceptions affect the design of actual tax policy. Developing a positive theory of tax policy, it shows that agents' inattention interacts with policymaking and induces the government to implement

inefficiently high tax rates. It then quantifies the magnitude of this policy distortion for the US economy. Overall, the findings suggest that existing information frictions – and thereby tax complexity – lead to undesirable, large and regressive tax increases.

The fourth chapter is an empirical contribution to the macroeconomic literature on information frictions. Using the ECB survey of professional forecasters, it estimates a two margin forecast formation process that allows for forecast rounding on individual and consensus forecast data. Forecasters decide when to revise their forecast (extensive margin). When they do, they slowly incorporate new information (intensive margin) and may report a rounded value for their new forecast (rounding). It finds that these three rigidities simultaneously exist and estimate their respective contribution. The overall forecast stickiness is almost exclusively the consequence of the rigidities at the intensive margin. It then derives quarterly time series for the evolution of information frictions and proposes a simple mapping to account for these variations in economic models.

**Field:** Economics.

**Subfields:** Macroeconomics, Public Economics, Behavioral Economics.

**Keywords:** Information frictions, inattention, consumption, optimal taxation.

# Résumé

Cette thèse est un recueil de trois essais sur les implications économiques de l'attention limitée. Elle est complétée par une introduction générale (chapitre 1).

Le premier chapitre introduit un changement de paradigme, portant sur le processus de formation des anticipations, qui a actuellement lieu en macroéconomie. Il décrit ensuite les caractéristiques principales d'une nouvelle génération de modèles, dont l'objet est de rationaliser l'inattention des agents économiques, afin d'identifier quelques défis actuels dans cette littérature qui sont à la base du travail de cette thèse.

Le deuxième chapitre est une contribution à la théorie de la consommation. Il étudie le problème de consommation et d'épargne d'un consommateur qui fait face à un coût fixe pour être attentif à de l'information imparfaite. La stratégie d'attention de ce consommateur, c'est-à-dire son choix d'être attentif ou non, peut dépendre du contenu de l'information qu'il reçoit. À l'optimum, on trouve que les consommateurs décident de devenir attentifs lorsque suffisamment d'évidences se sont accumulées à l'encontre de leur croyance antérieure (ou *prior*). Le modèle permet ainsi d'expliquer quatre conclusions empiriques déroutantes sur la consommation et les anticipations des consommateurs : (i) l'attention des consommateurs dépend du contenu d'information qu'ils observent, (ii) les rigidités d'information agrégées varient au cours du cycle économique, (iii) les consommateurs ne réagissent qu'aux chocs anticipés de revenu qui sont suffisamment importants et ignorent donc les plus petits et (iv) la dynamique de la consommation agrégée varie au cours du cycle économique.

Le troisième chapitre est une contribution théorique à la littérature en économie publique comportementale. Il étudie la façon dont les rigidités d'information, dans la perception des taxes par les contribuables, affectent la conduite des politiques fiscales. Il développe une théorie positive de la politique de taxation du revenu et montre que l'inattention des contribuables interagit avec la conduite de la politique fiscale, conduisant ainsi le gouvernement à appliquer des taux de taxation inefficacement élevés. Il quantifie ensuite la magnitude de cette distorsion de la politique fiscale aux Etats-Unis. Dans l'ensemble, les résultats de ce chapitre suggèrent que les frictions existantes en matière d'information, et donc la complexité fiscale, conduisent à des hausses d'impôt indésirables, considérables et régressives.

Le quatrième chapitre est une contribution empirique à la littérature macroéconomique traitant des rigidités d'information. Il estime un processus de formation des prévisions à deux marges tenant compte des arrondis sur les données individuelles et agrégées de l'enquête des prévisionnistes de la Banque Centrale Européenne. Les prévisionnistes décident quand réviser leur prévision (marge extensive). Quand ils le font, ils incorporent lentement les nouvelles informations qu'ils observent (marge intensive) et peuvent déclarer une valeur arrondie pour leur nouvelle prévision (arrondis). Le chapitre montre que ces trois forment de rigidités existent simultanément et estime leur contribution respective. La rigidité globale des prévisions est presque exclusivement la conséquence des rigidités à la marge intensive. Le chapitre calcule finalement des séries chronologiques trimestrielles de l'évolution des frictions informationnelles et propose une méthode simple permettant d'incorporer ces variations temporelles dans les modèles économiques.

**Discipline :** Économie.

**Sous-disciplines :** macroéconomie, économie publique, économie comportementale.

**Mots-clés :** rigidités d'information, inattention, consommation, taxation optimale.

## Notes to the reader

The last three chapters of this dissertation are self-contained research articles and can be read separately. They are preceded by a general introduction (Chapter 1) which contextualizes and summarizes the research papers in this dissertation. The term 'paper' is used to refer to chapters (excepted in the general introduction). Chapter 3 is a coauthored paper.

## Remarques pour le lecteur

Les trois chapitres principaux de cette thèse sont des articles de recherche rédigés en anglais et dont la structure est autonome. Ils peuvent donc être lus séparément. Ces chapitres sont précédés d'une introduction générale qui contextualise et résume les contributions de ces articles de recherche. Cette introduction générale est rédigée en français et en anglais. Les termes « papier » et « article » sont utilisés pour faire référence aux différents chapitres (sauf dans l'introduction générale). Le Chapitre 3 est un article de recherche co-écrit.

# Contents

# Chapter 1

# Introduction

## 1.1 General introduction (English)

This Ph.D. thesis is a collection of three original research papers on the economic implications of limited attention. This introduction briefly discusses an ongoing paradigm shift in the macroeconomic literature from full-information rational expectations to rationally inattentive economic agents. It then presents the main characteristics of this new class of models and some current challenges in the literature that motivates the work in this thesis. The introduction ends with an outline that summarizes the content of the subsequent chapters.

### 1.1.1 Rationally Inattentive Agents in Macroeconomics

This section briefly contextualizes a recent shift in the macroeconomic literature from full-information rational expectation models towards models of rationally inattentive economic agents. It emphasizes the role played by the Lucas critique along this evolution.

#### From Full-Information Rational Expectations ...

Understanding and modeling how decisionmakers form their expectations is a crucial methodological question in economics and, perhaps, even more in macroeconomics. The realization that expectations are potentially important for short-run fluctuations in macroeconomic variables is not new and was already the focus of an early economic literature

(Beveridge, 1909; Pigou, 1927; Keynes, 1936). Similarly, it has been rapidly recognized that expectations are also essential for the conduct and efficiency of monetary and fiscal policies.

In the 1960s, the common approach in macroeconomic models was to assume that economic agents had adaptive (or naive) expectations. A major prediction of these models, referred to as the causal interpretation of the Philips curve, was that monetary and fiscal policies could be used to exploit the inflation-unemployment trade off. However, the puzzling stagflation observed in the 1970s was at odds with this conclusion and the Keynesian macroeconomic models developed during the 1960s, that were often supplemented with adaptive expectations, unfortunately proved to be unable to provide useful policy guidance. This period called for a reconsideration of macroeconomic theory. It was the time for macroeconomists to elucidate the role of expectations and their implications for the conduct of economic policies.

> “Muth's idea was that if you take a policy that changes the time series characteristics and some of the variables that you were trying to forecast, people are going to be changing in their forecast rule, and you better have a model that explains exactly how that change occurs. [...]
>
> Now, then came the macro stuff: [...] Neil [Wallace] and Tom [Sargent] took an IS-LM model and just changed the expectations and nothing else, and just showed how that seemingly modest change completely, radically alters the operating characteristics of the system. People noticed at that point. Now we were applying Jack [John Muth]'s ideas to something that wasn't a straw man. It was something a lot of people had invested in, cared a lot about. It was helping to answer some real questions about macro policy, and his, Muth's, ideas start to really matter.”
>
> R. Lucas, panel session, 2011.

A major lesson that emerged from this period was that expectations are endogenous and directly influenced by policy choices (Lucas, 1972). This paradigm shift, that occurred in macroeconomics theory during the 1970s, was centered around what is nowadays known as the Lucas critique (Lucas, 1976). The Lucas critique is significant in the history of economic thought and it suggests that the use of parameters based on previous experience (e.g. adaptive expectations) can be misleading. Intuitively, Lucas claims that agents should incorporate changes in economic policies into their expectations. Accounting for this endogeneity of expectations with respect to actual policies in turn implies that policy-makers cannot systematically fool private agents' expectations and, thereby, e.g., cannot systematically exploit the inflation-unemployment trade off anymore.

From a modeling perspective, this endogeneity of expectations has been introduced through the hypothesis of model-consistent or rational expectations. The rational expectations hypothesis was first proposed by Muth (1961). It is the assumption that people have probability beliefs that coincide with the probabilities predicted by one's model. This hypothesis is actually a double hypothesis. First, it implies that economic agents have a full knowledge of the model-economy and are able to deduce the model consistent distribution of outcomes given the information they have. Second, it requires that agents have full-information on past, present and forecastable events. To emphasize these two hypotheses, the literature sometimes refers to the full-information rational expectations (FIRE) hypothesis.

Over the last 50 years, the FIRE hypothesis was, and remains, the dominant approach to model expectations in the macroeconomic literature. Full-information rational expectations have shaped modern macroeconomic theories and are a building block of, e.g., consumption, price setting, DSGE, asset pricing and financial macroeconomic models. The rational expectations revolution was not limited to the academic sphere. It has also largely influenced actual policy making (Taylor, 2001).

> The [FI]RE benchmark is a natural one to consider, and its use has allowed a tremendous increase in the sophistication of the analysis of dynamics in the theoretical literature in macroeconomics. Nonetheless, the assumption is a strong one, and one may wonder if it should be relaxed, especially when considering relatively short-run responses to disturbances, or the consequences of newly adopted policies that have not been followed in the past — both of which are precisely the types of situations which macroeconomic analysis frequently seeks to address.
>
> M. Woodford, 2013.

Because of its fundamental role, it is essential to critically asses the FIRE hypothesis. On a conceptual ground, the FIRE hypothesis is sometimes seen as an expression of intellectual modesty as it imposes a common knowledge between private economic agents, policymakers and economists.[1] The FIRE hypothesis is, therefore, the assumption of equal knowledge and sophistication: the modeler sets the rules of a model-economy (and an history of economic shocks) and passes on her knowledge to all economic agents in the model-economy without distinction.

**... to Rationally Inattentive Agents**

Yet, at the individual level, FIRE are often criticized for being a heroic assumption. Indeed, it remains a strong definition of rationality as it implicitly requires that agents understand subtle economic mechanisms and are able to efficiently process all the relevant information. It therefore disregards any potential cognitive limitations of individual private agents. One of such potential cognitive limitations that has drawn the attention of researchers is one's limited capacity to collect and process (available) information. Namely, the 'full-information' assumption within the FIRE hypothesis.

[1] "Muth's notion was that the professors [of economics], even if correct in their model of man, could do no better in predicting than could the hog farmer or steelmaker or insurance company. The notion is one of intellectual modesty. The common sense is 'rationality': therefore Muth called the argument 'rational expectations'." McCloskey (1998), p. 53.

> Everyone ignores or reacts sporadically and imperfectly to some information that they 'see'. I page through the business section of the New York Times most mornings, 'seeing' charts and tables of a great deal of information about asset markets. I also most days look at `ft.com`'s charts of within-day movements of oil prices, stock indexes, and exchange rates once or twice. But most days I take no action at all based on this information I've viewed. In fact, if you asked me a half hour after I looked at the paper or the web site what the numbers were I'd viewed, I would usually be able to give at best a rough qualitative answer.
>
> C. A. Sims, 2010.

At an early stage of the rational expectations revolution, some of its initiators considered situations in which economic agents face limited information (Lucas, 1972; Kydland and Prescott, 1982) and clearly illustrated that information constrains matter. For instance, in Lucas' island model (1972) output fluctuations are the sole consequence of individuals' misperception of unpredictable real shocks arising because of their imperfect information.

Recent work in the early 2000s revived this interest on the macroeconomic implications of information frictions (e.g. Mankiw and Reis (2002); Carroll (2003); Sims (2003)). This renewed interest in the expectations formation process has been spurred by several failures of full-information models (see the recent review by Mackowiak et al. (2018)). The end of the decade was also a period when the zero lower bound was binding in most developed economies, and central bankers needed to better understand unconventional policy tools directly targeting private agents' expectations (e.g. forward guidance and communication).

An illustration – that is of particular interest for the present Ph.D. thesis – of the failures of full-information models is the difficulty of dynamic stochastic general equilibrium (DSGE) models with FIRE to match the persistence of macroeconomic variables (Sims, 1998). In the presence of information frictions, economic agents slowly update their beliefs. Ultimately, these frictions generate a sluggish response to economic shocks and are, thereby, an explanation to the observed persistence of macroeconomic variables. Mankiw and Reis (2006) and Maćkowiak and Wiederholt (2015) show that these information frictions are able to reproduce the sluggishness of business cycle fluctuations without having to rely on the usual sources of slow adjustments that are found in the New Keynesian literature (e.g. habit formation, menu costs and Calvo wage setting). In sum, accounting for information frictions is essential to apprehend the sluggish dynamics of aggregate economic outcomes.

Another factor fostering the renewed interest in the analysis of information frictions is the recent abundance of surveys of consumers, firms and professional forecasters' expectations that provide direct empirical evidence regarding the expectation formation process of economic agents. These data clearly reject the hypothesis of full-information in favor of models with slow adjustments in beliefs (Coibion and Gorodnichenko, 2012, 2015; Andrade and Le Bihan, 2013).[2] Also, they highlight the heterogeneity of expectations, another feature that naturally arises in the presence of imperfect information.

> **"In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.**
>
> **H. A. Simon, 1971."**

But in the end, what truly represented a revolution in the last two decades was a conceptual shift within the behavioral macroeconomic literature. A first generation of behavioral macroeconomic models incorporates information frictions through ad-hoc behavioral rules. While these behavioral rules are sometimes supported by empirical evi-

[2]Chapter 4 provides a review of these studies.

dence and experiments, they are not robust to one of the fundamental requirements of macroeconomic modeling: the Lucas critique. Consequently, the conclusions drawn from this first generation of behavioral macroeconomic models focusing on the implications of information frictions were globally disregarded by mainstream macroeconomists.

It is only at the beginning of the century that a new generation of macroeconomic models with endogenous information frictions emerged. This endogeneity arises from the realization that one's attention is scarce. Therefore, each agent faces a trade off governing *when*, *how much* and *which* economic news to be attentive to. These new models share two substantial characteristics.

First, they apprehend one's attention as the result of a cost-benefit arbitrage. Therefore, this new class of models predicts that economic agents are rationally inattentive: when attention is scarce, it is not possible, nor worthwhile, to process each and every bit of available information. As a result, these models explain why economic agents may remain inattentive in a world where information is abundant and cheap.

Second, rationally inattentive agents adjust their attention to changes in the economic environment. Consequently, these new models are, in principle, robust to the Lucas critique.

Thanks to these two characteristics, macroeconomists were less reluctant to listen to the predictions from macroeconomic models accounting for information frictions. Furthermore, and as was already mentioned, these predictions turned out to be important for the study of macroeconomic phenomenons and supported by direct and indirect empirical evidence. All the ingredients were therefore gathered for models of rationally inattentive economic agents to become a prominent tool in the macroeconomic literature.

**Determinants of Attention**

The literature offers a wide range of models focusing on the optimal allocation of attention. Gabaix (2019) provides an overview of these models. The Chapters of this Ph.D. thesis build on an intuitive and widely used model of endogenous attention with stochastic actions: the Noisy information model with endogenous precision. In the following, I briefly discuss some basic elements of this class of models. Importantly, while I consider a simple and static model to highlight some important determinants of agents' attention in this introduction, the intuitions gained in this section are general and would continue to hold with more sophisticated specifications (e.g. dynamic setups, different costs of attention, different objective function).

Consider an economic agent who chooses her action $y$ to maximize her utility. For illustration, assume that her optimal action with full-information rational expectations depends linearly on a stochastic state variable $x \sim \mathcal{N}(0, \sigma_x^2)$. That is,

$$y^{FIRE} = ax \tag{1.1}$$

where the parameter $a$ denotes the elasticity of the FIRE optimal action to the state $x$.

When this agent is imperfectly informed, she cannot directly observe the stochastic state $x$ as she would with perfect information and must, therefore, form an expectation about this state. When forming this expectation, she has some knowledge about the state variable $x$. First, she is endowed with a prior belief about the potential outcomes for the state variable $x$. It is assumed that this prior belief is accurate – that is, we keep on maintaining the 'rational expectations' assumption within the FIRE hypothesis. Here, it implies that the agent knows that the state variable follows a Gaussian distribution with mean 0 and variance $\sigma_x^2$.

Second, she can decide to gain some information about the realization of the state variable. In Noisy information models, we consider that she can observe a signal $s = x + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a Gaussian white noise. The signal precision $1/\sigma_\epsilon^2$ is a proxy for the agent's attention. When the signal precision maps to infinity (i.e. $\sigma_\epsilon^2 = 0$), the signal is always equal to the realization $x$ and we say that the agent is perfectly attentive as she perfectly processes all the information relevant to determine $x$. This extreme case coincides with full-information rational expectations. On the other hand, as the signal precision gets smaller, the agent becomes less attentive and the signal she receives is less informative about $x$.

Then, *what are the determinants of an agent's attention?* To provide an intuitive answer to this question, let's assume that she is aiming to minimize the expected squared deviation from her optimal action under full-information plus an attention cost $\mathcal{C}(1/\sigma_\epsilon^2)$ that increases with the signal precision. Then, for a given signal precision, it is optimal to choose the action[3]

$$y^\star = aE[x|s] \tag{1.2}$$

where $E[x|s]$ is the mathematical expectation of $x$ given the signal $s$. As a result, she chooses the signal precision or, equivalently, the variance of the noise $\sigma_\epsilon^2$ such that[4]

$$\min_{\sigma_\epsilon^2 \geq 0} \quad a^2 \frac{\sigma_\epsilon^2 \sigma_x^2}{\sigma_\epsilon^2 + \sigma_x^2} + \mathcal{C}(1/\sigma_\epsilon^2) \tag{1.3}$$

where the ratio $\sigma_\epsilon^2 \sigma_x^2 / (\sigma_\epsilon^2 + \sigma_x^2)$ is the posterior uncertainty, i.e., the expected uncertainty about $x$ that remains after observing a signal $s$. Solving for the first order condition, the optimal signal precision is

$$\frac{1}{\sigma_\varepsilon^2} = \frac{a}{\sqrt{\mathcal{C}'(\cdot)}} - \frac{1}{\sigma_x^2} \tag{1.4}$$

When a solution exists, attention is thus governed by three parameters in this simple framework:

- It increases with the elasticity of the FIRE optimal action ($a$).
- It increases in uncertain environment ($\sigma_x^2$).
- It decreases with the marginal cost of attention ($\mathcal{C}'(\cdot)$).

Empirical evidence confirms that these factors are important determinants of attention allocation.[5] Importantly, the agent's attention only depends on deep parameters. That is, exogenous and time-invariant factors that describes the structural characteristics of the model-economy. These are unaffected by economic shocks.

[3] The conditional expectation is MSE-optimal.
[4] See Mackowiak et al. (2018) section 2.1 for the derivation of this expression.
[5] See the literature reviews in Gabaix (2019) and the following chapters of the present thesis.

### 1.1.2 Current Challenges in the Literature

This section summarizes some important challenges related to inattention and its implications in economics. The following chapters of this Ph.D. thesis are original contributions to these research agendas.

#### Determinants of Attention: The Lucas Critique Strikes Back

Empirical evidence indicates that other factors also directly affect agents' attention. For instance, we observe significant variations in attention which are correlated to business cycle fluctuations. Coibion and Gorodnichenko (2015) and the Chapter 4 of this Ph.D. thesis show that agents' attention increases persistently in the aftermath of a recession. Moreover, recent experiments identify a causal relationship between the information content and the attention paid to this information (Abe and Ueno, 2015; Armantier et al., 2016; Khaw et al., 2017).

Such state-dependence and/or information-dependence of attention is hardly reconcilable with standard models of attention allocation. This can be illustrated from the first order condition (1.4).[6] The model predicts that attention only depends on deep parameters, i.e., those parameters that govern the model-economy and are time-invariant in the absence of sudden and unexpected structural breaks. As a result, any variation in attention over time should be predetermined and unrelated to variations in the state variable $x$ (i.e. state-dependence) or the signal $s$ (i.e. information-dependence).

Yet, accounting for the observed state-dependence and/or information-dependence of attention is essential as it directly affects agents' response to changes in the economic environment. Indeed, the conditional mathematical expectation $E[x|s]$ in equation (1.2) depends on attention: the more attentive an agent is, the greater the weight she puts on her signal $s$ to form her expectation. More specifically, we have[7]

$$E[x|s] = \xi s = \xi(x + \varepsilon) \tag{1.5}$$

[6]The simple static model presented in this section remains helpful in this regard and continues to provide the essential intuitions that we would get in a dynamic setup.

[7]When the prior and signal distributions are Gaussian, the conditional expectation is a linear combinaison of both the prior average (which is nil here) and the signal.

where $\xi = \sigma_x^2/(\sigma_x^2 + \sigma_\varepsilon^2) \in [0,1]$ is the weight on the signal that reflects the level of attention. By not being able to capture the dependence of the weight $\xi$ on $s$ and/or $x$, most models of endogenous attention[8] are implicitly relying on reduced-form expectations that do not fully internalize the impact of changes in the economic environment. This is, in essence, the same critique that Lucas raised against Keynesian models. Hence, a new research agenda is aiming to account and provide a rational for the observed state-dependence of attention. Hopefully, this research agenda will improve the ability of macroeconomic models with rationally inattentive agents to provide accurate policy recommendations.

*Related Contributions: Chapters 2 and 4.*

**Interactions between Attention and the Dynamics of Economic Aggregates**

The literature argues that inattention is a source of the sluggishness of business cycle fluctuations (Sims, 2003; Mankiw and Reis, 2006; Maćkowiak and Wiederholt, 2015). Yet, empirical evidence suggests that inattention is not constant over time and is correlated to business cycle fluctuations (Coibion and Gorodnichenko, 2015). A promising challenge is therefore to understand how inattention and business cycle fluctuations interact with one another.

Chapter 2 demonstrates that consumption growth dynamics is a prominent illustration of this interaction. The persistence of aggregate consumption growth dropped in the aftermath of the 2008 recession (Kumar and Jia, 2019). This sudden change in the dynamics of consumption is puzzling; it is also unfortunate as the persistence of consumption is supposed to act as a smoothing mechanism when the economy enters in a recessionary period. Even though this shift in the dynamics of consumption was particularly brutal during the 2008 recession, Kumar and Jia (2019) report similar patterns during most of the recessions in the U.S. since 1980. In Chapter 2, I argue that this systematic drop in the persistence of aggregate consumption growth is the consequence of consumers' attention. The intuition is as follows. Accounting for the information-dependence of attention, consumers becomes (endogenously) more attentive on average when the economy

[8]See the literature review in Chapter 2.

enters in a recession. When consumers are more attentive, they react more promptly to economic shocks and the persistence of aggregate consumption is lower. The increase in consumers' attention during recession therefore explains the decrease in consumption persistence during these periods.

Chapter 2 is one of the very few papers that focuses on the *interaction* between agents' attention and the dynamics of economic aggregates. Importantly, it shows that this interaction matters for the study of macroeconomic phenomenons.

*Related Contributions: Chapters 2 and 4.*

**Economic Policies Shape Attention**

The design of economic policies shapes agents' attention to these policies which, then, governs their response to these policies. To illustrate this statement, consider the following three examples discussed in greater details in the following chapters of this Ph.D. thesis.

Income transfers are a straightforward illustration of a situation where the design of an economic policy shapes consumers' attention to this policy. Consider a government which unexpectedly annonces a lump-sum transfer in the near future. Following from the permanent income hypothesis, a consumer should adjust her consumption immediately after hearing the government's announcement,[9] independently of the size of the announced transfer. However, in the data, we observe that such adjustment is likely to be observed only when the announced transfer is large (Browning and Collado, 2001; Scholnick, 2013; Kueng, 2018). Chapter 2 demonstrates that this dependence of the consumption response with respect to the size of the transfer could be the consequence of consumers' attention. This is because consumers are more likely to be attentive to large transfers and inattentive to small ones.

At the aggregate level, the same mechanism implies that consumers' attention interact in nonlinear and non-trivial ways with stabilizing policies. During recessions, consumers are less likely to be attentive to positive news. As a result, the consumption channel

[9] In the absence of market frictions preventing her to do so.

is partially muted during these periods and, thereby, monetary and fiscal expansionary policies are less likely to be effective. One way to overcome this muted consumption channel is to implement aggressive expansionary policies that will trigger consumers' attention.

In a similar vein, Chapter 3 shows that when setting an income tax, the government faces tradeoffs related to taxpayers' endogenous attention. For instance, an increase in the tax rate makes taxpayers willing to be more attentive to the tax. As a result, tax increases make the tax more salient and taxpayers more responsive to it. Ultimately, this endogeneity of attention limits the government ability to leverage taxpayers' inattention.

In each of these examples, we see that the design of an economic policy (e.g. the size of a policy intervention) shapes agents' attention to this policy and, thereby, how agents respond to this policy. Overall, these examples illustrate a major challenge for economists: the endogeneity of attention generates non trivial interactions between the design of economic policies and agents' perceptions and expectations. Gaining new insights on this interaction is essential to provide useful policy recommendations.

*Related Contributions: Chapters 2 and 3.*

**Inattention Alters the Conduct of Policies**

Another challenge that economists are facing is to understand how agents' inattention affects the conduct of economic policies. The idea that policymakers can capitalize on individuals' limited attention in a welfare enhancing way is emerging in, e.g., behavioral public finance and political economy.[10] However, little is known about the behavior of a government which has the power to leverage agents' inattention. Analyzing these questions requires new policy frameworks that use tools from game theory to analysis the equilibrium outcomes in games where policymakers may seek opportunities to benefit from agents' misperceptions.

Chapter 3 belongs to this research agenda and demonstrates that inattention resurges old debates about the discretion of economic policies in setups that were thought to be

[10] See the literature review in Chapter 3.

immune to it. Indeed, when agents are partially inattentive to economic policies, they anchor their expectations on their prior beliefs and, thereby, under-react to actual changes in policies. A discretionary government therefore has an incentive to leverage these under-reactions. However, in equilibrium, agents' prior adjusts and internalizes the government's willingness to deviate. The equilibrium outcome turns out to be suboptimal in a game where a discretionary government has the power to benefit from agents' inattention.

*Related Contributions: Chapters 3.*

### 1.1.3 Outline of the Thesis

This Ph.D. thesis is a collection of three independent research papers studying the economic implications of information frictions. Each chapter starts with an introduction that reviews the relevant literature and underlines the paper contributions in details. In this outline, I briefly summarize the content of each chapter.

**Chapter 2 – Costly Information Processing and Consumption Dynamics**

Chapter 2 is a theoretical contribution to the study of consumption dynamics. It is motivated by four puzzling empirical observations that are hardly reconcilable with existing theories of rationally inattentive consumers. First, consumers' attention depends on the information content. Using information experiments, Armantier et al. (2016), Abe and Ueno (2015) and Khaw et al. (2017) find that individuals are more likely to incorporate new information when it contradicts their prior beliefs. Second, information rigidities are not constant over time. Dräger and Lamla (2012) and Coibion and Gorodnichenko (2015) find that information rigidities drop persistently in the aftermath of a recession. Third, consumers only react to large anticipated shocks and neglect the impact of small ones. Jappelli and Pistaferri (2010) term this mechanism the *magnitude hypothesis* and empirical evidence in favor of the hypothesis is accumulating (Browning and Collado, 2001; Scholnick, 2013; Kueng, 2018). Fourth, aggregate consumption dynamics vary over the business cycle. Caballero (1995) finds that "in good times, consumers respond more promptly to positive than to negative wealth shocks, while the opposite is true in bad times" and Kumar and Jia (2019) report systematic decreases in consumption growth persistence during recessions.

This chapter proposes a model of attention allocation that can match these four empirical findings. In the model, there is a source which conveys noisy information about shocks to one's permanent income. Consumers face a fixed cost for paying attention to this information and their attention strategy, i.e., whether or not an agent pays attention, can be a function of the underlying information. Apart from this novel information structure, the consumption-saving problem considered in the paper is left as transparent as possible and coincides with Hall's (1978) random walk model with quadratic utility. The model predictions are thus isolated from other refinements of the textbook consumption theory.

The main findings of the chapter are as follows. First, consumers' attention depends on the information content. At the optimum, the consumer faces an inattention region where she disregards new information and does not adjust her consumption plan. It is only when evidence against her prior beliefs accumulates that the consumer is willing to pay attention to new information releases and revises her consumption plan accordingly.

Second, information rigidities are time-varying. Consumers adopt an information-dependent attention strategy. Therefore, and starting from the stationary cross-sectional distribution of consumers, an aggregate income shock prompts more consumers to be attentive. Information being noisy, the impact of an aggregate shock disseminates slowly in the economy and the increase in the share of attentive consumers is persistent.

Third, the model predicts a positive correlation between the size of an income shock and the marginal propensity to consume. It thus provides a rationale to explain *the magnitude hypothesis*. The intuition is as follows. For a consumer who was last attentive one period ago, a small permanent income shock (in absolute value) is unlikely to be significant enough to trigger her attention. However, as the size of the shock gets larger, the shock becomes more likely to trigger the consumer's attention and to prompt her to revise her consumption path. More generally, the expected marginal propensity to consume out of a shock to permanent income is found to be both history-dependent and shock-dependent.

Fourth, aggregate consumption dynamics vary over the business cycle. The model predicts that during economic busts (respectively booms) a positive (resp. negative) shock generally lowers the share of attentive consumers and generates a smoother response, while a negative (resp. positive) shock generates a sharper response. Moreover, the persistence of aggregate consumption growth depends on the endogenously time-varying share of attentive consumers. In normal times, information rigidities are near their steady state level and consumption persistence relatively constant. However, during unusual times such as recessions, information rigidities decrease a lot and so does aggregate consumption persistence.

This chapter therefore deviates from the existing literature in that it captures some non linear features of information frictions. These non linearities conceal important implications for the propagation of economic shocks in macroeconomic models. They suggest that information frictions are large in normal times, thus resulting in sluggish hump-shaped responses of macroeconomic outcomes. However, these frictions partially vanish as the economy gets farer away from its steady state, and so does the persistence in the dynamics of macroeconomic outputs. In the latter situations, the dynamics of macroeconomic outcomes are much less sluggish and gets closer to that predicted by frictionless economies. Therefore, information frictions also act as a stabilizing mechanism toward steady state dynamics. These novel findings also hold significant implications for the effectiveness of monetary and fiscal policies during recessions.

In sum, Chapter 2 provides an explanation to the aforementioned four facts indicating that consumers' attention depends on the information content, information rigidities are time-varying, consumers only react to large anticipated shocks, and aggregate consumption dynamics vary over the business cycle. Doing so, it directly contributes to the first three research agendas mentioned in Section 1.1.2.

**Chapter 3 – Inattention and the Taxation Bias**

Chapter 3, a joint work with Antoine Ferey, studies how information frictions in agents' tax perceptions affect the design of actual tax policy. A growing body of evidence documents substantial information frictions in agents' tax perceptions (Chetty, 2015; Bernheim and Taubinsky, 2018; Stantcheva, 2019). In light of this evidence, a burgeoning normative literature analyzes the design of optimal tax policy in the presence of information frictions in agents' tax perceptions. This literature characterizes optimal tax policies in terms of sufficient statistics that capture agents' earnings responses to tax changes and perception biases at the optimum. Doing so, it generally sidesteps the issue that agents' tax perceptions may adjust to changes in tax policy and remains agnostic about the mechanisms behind these adjustments.

In practice, tax policy is likely influenced by the way agents' perceptions adjust to tax changes. Policymakers may for instance be tempted to increase taxes if agents are inattentive and only perceive a fraction of tax increases. In contrast to their normative counterparts, such positive policy implications remain surprisingly unexplored. Chapter 3 aims at filling this gap by studying how information frictions in agents' tax perceptions affect the design of actual tax policy.

Chapter 3 proposes a positive theory of tax policy in a setting where agents' labor supply is determined by their tax perceptions. It shows that the adjustment of agents' tax perceptions interacts with policymaking and generates a distortion in actual tax policy. Specifically, inattention leads the government to implement inefficiently high tax rates: this is the taxation bias.

Central to this result is a dichotomy between direct and indirect adjustments in perceptions upon changes in tax policy. Indeed, as agents' tax perceptions are determined by a combination of the actual tax rate and their prior, there are two margins through which perceptions may adjust: a *direct* margin capturing the attention agents devote to observing taxes and thus changes in tax policies, and an *indirect* margin capturing variations in the prior. For a given prior, inattentive agents only perceive a fraction of the change in tax policy which dampens their earnings responses. The government thus targets a higher tax rate than if agents were perfectly attentive. In equilibrium, agents' priors must however be consistent with the government's choice of tax policy. As a result, ex post earnings responses are larger than what anticipated ex ante. The government implements inefficiently high tax rates because it fails to internalize the indirect adjustment of the prior (arising as an equilibrium mechanism) in its choice of tax policy. In a nutshell, taxpayers' inattention to taxes creates the illusion that tax reforms induce lower efficiency costs than they actually do and ultimately prompts the government to misbehave from a normative perspective. Fundamentally, this reflects a commitment problem.

We then seek to illustrate the implications of this policy distortion and to quantify its magnitude. First, we fit a linear tax model to US tax data. Further relying on the existing empirical literature to calibrate our sufficient statistics, we estimate that the taxation bias is approximately equal to 3.7 percentage points. This means that the linearized US income tax rate is more than 12% higher than what would be optimal holding the government's objective constant: the taxation bias is large.

Second, we extend our analysis to nonlinear tax schedules. The government's incentive to increase the marginal tax rate at a given earnings level then depends on agents' attention at (or close to) this earnings level. As a consequence, the positive correlation between income and attention results in an income-specific taxation bias that is globally decreasing with income: the taxation bias is large at low income levels and virtually nonexistent at top income levels. The taxation bias thus attenuates the U-shape pattern of marginal tax rates (Saez, 2001) and reduces the progressivity of actual income tax schedules.

In sum, Chapter 3 develops a positive theory of tax policy and shows that agents' inattention interacts with policymaking and induces the government to implement inefficiently high tax rates. This taxation bias is found to be undesirable, large and regressive. Doing so, this chapter directly contributes to the last two research agendas discussed in Section 1.1.2.

**Chapter 4 – Forecast Stickiness along the Business Cycle**

The excess smoothness in economic agents' expectations, which arises as a consequence of information frictions, drives the persistence of aggregate variables and the transmission of monetary and fiscal policies. It is therefore a central topic in macroeconomics. However, economists have yet to concur on the source(s) of these frictions, whether they are a characteristics of individual data, their evolution over time at high frequencies, and their respective contribution to the excess smoothness of expectations.

Chapter 4 provides an empirical contribution to each of these important questions. It estimates a two margin forecast formation process that allows for forecast rounding. Forecasters decide when to revise their forecast (extensive margin, e.g. Sticky information). When they do, they slowly incorporate new information (intensive margin, e.g. Noisy information) and may report a rounded value for their new forecast (rounding, e.g. consideration sets). Simultaneously accounting for these three forms of rigidities allows to encompass a broad class of potential frictions in forecast formation and to asses their relative contribution.

The main contributions of the chapter are as follows. First, the previous literature measures the rigidities at the extensive margin from the observed share of zero forecast revisions. The chapter develops a simple test for the validity of this measure when rigidities exist at both the extensive and intensive margins. Applying this test on individual forecasts for inflation, GDP growth and unemployment from the ECB Survey of Professional Forecasters, the data clearly reject the validity of this measure.

Second, the chapter extends the two margin forecast model to allow for forecast rounding at both margins. It estimates this model on individual forecast data and finds that the three forms of rigidities coexist. About 80% of the zero-revision that were previously attributed to Sticky information are actually the consequence of forecast rounding. The rigidities at the extensive are therefore much smaller than previously reported: about 95% of European forecasters revise their forecast each quarter. Importantly, this new measure of the rigidities at the extensive margin is not rejected by the data anymore. The rigidities at the intensive margin are large: forecasters incorporate only 55-60% of the information they receive each quarter.

Third, the chapter assesses the relative contribution of each form of rigidities. In macroeconomic models with representative agents, information frictions generally affect macroeconomic outputs through their impact on the excess smoothness of average expectations. Therefore, the chapter shows how the three forms of rigidities estimated in the data map into a single parameter that drives the excess smoothness of the consensus (average) forecast under a mild approximation. Decomposing the average excess smoothness, we find that it is almost exclusively the consequence of the rigidities at the intensive margin. Rigidities at the extensive margin only account for 4 to 5% of this excess smoothness and rounding for less than 0.04%.

Fourth, the chapter derives quarterly time series for the evolution of the excess smoothness in forecasts. The trend in information acquisition has remained relatively constant over the 1999-2019 period and there is no sudden structural change over this period. We nevertheless observe medium term variations that are common to all three series (inflation, GDP growth and unemployment). Information acquisition has slowly decreased over the 2000-2008 period, a period of relative economic stability in the euro area. It started to increase afterward and remained stable in between the two most recent recessions. Since 2016, the trend in information acquisition seems to be on an increasing path again. Notably, information acquisition has become more volatile in the post 2008 area. Regarding the contribution of each form rigidities to these variations over time, we reach a similar conclusion than for the average: variations in information rigidities are almost exclusively the consequence of the intensive margin.

Fifth, the chapter studies the state-dependence of information frictions by studying its dynamics during recessions. Similarly to previous studies (Coibion and Gorodnichenko, 2015), we observe an overall decrease in information frictions in the aftermath of a recession. However, the dynamics are very different for the three variable-specific time series. More specifically, we find a sharp increase in information rigidities for inflation and GDP growth that is more than compensated by a decrease in information frictions for unemployment.

In sum, Chapter 4 proposes a novel methodology to simultaneously estimate a large class of information rigidities on individual forecast data. This methodology is then used to show that while many forms of rigidities are indeed a characteristics of the data, only the inattention to new information while revising a forecast is relevant for macroeconomic models with representative agents. It then derives quarterly time series to gain new insights on the time variation of information frictions. Doing so, this chapter directly contributes to the first two research agendas discussed in Section 1.1.2.

## 1.2 Introduction générale (Français)

Cette thèse de doctorat rassemble trois articles de recherche originaux traitant des implications économiques de l'attention limitée. L'introduction générale de cette thèse commence par brièvement discuter un changement de paradigme, portant sur le processus de formation des anticipations, qui a actuellement lieu en macroéconomie. Elle décrit ensuite les caractéristiques principales d'une nouvelle génération de modèles, dont l'objet est de rationaliser l'inattention des agents économiques, afin d'identifier quelques défis actuels dans cette littérature qui sont à la base du travail de cette thèse. Finalement, l'introduction se termine par un plan de thèse qui résume le contenu des chapitres suivants.

### 1.2.1 Les agents rationnellement inattentifs en macroéconomie

Cette première partie contextualise brièvement l'évolution récente des modèles à anticipations rationnelles avec information parfaite vers des modèles à inattention rationnelle dans la littérature macroéconomique. L'accent est mis sur le rôle joué par la critique de Lucas à travers cette évolution.

#### Des anticipations rationnelles avec information parfaite...

Comprendre et modéliser la façon dont les agents économiques forment leurs anticipations est une question méthodologique cruciale en économie et, peut-être encore plus, en macroéconomie. La réalisation que ces anticipations sont déterminantes pour les fluctuations macroéconomiques de court terme n'est pas nouvelle et est le résultat d'une littérature relativement ancienne (Beveridge, 1909; Pigou, 1927; Keynes, 1936). Cette même littérature indique également que les anticipations des agents économiques privés sont essentielles pour la conduite et l'efficacité des politiques monétaires et fiscales.

Durant les années 1960, l'approche dominante en macroéconomie était de modéliser les anticipations des agents économiques comme étant adaptatives ou naïves (c'est-à-dire uniquement basées sur l'observation de données passées). Une des prédictions les plus influentes de ces modèles, que l'on nomme la courbe de Phillips originelle, est que les politiques fiscales et monétaires peuvent être utilisées afin d'exploiter l'arbitrage entre inflation et chômage. Néanmoins, la surprenante *stagflation* observée durant les années

1970 s'est avérée être en opposition avec les conclusions de la courbe de Phillips originelle et, plus généralement, avec les modèles Keynésiens qui reposaient généralement sur des anticipations adaptatives. Cette période motiva une reconsidération en profondeur de la théorie macroéconomique. Il était temps pour les macroéconomistes d'élucider le rôle joué par les anticipations et leurs implications pour les politiques économiques.

> “ L'idée de Muth était que si l'on prend une politique qui change les caractéristiques des séries temporelles et certaines variables que l'on essaie de prédire, les individus vont ajuster leurs règles de prévision et il vaudrait mieux avoir un modèle qui explique exactement comment cet ajustement a lieu [...]
>
> Puis sont venus les trucs macro : [...] Neil [Wallace] et Tom [Sargent] ont pris un modèle IS-LM et ont juste changé les anticipations des agents. Rien de plus. Ils ont montré que ce changement apparemment modeste modifie radicalement les mécanismes élémentaires du modèle. Les gens ont été attentifs à partir de ce moment-là. Nous étions maintenant en train d'appliquer les idées de Jack [John Muth] à quelque chose qui n'était plus qu'un grain de sable jeté dans l'océan. Le modèle IS-LM était une théorie dans laquelle beaucoup de gens avaient investi et dont ils se souciaient. Ce modèle permettait de répondre à des questions réelles concernant les politiques macroéconomiques. C'est à partir de ce moment là que les idées de Muth ont commencé à être vraiment importantes. ”
>
> R. Lucas, panel session, 2011. Traduit de l'anglais.

Une leçon essentielle qui ressortit de cette reflexion fût que les anticipations sont endogènes et directement influencées par les politiques économiques (Lucas, 1972). Ce changement de paradigme durant les années 1970 était centré autours de ce que l'on appelle maintenant la critique de Lucas (Lucas, 1976). Cette critique est déterminante dans l'histoire de la pensée économique et suggère que l'utilisation de paramètres basés sur des données passées, tels que les anticipations adaptatives, peut être captieux. Intuitivement, Lucas affirme que les agents économiques doivent incorporer les changements de politiques économiques dans leurs anticipations. La prise en compte de cette endogénéité des anticipations vis-à-vis des politiques économiques implique que les décideurs politiques ne peuvent pas systématiquement tromper les attentes des agents privés et, par conséquent et contrairement à ce qui avait été affirmé, qu'ils ne peuvent par exemple pas librement exploiter l'arbitrage entre inflation et chômage.

En termes de modélisation, cette endogénéité des anticipations a été introduite à travers l'hypothèse d'anticipations rationnelles (ou d'anticipations cohérentes avec le modèle). L'hypothèse d'anticipations rationnelles a été proposée par Muth (1961). Elle suppose que les individus ont des croyances probabilistiques qui coïncident avec les probabilités du modèle. De fait, l'hypothèse d'anticipations rationnelles est une hypothèse double. D'une part, elle implique que les individus ont une connaissance parfaite du système économique dans lequel ils évoluent et qu'ils sont capables d'en déduire une distribution des issues potentielles qui sera cohérente avec ce système étant donné l'information dont ils disposent. D'autre part, cette même hypothèse admet que les individus ont également une information parfaite des événements passés, présents et prévisibles dans le futur. Afin de souligner ces deux aspects sous-jacents à l'hypothèse d'anticipations rationnelles, la littérature fait parfois référence à la dénomination plus explicite d'anticipations rationnelles avec information parfaite (ou à l'acronyme FIRE pour *Full-Information Rational Expectations*).

Au cours des cinq dernières décennies, l'hypothèse FIRE a été, et reste encore, l'approche dominante en macroéconomie afin de modéliser les anticipations. Les anticipations rationnelles avec information parfaite ont façonné les théories macroéconomiques et sont, par exemple, une composante essentielle des modèles de consommation, de détermination des prix, dynamiques stochastiques d'équilibre général (*DSGE*) et de macroéconomie financière. La révolution des anticipations rationnelles ne s'est pas limitée à la sphère académique. Elle a également eu un écho important parmi les décideurs politiques (Taylor, 2001).

> L’hypothèse d’anticipations rationnelles [avec information parfaite] est une référence naturelle qui se doit d’être considérée. Son utilisation a permis un formidable raffinement de l’analyse des [processus] dynamiques dans la littérature théorique en macroéconomie. Néanmoins, cette hypothèse est une hypothèse forte et on peut se demander si elle doit être assouplie, en particulier lorsque l’on considère les réponses de relativement court terme suite à des perturbations [économiques] ou les conséquences de politiques nouvellement adoptées et qui n’ont jamais été observées dans le passé. Deux types de situations auxquelles les analyses macroéconomiques sont régulièrement confrontées.
>
> M. Woodford, 2013. Traduit de l’anglais.

À cause de son rôle fondamental, il est essentiel d’évaluer l’hypothèse FIRE de façon critique. D’un point de vue conceptuel, cette hypothèse est parfois considérée comme une expression de modestie intellectuelle puisqu’elle impose un savoir commun entre les agents économiques privés, les décideurs politiques et les économistes[11]. L’hypothèse FIRE est, ainsi, l’hypothèse d’égalité de la connaissance et de sophistication : le modélisateur définit les règles du système économique (et une histoire de chocs économiques) et transmet son savoir à tous les agents économiques qui évoluent dans ce système sans distinction.

### ...à des agents rationnellement inattentifs

Pourtant, d’un point de vue individuel, il est souvent reproché à l’hypothèse FIRE d’être héroïque. En effet, elle reste une définition forte de la rationalité puisqu’elle nécessite que les individus comprennent des mécanismes économiques subtiles et qu’ils soient capables de traiter efficacement toutes les informations pertinentes. Elle néglige ainsi toute

[11] « La notion de Muth était que les professeurs [d’économie], même s’ils sont corrects dans leur modèle de l’homme, ne pouvaient pas mieux prévoir que l’éleveur de porcs, le sidérurgiste ou la compagnie d’assurance. Cette notion est une notion de modestie intellectuelle. Le sens commun est ‘la rationalité’ : par conséquent, Muth nomma cet argument ‘les anticipations rationnelles’. » McCloskey (1998), p. 53. Traduit de l’anglais.

contrainte cognitive potentielle de la part des agents économiques privés. Une de ces contraintes cognitives potentielles, qui a attiré l'attention des économistes, est la capacité limitée des individus à rassembler et traiter les informations (pourtant disponibles). C'est-à-dire, le caractère d' « information parfaite » des « anticipations rationnelles avec information parfaite ».

> Tout le monde ignore ou réagit sporadiquement et imparfaitement à de l'information qu'il 'voit'. Je feuillette la section économie du *New York Times* la plupart des matins, 'voyant' des graphiques et tableaux contenant beaucoup d'informations sur les marchés d'actifs. Je regarde également les graphiques de `ft.com` concernant les variations intra-journalières du prix du pétrole, des indices boursiers et des taux de change, généralement une à deux fois par jour. Mais la plupart des jours, je n'agis pas en conséquence de ces informations que j'ai vues. En fait, si vous me demandiez 30 minutes après que j'ai regardé le journal ou le site internet quels sont les chiffres que j'ai vus, je ne serais, d'ordinaire, capable que de vous donner au mieux une réponse qualitative approximative.
>
> C. A. Sims, 2010. Traduit de l'anglais.

À un stade précoce de la révolution des anticipations rationnelles, certains de ses initiateurs ont considéré des situations dans lesquelles les agents économiques ne disposent que d'une information limitée (Lucas, 1972; Kydland and Prescott, 1982). Ces travaux montrent indiscutablement que les contraintes d'information ont de l'importance. Par exemple, dans le modèle des îlots de Lucas (1972), les fluctuations du cycle économique sont la seule conséquence des erreurs de perception des chocs réels imprévisibles engendrées par l'information imparfaite dont disposent les individus.

Au début des années 2000, plusieurs travaux ont ressuscité cet interêt pour l'analyse des implications macroéconomiques des frictions d'information (e.g. Mankiw and Reis (2002); Carroll (2003); Sims (2003)). Cet intérêt renouvelé pour le processus de formation des anticipations a été motivé, en partie, par plusieurs échecs des modèles avec information parfaite (voir la revue de littérature récente de Mackowiak et al. (2018)). La fin de la décennie a également été marquée par une période où les taux d'intérêt directeurs de la plupart des pays développés se sont rapprochés de zéro. Les banquiers centraux avaient ainsi également besoin de mieux comprendre le fonctionnement des outils monétaires non-conventionnels ciblant les anticipations des agents privés (comme par exemple la *forward guidance* et la communication des banques centrales).

Une illustration de la défaillance des modèles macroéconomiques avec information parfaite, qui sera fréquemment discutée au cours de cette thèse, est la difficulté des modèles dynamiques stochastiques d'équilibre général (DSGE) avec l'hypothèse FIRE à reproduire la persistance de la dynamique des variables macroéconomiques (Sims, 1998). Lorsqu'ils font face à des rigidités d'information, les agents économiques révisent lentement leurs croyances. Cette lente révision des croyances implique que les agents répondent également lentement aux chocs économiques. Les rigidités d'information peuvent, par conséquent, expliquer la forte persistance des variables macroéconomiques que l'on retrouve dans les données. Mankiw and Reis (2006) et Maćkowiak and Wiederholt (2015) montrent qu'il est possible de reproduire l'inertie des cycles économiques en tenant compte de ces rigidités d'information, sans pour autant avoir à recourir aux nombreuses rigidités d'ajustement que l'on retrouve dans les modèles des nouveaux Keynésiens (habitudes de consommation, coûts de menu, fixation des salaires à la Calvo, etc.). En conclusion, tenir compte des rigidités d'information est essentiel afin d'appréhender l'inertie qui caractérise la dynamique de nombreuses variables macroéconomiques.

Un autre facteur qui a favorisé cet interêt renouvelé pour l'analyse des rigidités d'information est l'abondance récente d'enquêtes sur les anticipations des consommateurs, des entreprises et des prévisionnistes. Ces enquêtes permettent d'obtenir des évidences empiriques concernant la façon dont les agents économiques forment leurs anticipations qui sont directement fondées sur les anticipations reportées par ces mêmes agents. Les évidences ainsi obtenues rejettent clairement l'hypothèse d'information parfaite en faveur de modèles dans lesquels les agents révisent lentement leurs croyances (Coibion and Gorodnichenko, 2012, 2015; Andrade and Le Bihan, 2013). De plus, elles mettent en avant le caractère hétérogène des anticipations, une autre spécificité qui émerge naturellement en présence de rigidités d'information[12].

> “ **Dans un monde où l'information est partout, l'abondance d'information signifie une pénurie de quelque chose d'autre : une rareté de quoi que ce soit que l'information consomme. Ce que consomme l'information est plutôt évident : elle consomme l'attention de son bénéficiaire. Ainsi, une richesse d'information crée une pauvreté d'attention et un besoin de distribuer efficacement cette attention à travers la surabondance des sources d'information qui peuvent la consommer.**
>
> **H. A. Simon, 1971. Traduit de l'anglais.** ”

Mais finalement, ce qui a vraiment représenté une révolution au cours des deux dernières décennies est un changement conceptuel drastique dans la littérature macroéconomique comportementale. Une première génération de modèles macroéconomiques comportementaux introduisait des rigidités d'information via des règles comportementales *ad-hoc*. Bien que ces règles comportementales étaient parfois supportées par des évidences empiriques ou des experiences en laboratoire, elles ne sont pas robustes à l'un des requis fondamentaux de la modélisation macroéconomique : la critique de Lucas. Dès lors, les conclusions issues de cette première génération de modèles ont été globalement ignorées par le reste des macroéconomistes.

C'est seulement au début du siècle qu'une nouvelle génération de modèles macroéconomiques avec des rigidités d'information endogènes s'est développée. Cette endogénéité est la conséquence du caractère limité de l'attention. En effet, puisque cette dernière est rare, chaque individu fait face à un arbitrage afin de determiner *le moment*, *l'intensité* et *le type* d'information auxquelles il veut être attentif. Tous ces nouveaux modèles, dits de modèles à agents rationnellement inattentifs, cherchent à tenir compte de cet arbitrage. Ils partagent deux caractéristiques essentielles.

Premièrement, ils appréhendent l'allocation de l'attention comme le résultat d'un arbitrage entre couts et bénéfices. Par conséquent, cette nouvelle classe de modèles consi-

[12]Le Chapitre 4 propose une revue de littérature de ces travaux empiriques.

dère que les agents économiques sont rationnellement inattentifs : lorsque l'attention est rare[13], il n'est pas possible, ni intéressant, d'intégrer chaque bit d'information disponible. Par conséquent, ces modèles expliquent pourquoi les agents économiques peuvent être inattentifs, et donc mal informés, dans un monde où l'information est pourtant abondante et peu coûteuse.

Deuxièmement, un individu rationnellement inattentif ajuste constamment son attention dans un environnement économique changeant. Ainsi, ces nouveaux modèles sont, en principe, robustes face à la critique de Lucas.

Grace à ces deux caractéristiques, les macroéconomistes ont été moins réticents à l'idée d'écouter les prédictions de cette nouvelle génération de modèles macroéconomiques traitant des rigidités d'information. Comme nous l'avons déjà vu, ces prédictions se sont avérées importantes pour l'analyse des phénomènes macroéconomiques et largement supportées par des évidences empiriques directes et indirectes. Tous les ingredients étaient ainsi rassemblés pour que les modèles à agents rationnellement inattentifs deviennent un outil majeur de la littérature macroéconomique.

**Déterminants de l'attention**

Il existe une large variété de modèles se concentrant sur l'allocation optimale de l'attention. Gabaix (2019) propose une revue de ces modèles dans la littérature. Les chapitres de cette thèse de doctorat s'appuient sur un modèle intuitif et largement utilisé d'allocation endogène de l'attention avec des actions stochastiques : le modèle d'information bruitée avec précision endogène. Dans ce qui suit, je présente brièvement quelques éléments fondamentaux de cette classe de modèles. Bien que je considère un modèle simple et statique pour mettre en avant des déterminants importants de l'allocation de l'attention dans cette introduction, les intuitions qui seront gagnées dans cette section sont générales et se retrouveraient également avec des spécifications plus complexes (par exemple dans un cadre dynamique, avec différents coûts pour l'attention, avec différentes fonctions d'objectif, etc.).

[13]L'utilisation de l'adjectif rare pour qualifier l'attention dans cette partie peut surprendre. Elle n'est cependant pas anodine et renvoie à la terminologie économique que l'on utilise pour parler d'un bien. De ce fait, l'attention est implicitement considérée comme un bien tout à fait standard ici et l'objectif de l'économiste est donc d'en étudier l'allocation optimale.

Considérons un agent économique qui choisit son action $y$ dans le but de maximiser son utilité. À titre d'illustration, supposons que son action optimale avec des anticipations rationnelles et une information parfaite dépende linéairement d'une variable aléatoire d'état $x \sim \mathcal{N}(0, \sigma_x^2)$. On a donc,

$$y^{FIRE} = ax \tag{1.6}$$

où le paramètre $a$ est l'élasticité de l'action optimale avec des anticipations FIRE par rapport à la variable d'état $x$.

Lorsque cet agent est imparfaitement informé, il ne peut pas directement observer la variable aléatoire d'état $x$ comme il le ferait avec une information parfaite. Il doit, dès lors, former une anticipation concernant cette variable. Afin de former cette anticipation, l'agent dispose de certaines connaissances. Premièrement, il est doté d'une croyance initiale portant sur les réalisations potentielles de la variable aléatoire d'état $x$. Il est généralement supposé que cette croyance initiale est correcte. En d'autres termes, on maintient généralement l'hypothèse d' « anticipations rationnelles » de l'hypothèse FIRE. Dans notre cadre, cela signifie que l'agent sait que la variable aléatoire d'état suit une distribution Gaussienne dont l'espérance est nulle et la variance donnée par $\sigma_x^2$.

Deuxièmement, s'il le souhaite, cet agent peut observer de l'information portant sur la réalisation de la variable d'état. Dans les modèles d'information bruitée, cela signifie qu'il peut observer un signal $s = x + \epsilon$ où $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ est un bruit blanc Gaussien. La précision du signal, donnée par $1/\sigma_\epsilon^2$, est ici une mesure de l'attention de l'agent. Lorsque la précision du signal tend vers l'infini (c-à-d lorsque $\sigma_\epsilon^2 = 0$), le signal $s$ est systématiquement égale à la réalisation de la variable aléatoire d'état $x$. On dit alors que l'agent est parfaitement attentif puisqu'il traite toute l'information utile afin de determiner la réalisation de $x$. Ce cas extrême correspond aux anticipations rationnelles avec information parfaite. À l'inverse, lorsque la précision du signal est plus faible, l'agent est moins attentif et le signal qu'il reçoit moins informatif sur la réalisation de $x$.

*Quels sont alors les déterminants de l'attention de l'agent ?* Afin de fournir une réponse intuitive à cette question, supposons que l'agent cherche à minimiser l'espérance des déviations quadratiques en comparaison de son action optimale avec information parfaite plus un coût d'attention $\mathcal{C}(1/\sigma_\epsilon^2)$ qui augmente avec la précision du signal. Ainsi, pour une précision donnée du signal, il est optimal de choisir l'action suivante[14]

$$y^\star = aE[x|s] \tag{1.7}$$

où $E[x|s]$ est l'espérance mathématique de $x$ étant donné le signal $s$. Par conséquent, l'agent choisit la précision du signal ou, de façon équivalente, la variance du bruit $\sigma_\epsilon^2$ telle que[15]

$$\min_{\sigma_\epsilon^2 \geq 0} \quad a^2 \frac{\sigma_\epsilon^2 \sigma_x^2}{\sigma_\epsilon^2 + \sigma_x^2} + \mathcal{C}(1/\sigma_\epsilon^2) \tag{1.8}$$

où le ratio $\sigma_\epsilon^2 \sigma_x^2/(\sigma_\epsilon^2 + \sigma_x^2)$ est l'incertitude *a posteriori*, c'est-à-dire l'incertitude moyenne concernant $x$ qui reste après avoir observé un signal $s$. En utilisant la condition du premier ordre, on peut montrer que la précision optimale du signal est solution de

$$\frac{1}{\sigma_\varepsilon^2} \quad = \quad \frac{a}{\sqrt{\mathcal{C}'(\cdot)}} - \frac{1}{\sigma_x^2} \tag{1.9}$$

Lorsqu'une solution existe, l'attention dépend ainsi de trois paramètres dans ce cadre simple :

- Elle augmente avec l'élasticité de l'action optimale avec FIRE ($a$).
- Elle augmente dans des environnements incertains ($\sigma_x^2$).
- Elle baisse avec le coût marginal de l'attention ($\mathcal{C}'(\cdot)$).

Les évidences empiriques confirment que ces trois facteurs sont des déterminants importants de l'allocation de l'attention[16]. Il est essentiel de remarquer que l'attention ne dépend ici que de paramètres profonds, c'est-à-dire de facteurs exogènes et invariants dans

[14] L'espérance conditionnelle est optimale au sens de la *MSE* (erreur quadratique moyenne).

[15] Voir Mackowiak et al. (2018) section 2.1 pour la dérivation de cette expression.

[16] Voir les revues de littérature dans Gabaix (2019) et les chapitres suivants de cette thèse.

le temps qui décrivent la structure fondamentale de l'économie modélisée. En particulier, ces paramètres ne sont pas affectés par les chocs économiques.

### 1.2.2 Défis actuels dans la littérature

Cette partie présente quelques défis liés à l'inattention et ses implications en économie. Les chapitres suivants de cette thèse de doctorat sont des contributions originales dans ces programmes de recherche.

**Déterminants de l'attention : la critique de Lucas contre-attaque**

Les études empiriques indiquent que d'autres facteurs affectent directement l'allocation de l'attention. Par exemple, on observe des variations notables de l'attention qui sont corrélées aux fluctuations du cycle économique. Coibion and Gorodnichenko (2015) et le Chapitre 4 de cette thèse montrent que l'attention des individus augmente durablement lorsque l'économie entre dans une récession. En outre, des expériences récentes identifient une relation causale entre le contenu de l'information et l'attention qui est portée à cette information (Abe and Ueno, 2015; Armantier et al., 2016; Khaw et al., 2017).

Cette dépendance (à certaines variables) d'état ou (au contenu) d'information est difficilement réconciliable avec les modèles standards d'allocation de l'attention. Cela se retrouve dans la condition du premier ordre[17] (1.9). En effet, elle prédit que l'attention dépend seulement de paramètres profonds, c'est-à-dire ces paramètres qui gouvernent la structure de l'économie et sont invariants dans le temps en l'absence de rupture structurelle non-anticipée. Ainsi, toute variation dans le temps de l'attention est prédéterminée dans ce modèle et n'est pas liée aux variations des variables d'état (la variable $x$ dans le modèle).

Pourtant, tenir compte de la dépendance d'état ou d'information que l'on observe dans les données est essentiel puisqu'elle affecte directement les réponses des agents aux changements de l'environnement économique. En effet, l'espérance conditionnelle mathématique $E[x|s]$ dans l'équation (1.7) dépend du niveau d'attention : plus un agent est attentif, plus

[17] Le modèle statique présenté dans l'introduction reste utile à cet égard et continue de fournir les intuitions essentielles que l'on obtiendrait dans un modèle dynamique.

il s'adosse sur le signal qu'il reçoit pour former son anticipation. Plus spécifiquement, nous avons[18]

$$E[x|s] = \xi s = \xi(x + \varepsilon) \tag{1.10}$$

où $\xi = \sigma_x^2/(\sigma_x^2 + \sigma_\varepsilon^2) \in [0, 1]$ est le poids que l'agent met sur le signal qu'il reçoit et qui est une mesure de son attention. En ne permettant pas de reproduire la dépendance de ce poids $\xi$ avec $s$ et/ou $x$, la plupart des modèles d'attention endogène[19] reposent implicitement sur des formes réduites pour modéliser les anticipations qui ne tiennent que partiellement compte de l'impact des changements de l'environnement économique. Cette remarque est, en substance, similaire à la critique que Lucas a levée à l'encontre des modèles Keynésiens. Ainsi, un récent programme de recherche a pour objectif de tenir compte et de rationaliser la dépendance d'état de l'attention que l'on observe dans les données. Il devrait améliorer la capacité des modèles macroéconomiques à agents rationnellement inattentifs à fournir des recommandations politiques plus satisfaisantes.

*Contributions connexes : Chapitres 2 et 4.*

**Interactions entre attention et les dynamiques des agrégats économiques.**

La littérature soutient que l'inattention est une source d'inertie des fluctuations du cycle économique (Sims, 2003; Mankiw and Reis, 2006; Maćkowiak and Wiederholt, 2015). En parallèle, les évidences empiriques suggèrent que l'inattention n'est pas constante dans le temps et corrélée aux fluctuations du cycle (Coibion and Gorodnichenko, 2015). Un défi majeur pour les économistes est ainsi de comprendre comment l'inattention et les fluctuations économiques interagissent.

Le Chapitre 2 montre que la dynamique de la croissance de la consommation est une illustration flagrante de cette interaction. La persistance de la croissance de la consommation agrégée a baissé drastiquement durant la récession de 2008 (Kumar and Jia, 2019). Ce changement soudain de la dynamique de la consommation est surprenant. Il est éga-

[18]Lorsque le prior et le signal sont Gaussiens, l'espérance conditionnelle est une combinaison linéaire de le moyenne du prior (qui est nulle ici) et du signal.

[19]Voir la revue de littérature dans le chapitre 2.

lement malheureux puisque la persistence de la consommation joue normalement un rôle stabilisateur lorsque l'économie entre en récession. Bien que ce changement ait été particulièrement brutal durant la récession de 2008, Kumar and Jia (2019) identifient une baisse de la persistence de la consommation agrégée durant quasiment tous les épisodes de recession aux U.S. depuis 1980. Dans le Chapitre 2, je soutiens que cette baisse systématique est la conséquence de l'attention des consommateurs. L'argument est le suivant. En tenant compte de la dépendance d'information de l'attention que nous avons précédemment mentionnée, les consommateurs deviennent mécaniquement plus attentifs en moyenne quand l'économie entre en récession. Par ailleurs, lorsqu'ils sont plus attentifs, les consommateurs répondent plus rapidement aux chocs économiques et la persistance de la consommation agrégée est plus faible. Ainsi, la hausse de l'attention des consommateurs durant les recessions cause la baisse de la persistence de la consommation durant ces périodes.

Le Chapitre 2 est l'un des rares articles qui se concentre sur l'*interaction* entre l'attention des agents et les dynamiques des agrégats économiques. De façon importante, il montre que cette interaction est importante pour l'étude des phénomènes macroéconomiques.

*Contributions connexes : Chapitres 2 et 4.*

**Les politiques économiques façonnent l'attention**

Les caractéristiques des politiques économiques affectent l'attention que les agents portent à ces politiques et, par conséquent, leurs réponses face à ces politiques. Ci-dessous, je présente trois exemples, traités plus précisément dans les chapitres suivants de cette thèse, qui illustrent ce mécanisme.

Les transferts de revenu sont une illustration manifeste d'une situation où les caractéristiques d'une politique économique façonnent l'attention que les consommateurs portent à cette politique. Soit un gouvernement qui annonce de façon inattendue un transfert forfaitaire dans un futur proche. D'après l'hypothèse du revenu permanent, un consommateur doit réagir immédiatement suite à cette annonce et ajuster sa consommation en conséquence[20], cela indépendamment de la taille du transfert qui est annoncé. Cependant, dans les données, on observe que ces ajustements de consommation sont plus probables lorsque le transfert qui est annoncé est large (Browning and Collado, 2001; Scholnick, 2013; Kueng, 2018). Le Chapitre 2 démontre que cette dépendance de la réponse de consommation par rapport à la taille du transfert peut être la conséquence de l'attention des consommateurs. Ceci est dû au fait que les consommateurs ont plus de chances d'être attentifs aux transferts importants et inattentifs aux transferts plus petits.

Au niveau agrégé, ce même mécanisme implique que l'attention des consommateurs interagit de façon non linaire et non triviale avec les politiques économiques de stabilisation. Durant les récessions, les consommateurs sont généralement moins attentifs aux bonnes nouvelles. Par conséquent, le canal de consommation est atténué durant ces périodes et les politiques monétaires et fiscales expansionnistes moins efficaces. Une façon de réactiver ce canal de consommation est de mettre en place des politiques expansionnistes agressives qui vont attirer l'attention des consommateurs.

Dans la même veine, le Chapitre 3 montre que lorsque le gouvernement fixe une taxe proportionnelle sur le revenu, il fait face à de nouveaux arbitrages directement liés à l'attention endogène des contribuables. Par exemple, lorsque le gouvernement décide d'augmenter le taux de taxation, les contribuables deviennent plus attentif à ce taux. Ainsi, les hausses de taxe rendent les taxes plus saillantes et les contribuables plus réactifs à ces dernières. En fin de compte, cette endogénéité de l'attention limite la capacité du gouvernement à tirer parti de l'inattention des contribuables.

Dans chacun des ces exemples, nous remarquons que les caractéristiques d'une politique économique (ici la taille de l'intervention) façonnent l'attention que les agents portent à cette politique et, par là même, la façon dont ils y réagissent. Dans l'ensemble, ces exemples illustrent un défi majeur pour les économistes : l'endogénéité de l'attention génère des interactions non triviales entre la conception des politiques économiques et les perceptions et anticipations qu'en ont les agents. Il est essentiel d'accroître nos connaissances sur cette interaction afin de fournir des recommandations utiles portant sur la mise

[20] En l'absence de frictions de marché l'empêchant.

en œuvre des politiques économique.

*Contributions connexes : Chapitres 2 et 3.*

**L'inattention modifie la conduite des politiques économiques**

Un autre défi auquel les économistes sont confrontés est de comprendre comment l'inattention des agents affecte la conduite des politiques économiques. L'idée selon laquelle les décideurs politiques peuvent tirer parti de l'attention limitée des individus, de sorte à améliorer le bien-être social, émerge, notamment, en finances publiques comportementales et en économie politique[21]. Néanmoins, on sait peu de choses quant au comportement d'un gouvernement qui a le pouvoir de tirer parti de l'inattention des agents. L'analyse de ces questions requière de développer de nouveaux cadres pour l'analyse des politiques économiques. Ces derniers doivent recourir à la théorie des jeux afin d'analyser les comportements d'équilibre dans des jeux où les décideurs politiques peuvent bénéficier des perceptions erronées des agents.

Le Chapitre 3 appartient à ce programme de recherche et montre que l'inattention fait ressurgir d'anciennes questions portant sur la discrétion des politiques économiques dans des situations où l'on ne s'y attendait pas. En effet, lorsque les agents sont partiellement inattentifs aux politiques économiques, ils ancrent leurs anticipations sur leurs convictions antérieures et, par conséquent, sous-réagissent aux changements réels des politiques. Un gouvernement discrétionnaire est donc incité à tirer parti de ces sous-réactions. Cependant, à l'équilibre, les convictions des agents s'ajustent afin de prendre en compte la volonté du gouvernement de dévier. À cause de cet ajustement d'équilibre, les politiques économiques d'équilibre s'avèrent sous-optimales dans un jeu où un gouvernement discrétionnaire a le pouvoir de bénéficier de l'inattention des agents.

*Contributions connexes : Chapitre 3.*

[21] Voir la revue de littérature dans le Chapitre 3.

### 1.2.3 Plan de la thèse

Cette thèse de doctorat est une collection de trois articles de recherche indépendants. Ces articles étudient les implications économiques des rigidités d'information. Ils sont reportés dans les chapitres 2 à 4 de cette thèse. Chaque chapitre commence par une introduction qui discute la littérature pertinente et présente les contributions du chapitre en détail. Ces chapitres sont rédigés en anglais. Dans ce plan de thèse, je résume brièvement le contenu de chaque chapitre.

**Chapitre 2 – Traitement coûteux de l'information et dynamique de la consommation**

Le Chapitre 2 est une contribution théorique à l'analyse de la dynamique de la consommation. Il est motivé par quatre faits empiriques déconcertants, car difficilement conciliables avec les théories existantes de consommateurs rationnellement inattentifs. Premièrement, l'attention des consommateurs dépend du contenu de l'information. En se basant sur des expériences sur le traitement de l'information, Abe and Ueno (2015), Armantier et al. (2016) et Khaw et al. (2017) constatent que les individus sont plus susceptibles d'incorporer de nouvelles informations lorsqu'elles contredisent leurs croyances initiales. Deuxièmement, les rigidités d'information ne sont pas constantes dans le temps. Dräger and Lamla (2012) et Coibion and Gorodnichenko (2015) trouvent que les rigidités d'information chutent durablement lors des récessions. Troisièmement, les consommateurs réagissent uniquement aux chocs de revenu anticipés qui sont importants et négligent ceux qui sont faibles. Jappelli and Pistaferri (2010) nomment ce mécanisme l'*hypothèse de magnitude* et les preuves empiriques en faveur de cette hypothèse s'accumulent dans la littérature empirique (Browning and Collado, 2001; Scholnick, 2013; Kueng, 2018). Quatrièmement, la dynamique de la consommation agrégée change au cours du cycle économique. En particulier, Caballero (1995) montre que « durant les périodes d'expansion économique, les consommateurs réagissent plus rapidement aux chocs de richesse positifs qu'aux négatifs, tandis que l'inverse est vrai durant les périodes de compression » et Kumar and Jia (2019) reportent des diminutions systématiques de la persistance de la croissance de la consommation pendant les récessions.

Ce chapitre propose un modèle d'allocation de l'attention des consommateurs qui peut expliquer ces quatre faits empiriques. Dans le modèle, il existe une source qui transmet des informations bruitées sur les chocs de revenu permanent. Les consommateurs doivent s'acquitter d'un coût fixe pour être attentif à cette information et leur stratégie d'attention, c'est-à-dire le choix d'être attentif ou non, peut être une fonction de l'information sous-jacente. Mis à part cette nouvelle structure d'information, le problème de consommation et d'épargne considéré dans ce chapitre est aussi transparent que possible et coïncide avec le modèle de marche aléatoire de Hall (1978) avec une utilité quadratique. Ainsi, les prédictions du modèle ne dépendent pas de l'une des multiples extensions de la théorie standard de la consommation.

Les conclusions principales du chapitre sont les suivantes. Premièrement, l'attention des consommateurs dépend bien du contenu de l'information. À l'optimum, le consommateur est caractérisé par une zone d'inattention dans laquelle il ne prête pas attention aux nouvelles informations et n'ajuste pas son plan de consommation. Ce n'est seulement qu'une fois que les évidences à l'encontre de ses croyances initiales se sont suffisamment accumulées, qu'il devient attentif à celles-ci et révise son plan de consommation en conséquence.

Deuxièmement, les rigidités d'information varient dans le temps. Les consommateurs adoptent une stratégie d'attention qui dépend du contenu de l'information. Par conséquent, en partant de la distribution stationnaire des consommateurs, un choc de revenu agrégé incite davantage de consommateurs à être attentifs. L'information étant bruitée, l'impact d'un choc agrégé se diffuse lentement dans l'économie et l'augmentation de la part des consommateurs attentifs est persistante.

Troisièmement, le modèle prédit une corrélation positive entre la taille d'un choc de revenu et la propension marginale à consommer. Il fournit ainsi une justification à l'*hypothèse de magnitude*. L'intuition est la suivante. Pour un consommateur qui a été attentif à la période précédente, il est peu probable qu'un petit choc de revenu permanent (en valeur absolue) soit suffisamment important pour attirer son attention. Cependant, à mesure que la taille du choc augmente, le consommateur est plus susceptible d'être attentif à ce choc et de réviser son plan de consommation. Plus généralement, la propension marginale à consommer moyenne à la suite d'un choc sur le revenu permanent dépend à la fois des chocs passés et présents.

Quatrièmement, la dynamique de la consommation agrégée varie avec le cycle économique. Le modèle prédit que pendant les crises économiques (respectivement les booms économiques), un choc positif (resp. négatif) réduit généralement la part des consommateurs attentifs et génère une réponse plus douce de la consommation agrégée, tandis qu'un choc négatif (resp. positif) génère une réponse plus brutale. De plus, la persistance de la croissance de la consommation agrégée dépend de la proportion endogène, qui varie au cours du cycle économique, de consommateurs attentifs. En temps normal, les rigidités d'information sont proches de leur niveau stationnaire et la persistance de la consommation est relativement forte et constante. Cependant, pendant des périodes inhabituelles, comme les récessions, les rigidités de l'information diminuent beaucoup, réduisant ainsi également la persistance de la consommation agrégée.

Ce chapitre s'écarte donc de la littérature existante en ce qu'il saisit certaines caractéristiques non linéaires des frictions informationnelles. Ces non-linéarités cachent des implications importantes pour la propagation des chocs économiques dans les modèles macroéconomiques. Elles suggèrent que les frictions d'information sont importantes en temps normal, entraînant ainsi des réponses lentes, en forme de bosse, pour les agrégats macroéconomiques. Cependant, ces frictions disparaissent partiellement à mesure que l'économie s'éloigne de son état stationnaire. Dans ces dernières situations, la dynamique des agrégats macroéconomiques est beaucoup moins lente et se rapproche de celle prédite par des économies sans friction. Par conséquent, les frictions d'information agissent comme un mécanisme de stabilisation vers la dynamique d'état stationnaire. Ces nouvelles conclusions ont également des implications importantes pour l'efficacité des politiques monétaires et budgétaires pendant les récessions.

En résumé, le chapitre 2 fournit une explication aux quatre faits empiriques susmentionnés indiquant, respectivement, que l'attention des consommateurs dépend du contenu de l'information, que les rigidités d'information varient dans le temps, que les consommateurs ne réagissent qu'aux chocs de revenu anticipés qui sont importants et que la dynamique de la consommation agrégée varie au cours du cycle économique. Ce faisant, il contribue directement aux trois premiers programmes de recherche mentionnés dans la section 1.2.2.

**Chapitre 3 – Inattention et le biais de taxation**

Le Chapitre 3, issu d'une collaboration avec Antoine Ferey, étudie la façon dont les rigidités d'information dans la perception des taxes affectent la conduite des politiques fiscales. Un nombre croissant d'études empiriques documentent des frictions importantes dans la perception des taxes par les contribuables (Chetty, 2015; Bernheim and Taubinsky, 2018; Stantcheva, 2019). À la lumière de ces études, une littérature normative florissante cherche à identifier les politiques fiscales optimales en présence de telles frictions d'information. Cette littérature caractérise les politiques fiscales optimales en termes de statistiques suffisantes. Ces dernières capturent l'élasticité du revenu des contribuables à la taxe et les biais de perception à l'optimum. En procédant ainsi, cette littérature ignore les mécanismes liés à l'ajustement des perceptions et en méprise les potentielles conséquences.

Pourtant, dans la pratique, la politique fiscale est très probablement influencée par la manière dont les perceptions des contribuables s'ajustent suite à une variation du taux de taxation. Par exemple, les décideurs politiques seront plus enclin à augmenter les impôts si les agents sont inattentifs et ne perçoivent qu'une fraction des augmentations d'impôt. Ces implications positives des frictions d'information pour la politique fiscale n'ont jusqu'alors pas été explorées, contrairement aux implications normatives. Le Chapitre 3 a pour objectif de combler ce manque en étudiant la façon dont les rigidités d'information dans la perception des taxes affectent la conduite des politiques fiscales.

Le Chapitre 3 propose une théorie positive de la politique de taxation du revenu dans un cadre où l'offre de travail des agents est déterminée par leur perception de la politique fiscale. Il démontre que l'ajustement des perceptions de la taxe sur le revenu interagit avec la conduite de la politique fiscale et altère cette dernière. Plus précisément, l'inattention des contribuables conduit un gouvernement discrétionnaire à appliquer des taux de taxation inefficacement élevés : c'est le biais de taxation qui résulte de l'inattention.

Ce résultat est la conséquence d'une dichotomie entre les ajustements directs et indirects des perceptions à la suite d'une réforme fiscale. En effet, en présence de frictions d'information, les perceptions des contribuables sont une moyenne pondérée du vrai taux de taxation et de la croyance *a priori* que les contribuables ont de ce taux. Il y a donc deux marges à travers lesquelles les perceptions s'ajustent : une marge *directe* qui capture l'attention que les contribuables portent à l'observation continue du taux de taxation et, ainsi, aux réformes de ce dernier, et une marge *indirecte* qui capture l'ajustement des croyances *a priori* des contribuables. Pour des croyances *a priori* fixées, les contribuables inattentifs ne perçoivent qu'une fraction du changement du taux de taxation, ce qui limite leur réponse en termes d'offre de travail. Un gouvernement discrétionnaire cible alors un taux de taxation plus élevé que celui qu'il aurait choisi si les agents étaient parfaitement attentifs. À l'équilibre, les croyances *a priori* des contribuables doivent nécessairement être cohérentes avec la politique fiscale choisie par le gouvernement. Par conséquent, la réponse de l'offre de travail est *ex post* plus forte que ce que le gouvernement avait perçu *ex ante*. Le taux de taxation mis en place par un gouvernement discrétionnaire est ainsi trop élevé parce que ce dernier n'arrive pas à prendre en compte l'ajustement indirect des croyances *a priori* (qui est le résultat d'un mécanisme d'équilibre) dans son choix de politique fiscale. En bref, l'inattention des contribuables crée l'illusion que les réformes fiscales induisent des coûts d'efficience inférieurs à ce qu'ils sont réellement, incitant ainsi le gouvernement à dévier de la politique qui serait optimale d'un point de vue normatif. Fondamentalement, cette inefficacité à l'équilibre reflète un problème d'engagement[22] de la part du gouvernement.

Nous cherchons ensuite à illustrer les implications de cette distorsion politique et à quantifier son ampleur. Dans un premier temps, nous calibrons un modèle linéaire de taxation du revenu pour les U.S. En nous appuyant également sur la littérature empirique existante pour calibrer nos statistiques suffisantes, nous estimons que le biais de taxation est approximativement égal à $3,7$ points de pourcentage. Cela signifie que le taux (linéaire) de taxation sur le revenu aux U.S. est au moins 12% plus élevé que ce qu'il devrait être en gardant l'objectif redistributif du gouvernement constant : le biais de taxation est considérable.

[22] Pour reprendre le terme anglais de *commitment* que l'on met en opposition à celui de *discretion* pour qualifier le caractère discrétionnaire des politiques économiques.

Dans un second temps, nous étendons notre analyse aux systèmes de taxation non linéaires. Lorsque la taxe est non linéaire, l'incitation du gouvernement à augmenter le taux d'imposition marginal à un niveau de revenu donné dépend de l'attention des agents à (ou proches de) ce niveau de revenu. Par conséquent, la corrélation positive entre le revenu et l'attention se traduit par un biais de taxation qui décroit globalement avec le revenu : le biais de taxation est important pour les revenus faibles et quasiment inexistant pour les revenus les plus élevés. Le biais de taxation atténue ainsi la forme en U des taux marginaux (Saez, 2001) et réduit la progressivité du barème d'imposition sur le revenu.

En résumé, le Chapitre 3 développe une théorie positive de la politique fiscale et montre que l'inattention des agents interagit avec la conduite de cette politique et incite le gouvernement à appliquer des taux d'imposition inefficacement élevés. Ce biais de taxation est indésirable, considérable et régressif. Ce faisant, ce chapitre contribue directement aux deux derniers programmes de recherche mentionnés dans la section 1.2.2.

**Chapitre 4 – Inertie des prévisions à travers le cycle économique**

L'excès d'inertie des anticipations des agents économiques, qui résulte de frictions informationnelles, est une cause majeure de la persistance des variables agrégées et affecte la transmission des politiques monétaires et budgétaires. C'est donc un thème central de la macroéconomie. Cependant, les économistes doivent encore s'accorder quant à la ou les sources de ces frictions, quant au fait qu'elles soient une caractéristique des données individuelles, quant à leur évolution dans le temps et quant à leur contribution respective à l'excès d'inertie des anticipations.

Le chapitre 4 fournit une contribution empirique à chacune de ces questions importantes. Il estime un processus de formation des prévisions à deux marges qui tient compte d'un potentiel arrondissement des prévisions. Les prévisionnistes décident quand réviser leurs prévisions (marge extensive, par exemple, les modèles d'information sporadique). Lorsqu'ils le font, ils incorporent lentement les nouvelles informations (marge intensive, par exemple, les modèles d'information bruitée) et peuvent déclarer une valeur arrondie pour leur nouvelle prévision (arrondi, par exemple, modèles traitant des ensembles de considération). La prise en compte simultanée de ces trois formes de rigidités permet d'englober une large classe de frictions potentielles dans la formation des prévisions et d'évaluer leur contribution respective.

Les principales contributions du chapitre sont les suivantes. Premièrement, la littérature existante mesure les rigidités à la marge extensive à partir de la part observée des révisions de prévision qui sont nulles. Le chapitre développe un test simple pour la validité de cette mesure lorsque des rigidités existent à la fois aux marges extensives et intensives. En appliquant ce test aux prévisions individuelles d'inflation, de croissance du PIB et du taux de chômage issues de l'enquête de la BCE auprès des prévisionnistes professionnels, les données rejettent clairement la validité de cette mesure.

Deuxièmement, le chapitre étend le modèle de prévision à deux marges afin d'également tenir compte que les prévisionnistes peuvent reportés des valeurs arrondies pour leurs prévisions. Il estime ce modèle sur des données de prévisions individuelles et constate que les trois formes de rigidités coexistent. Environ 80 % des révisions nulles qui étaient auparavant attribuées aux rigidités à la marge extensive sont en fait la conséquence de l'arrondissement des prévisions. Les rigidités à la marge extensive sont donc beaucoup plus faibles que celles précédemment reportées : environ 95 % des prévisionnistes européens révisent leur prévision chaque trimestre. Surtout, cette nouvelle mesure des rigidités à la marge extensive n'est plus rejetée par les données. Les rigidités à la marge intensive sont importantes : les prévisionnistes n'intègrent que 55 à 60 % des informations qu'ils reçoivent chaque trimestre.

Troisièmement, le chapitre évalue la contribution relative de chaque forme de rigidité. Dans les modèles macroéconomiques avec agents représentatifs, les frictions d'information affectent généralement les agrégats macroéconomiques via leur impact sur l'inertie excessive des anticipations moyennes. Par conséquent, le chapitre montre comment les trois formes de rigidités estimées à partir des données individuelles se retrouvent dans un paramètre unique qui mesure l'inertie excessive de la prévision moyenne. En décomposant ainsi l'inertie excessive moyenne, on constate qu'elle est presque exclusivement la conséquence des rigidités à la marge intensive. Les rigidités à la marge extensive ne représentent que 4 à 5 % de cet excès d'inertie et l'arrondi moins de 0,04 %.

Quatrièmement, le chapitre estime des séries chronologiques trimestrielles pour l'évolution de l'excès d'inertie des prévisions. La tendance globale de l'acquisition d'informations est restée relativement constante sur la période 1999-2019 et il n'y a pas de changement structurel soudain sur cette période. On observe néanmoins des variations à moyen terme communes aux trois séries (inflation, croissance du PIB et chômage). L'acquisition d'informations a lentement diminué au cours de la période 2000-2008, période de relative stabilité économique dans la zone euro. Elle a commencé à augmenter par la suite et est restée stable entre les deux dernières récessions. Depuis 2016, la tendance de l'acquisition d'information semble à nouveau sur une trajectoire croissante. Notamment, l'acquisition d'information est devenue plus volatile depuis la crise de 2008. Concernant la contribution de chaque forme de rigidités à ces variations dans le temps, nous arrivons à la même conclusion que pour la moyenne : les variations des rigidités de l'information sont presque exclusivement la conséquence de la marge intensive.

Cinquièmement, le chapitre analyse la dépendance d'état des frictions d'information en étudiant sa dynamique pendant les récessions. Comme dans les études précédentes (Coibion and Gorodnichenko, 2015), nous observons une diminution globale des frictions d'information au lendemain d'une récession. Cependant, la dynamique est différente pour chacune des trois séries chronologiques. Plus précisément, nous constatons une augmentation des rigidités d'information pour l'inflation et la croissance du PIB mais une forte diminution des frictions de l'information pour le chômage.

En résumé, le chapitre 4 propose une nouvelle méthodologie afin d'estimer simultanément une grande classe de rigidités d'information sur des données de prévisions individuelles. Cette méthodologie est ensuite utilisée pour montrer que si de nombreuses formes de rigidités sont en effet une caractéristique des données, seule l'inattention aux nouvelles informations lors de la révision d'une prévision est pertinente pour les modèles macroéconomiques avec agents représentatifs. Il dérive ensuite des séries chronologiques trimestrielles pour les frictions d'information afin d'en analyser les variation temporelles. Ce faisant, ce chapitre contribue directement aux deux premiers programmes de recherche mentionnés dans la section 1.1.2.

# Chapter 2

# Costly Information Processing and Consumption Dynamics

**Abstract:** This paper studies the consumption-saving problem of a consumer who faces a fixed cost for paying attention to noisy information and whose attention strategy, i.e., whether or not she pays attention, can be a function of the underlying information. At the optimum, consumers chose to be attentive when evidence accumulates far from their prior beliefs. The model provides an explanation for four puzzling empirical findings on consumption and expectations. First, consumers' attention depends on the information content. Second, aggregate information rigidities vary over the business cycle. Third, consumers only react to large anticipated shocks and neglect the impact of small ones. Fourth, aggregate consumption dynamics vary over the business cycle.[23]

Keywords: Consumption dynamics, information frictions and inattentive consumers.

JEL Classification: E21, E70, C61.

[23] I am extremely grateful to Xavier Ragot, Philippe Andrade, Olivier Compte, Lena Drager, Antoine Ferey, Gaetano Gaballo, Cars Hommes, Ricardo Reis, Mirko Wiederholt, anonymous referees and participants at the CEF and EEA annual meetings, the Second Behavioral Macroeconomics Workshop, and seminars at the University of Amsterdam, the Paris School of Economics, the banks of France and Italy, and the University of Bologna, for helpful comments and suggestions. The paper was awarded at the 2016 CEF Graduate Student Paper Contest.

## 2.1 Introduction

Theories of consumption play an important role in business cycle models and recently many authors have proposed changes to the modeling of consumption in DSGE models. In Heterogeneous Agents New Keynesian (HANK) models, hand-to-mouth consumers, precautionary savings, and heterogeneity of marginal propensities to consume shape the response of aggregate consumption to shocks (Jappelli and Pistaferri, 2017). In models with information frictions on the household side, the slow updating of beliefs over time explains the persistence of consumption growth (Reis, 2006a; Carroll et al., 2020) and generates hump-shaped responses of aggregate consumption to shocks (Mankiw and Reis, 2006; Maćkowiak and Wiederholt, 2015).

However, several empirical findings on consumption and expectations are not simultaneously explained by existing theories. First, consumers' attention depends on the information content. Using information experiments, Armantier et al. (2016), Abe and Ueno (2015) and Khaw et al. (2017) find that individuals are more likely to incorporate new information when it contradicts their prior beliefs. Second, aggregate information rigidities are not constant over time. They drop persistently in the aftermath of a recession (Dräger and Lamla, 2012; Coibion and Gorodnichenko, 2015) and large unexpected shocks (Baker et al., 2020). Third, consumers only react to large anticipated shocks and neglect the impact of small ones. Jappelli and Pistaferri (2010) term this mechanism the *magnitude hypothesis* and empirical evidence in favor of the hypothesis is accumulating (Browning and Collado, 2001; Scholnick, 2013; Kueng, 2018).[24] Fourth, aggregate consumption dynamics vary over the business cycle. Caballero (1995) finds that "in good times, consumers respond more promptly to positive than to negative wealth shocks, while the opposite is true in bad times" and Kumar and Jia (2019) report systematic decreases in consumption growth persistence during recessions.

This paper proposes a model that can match these four empirical findings, while

[24]Each of these papers controlled for the presence of credit constraints and concluded that it cannot explain the observed features of excess sensitivity. A priori, the magnitude hypothesis could arise as a consequence of consumption adjustment costs. In these models, consumption adjusts sporadically based on a state-dependent rule. Accordingly, Chetty and Szeidl (2016) demonstrate that in the presence of commitment costs, excess sensitivity and smoothness vanish for large shocks. However, empirical studies on the magnitude hypothesis have mainly focused on non-durable consumption goods, that is, those consumption goods which are the less likely to exhibit commitment costs.

preserving the aforementioned prediction of information friction models regarding the sluggish response of aggregate consumption to shocks. In the model, there is a source which conveys noisy information about shocks to one's permanent income. Consumers face a fixed cost for paying attention to this information and their attention strategy, i.e., whether or not an agent pays attention, can be a function of the underlying information. Apart from this novel information structure, the consumption-saving problem considered in the paper is left as transparent as possible and coincides with Hall's (1978) random walk model with quadratic utility. The model predictions are thus isolated from other refinements of the textbook consumption theory.

There are two significant departures from Reis (2006a) micro-founded model of inattentive consumers. First, consumers are not restricted to time-dependent attention strategies and *can*, for instance, prefer an information-dependent attention strategies. Second, they do not necessarily access perfect information when attentive. The first deviation allows to match the aforementioned empirical evidence that consumers do not update on a purely time-dependent basis while, the second captures the heterogeneity of expectations among attentive consumers (e.g. Armantier et al. (2016)). Consequently, the definition of attentiveness slightly differs from models of Sticky information (Gabaix and Laibson, 2001; Mankiw and Reis, 2002; Carroll, 2003; Reis, 2006a). In the latter, consumers access full-information rational expectations when attentive. In the present paper, attentive consumers observe imperfect information in the form of noisy signals that they rationally filter. They thus never reach full-information rational expectations.[25]

The main findings of the paper are as follows. First, consumers adopt an information-dependent attention strategy. At the optimum, the consumer faces an inattention region where she disregards new information and does not adjust her consumption plan. It is only when evidence against her prior beliefs accumulates that the consumer is willing to pay attention to new information releases and revises her consumption plan accordingly.

Second, information rigidities are time-varying and state-dependent. Because of their attention histories, consumers are ex post heterogenous. Starting from the stationary

[25] This paper thus builds the micro-foundations of the Noisy and Sticky information model evidenced in the data (Andrade and Le Bihan, 2013).

cross-sectional distribution of consumers, an aggregate income shock shifts the cross-sectional distribution and prompts more consumers out of their inattention region.[26] This increase in aggregate attention is persistent. Indeed, consumers' information being noisy (even when they are attentive), the impact of an aggregate shock disseminates slowly in the economy and the increase in the share of attentive consumers is persistent.

Third, the model predicts a positive correlation between the size of an income shock and the marginal propensity to consume. It thus provides a rationale to explain *the magnitude hypothesis*. The intuition is as follows. For a consumer who was last attentive one period ago, a small permanent income shock (in absolute value) is unlikely to be significant enough to trigger her attention. However, as the size of the shock gets larger, the shock becomes more likely to trigger the consumer's attention and to prompt her to revise her consumption path. More generally, the expected marginal propensity to consume out of a shock to permanent income is found to be both history-dependent and shock-dependent.

Fourth, aggregate consumption dynamics vary over the business cycle. The paper focuses on two aspects of aggregate consumption dynamics. First, the impulse response to aggregate income shocks. It finds that the latter depends on the state of the economy, and the magnitude and sign of aggregate shocks. It predicts that during economic busts (respectively booms) – that is periods when lagged consumption growth was below (resp. above) its steady state level – a positive (resp. negative) shock generally lowers the share of attentive consumers and generates a smoother response, while a negative (resp. positive) shock generates a sharper response. As the magnitude of the aggregate shock increases, the interaction between the sign and state of the economy becomes less relevant. Second, it analyzes the variation of the persistence of consumption growth along business cycles. It shows that the persistence of aggregate consumption growth depends on the endogenously time-varying share of attentive consumers. In normal times, information rigidities are

[26]Consumers are ex-post heterogenous because of their idiosyncratic income shocks, information noises and attention histories. To apprehend this multi-dimensional heterogeneity, I show that the cross-sectional distribution of consumers may be characterized by a function of the information that each consumer disregards when being inattentive. Starting from the stationary cross-sectional distribution, an aggregate income shock shifts the distribution of consumers and increases the average information that consumers would disregard by remaining inattentive. Thereby, more consumers are willing to pay attention and the share of attentive consumers increases.

near their steady state level and consumption growth persistence high. However, during unusual times such as recessions, information rigidities decrease a lot and so does aggregate consumption growth persistence.

In sum, the model provides an explanation to the aforementioned four facts indicating that consumers' attention depends on the information content, information rigidities are time-varying, consumers only react to large anticipated shocks, and aggregate consumption dynamics vary over the business cycle.

The model therefore deviates from existing ones in that it captures some non linear features of information frictions. These non linearities conceal important implications for the propagation of economic shocks in macroeconomic models. They suggest that information frictions are large in normal times, thus resulting in sluggish hump-shaped responses of macroeconomic outcomes. However, these frictions partially vanish as the economy gets farer away from its steady state, and so does the persistence in the dynamics of macroeconomic outputs. In the latter situations, the dynamics of macroeconomic outcomes are much less sluggish and gets closer to that predicted by frictionless economies.[27] Therefore, information frictions also act as a stabilizing mechanism toward steady state dynamics. These novel findings also hold significant implications for the effectiveness of monetary and fiscal policies during recessions.[28]

The consumption theory proposed in this paper also has some other attractive properties. First, in models of consumption adjustment costs (Caballero, 1993; Chetty and Szeidl, 2016) and near-rationality (Caballero, 1995; Kueng, 2018), household consumption is constant between adjustments and consumption adjustments are large; while here consumption is not constant whilst the consumer is inattentive and the consumption

[27]Because consumers never reach full-information, the model predicts that, in general, the economy is not frictionless at the limit. It nevertheless nests the 'no frictionless economy at the limit' as a special case. Our findings suggest that information friction never completely vanish at the limit since they are necessary to explain the persistence in the dynamics of information friction themselves. However, analyzing whether information frictions completely vanish at the limit remains an open empirical question.

[28]The framework considered in the paper is in partial equilibrium and does not explicitly introduce a monetary or fiscal authority. Nevertheless, Section 2.6.5 highlights some policy implications.

jumps are relatively modest.[29,30] Second, the model is tractable and a simple iterative method makes it possible to track the evolution of the cross-sectional distribution of consumers. It therefore enables us to analyze multiple aspects of consumption nonlinearities. Moreover, it offers a natural benchmark to asses the approximation loss from relying on time-invariant information rigidities. In the absence of aggregate shocks, the cross-sectional distribution of consumers is stationary and information rigidities constant. Predictions from time-invariant information-rigidity models however become less accurate as the variance of aggregate shocks increases relatively to that of idiosyncratic shocks. Finally, the model accurately matches the persistence of aggregate consumption growth and provides an explicit mapping between aggregate information frictions and consumption growth persistence.

This paper belongs to a growing literature analyzing the implications of consumers' information rigidities for consumption dynamics and, more specifically, to consumption models with rational inattention or sticky expectations. Papers building on Sims' (2003) rational inattention generally consider linear-quadratic Gaussian frameworks (Luo, 2008; Luo and Young, 2014) or assume ex-post Gaussian distributions of the true state and noise (Luo et al., 2017).[31] Sticky expectation models fall into one of two categories. The first corresponds to the Calvo-like models in which agents have a constant probability to update their expectations, e.g., Mankiw and Reis (2002), Carroll (2003) and Carroll et al. (2020), and the second category builds on microfounded models of sticky expectations (Gabaix and Laibson, 2001; Reis, 2006a,b) where agents update every $n$th period. Consequently, both rational inattention and sticky expectations imply that information rigidities are constant over time. In comparison, this paper proposes a model of rationally inattentive consumers that nests both forms of information rigidities as limiting cases (within a

[29]See Reis (2006a) and Carroll et al. (2020) for a discussion of these models. Unattractive predictions of these models are that (i) we do not observe adjustment costs for non-durable consumption goods, (ii) consumption must be constant between adjustments and (iii) consumption adjustments must be large in these models.

[30]I find that, in order to match the dynamics of aggregate consumption, the consumption jumps implied by inattention are less than half the consumption jumps implied by adjustment costs/near-rationality (Caballero, 1995).

[31]A notable exception is Tutino (2013) who numerically solves for the optimal discrete distribution of actions when utility is CRRA. In comparison, the setup considered here is substantially different, the model is solved analytically and the focus is on aggregate consumption.

LQG framework) and that naturally generates time-varying information rigidities. These variations in information rigidities are found to be large and to hold important implications for consumption dynamics at both the household and aggregate levels.

A few papers have already proposed mechanisms to generate time-varying aggregate information rigidities. Gorodnichenko (2008) and Woodford (2009) consider the price setting problem of firms and respectively find that information externalities and inattention between price reviews may result in time-varying information rigidities. Similarly, Cheremukhin and Tutino (2016) highlight that time variations in firms' exit rates and markups lead to counter-cyclical information rigidities. These explanations are based on information rigidities on the side of firms. Nimark (2014) and Larsen et al. (2020) argue that the media coverage of economic events results in time-varying information rigidities. However, most of the uncertainty faced by consumers is the consequence of idiosyncratic shocks, for which the media are unlikely to be the main source of information.[32] This paper thus adds to this literature by providing a novel mechanism to generate time-varying information rigidities. This mechanism is internal – it depends only on the endogenous attention that each consumer allocates to the variations in her own income – and does not arise as the consequence of an aggregate externality. It is thus more likely to be transferable to other settings.

Finally, this paper relates to the large literature studying the consumption response to income changes (see Jappelli and Pistaferri (2017)). In particular, it provides novel explanations to some puzzling consumption nonlinearities previously identified in the data. These nonlinearities are shown to be the consequence of household level adjustments in expectations. Hence, the paper also relates to the large literature analyzing the implications of microeconomic adjustments for aggregate dynamics, e.g., Caballero and Engel (2007). In particular, some of the predictions for consumption discussed in the present paper could also obtain from a model with consumption adjustment costs. For instance, Caballero (1995) and Chetty and Szeidl (2016) show that this class of models may also explain the magnitude hypothesis and some non linear dynamics in aggregate consumption.

[32] Using the terminology in Nimark (2014), our model implies that media only report 'man-bites-dog' news, i.e., unusual events, not because of editorial constraints such as limited space in newspapers, but because consumers, who demand media, only consider these unusual events as attention worthy.

In comparison to both, I have already argued that two key distinctions for consumption are that the model in this paper does not predict that household consumption remains constant between adjustments[33] and that household consumption adjustments are relatively small. Nevertheless, the main contribution of the present paper to this class of models is, perhaps, to provide a rational for their usefulness for studying nondurable consumption, for which we do not observe direct adjustment costs. Indeed, our model predicts that the adjustment-like behaviors in consumption are the consequence of adjustments in consumers' expectations that we observe in the data (e.g. the first two puzzles mentioned earlier). Therefore, any attempt to explain non durable consumption dynamics should not rely on assumed ad-hoc adjustment costs, but match key features of consumers' expectations that naturally restrain the set of admissible adjustment-like behaviors in consumption.

The paper is organized as follows. The consumption problem is presented and solved in Section 2.2. Section 2.3 further discusses consumers' attention strategy and derives implications for inattention lengths. Section 2.4 provides a quantitative application with an ARMA income process. Sections 2.5 and 2.6 respectively derive implications for consumption dynamics at the household and aggregate levels.

## 2.2 Optimization with costly information processing

### 2.2.1 The consumer problem

This section presents the consumer's problem. Importantly, it discusses the information structure which is the main innovation of this paper and explains how it interacts with the consumer's problem.

I consider the problem of a rationally inattentive consumer with memory who lives

[33]A similar distinction exists in the pricing literature. Burstein (2006) proposes a model where firms may choose a pricing plan at revision dates, instead of price level. Here, I similarly allow consumers to choose a consumption plan when attentive, instead of a consumption level. In both Burstein (2006) and the model in the present paper, a plan can be contingent on the information available at the adjustment date but cannot be contingent upon future realizations of economic shocks.

for $T$ periods, consumes $c_t$ each period and whose utility $u(c_t)$ is quadratic.[34] 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 therefore writes $a_{t+1} = (1+r)a_t - c_t + y_t$ where $a_t$ are the consumer's assets at time $t$.

For ease of exposition, I follow the literature (e.g. Luo et al. (2017)) and reformulate the consumer's problem in terms of permanent income $s_t \equiv a_t + \mu_t$ where $\mu_t \equiv \sum_{k=t}^{T-1} E_t[(1+r)^{(t-k-1)} y_k]$ is the discounted expected present value of current and future (labor) incomes, i.e., human wealth. The period budget constraint may accordingly be written in terms of permanent income and thus becomes $s_{t+1} = (1+r)s_t - c_t + \zeta_{t+1}$ with $\zeta_{t+1} = \mu_{t+1} - (1+r)\mu_t$ a Gaussian white noise with variance $\sigma_\zeta^2$.[35]

Because of information frictions, the consumer cannot perfectly observe the economic state $s_t$. Nevertheless, there exists a source that continuously conveys information about the evolution of $s_t$. This information channel stands for instance for the news from newspapers, TV and cheap talk agents get every day and that may potentially reflect an evolution of the economic environment. At any moment, the consumer may either be attentive to this information – in return for a fixed utility cost denoted $\lambda$ – or remain inattentive and put this information aside for a later use.

Formally, I follow the signal extraction literature (e.g. Luo and Young (2014) for a recent application to consumer theory) and model the continuous information using an additive noisy signal $z_t = s_t + \vartheta_t$ where $\vartheta_t$ is an i.i.d. Gaussian white noise with variance $\sigma_\vartheta^2$ at each period. The smaller the variance of the noises $\sigma_\vartheta^2$ is, the more informative these signals are. Moreover, I define a latent information set, denoted $\mathcal{I}_t$, that contains all signals since the beginning of time and past actions of the consumer. These actions consist in her consumption choices ($c_k$) and whether she was attentive ($\tau_k = 1$) or not

[34]The quadratic utility assumption allows to derive an analytical solution and is a widely used framework for the study of rationally inattentive consumers (e.g. Sims (2003), Luo and Young (2014)).

[35]$\zeta_{t+1} \equiv \sum_{k=t+1}^{T-1} (1+r)^{t-k} \big(E_{t+1} - E_t\big)[y_k]$ is the innovation to permanent income.

($\tau_k = 0$) at each period $k \in [0, 1, \ldots, t-1]$. That is, $\mathcal{I}_t$ is the $\sigma$-algebra defined as

$$\mathcal{I}_t \equiv \left\{z_0, \tau_0, c_0, \ldots, z_{t-1}, \tau_{t-1}, c_{t-1}, z_t\right\} \tag{2.1}$$

This information is not observable by the consumer and is used as a modeling tool. Consequently, and in opposition to this latent information set, let $\overline{\mathcal{I}}_t$ be the information set in the hands of the consumer. $\overline{\mathcal{I}}_t$ and $\mathcal{I}_t$ are generally different and they coincide only when the consumer is constantly attentive to each new signal release. More specifically, we have

$$\overline{\mathcal{I}}_t \equiv \left\{\bar{z}_0, \tau_0, c_0, \ldots, \bar{z}_{t-1}, \tau_{t-1}, c_{t-1}, \tau_t, \bar{z}_t\right\} \tag{2.2}$$

The consumer's information set also contains past actions. However, it is not necessarily incremented at each period by the signal $z_t$ as the consumer may rationally prefer not to pay attention to the information channel. Instead, let $\bar{z}_t$ be the novel information that the consumer gets at period $t$. Then, by definition, we have that $\bar{z}_t = \emptyset$ is empty whenever the consumer is inattentive ($\tau_t = 0$). I do not impose a specific form for $\bar{z}_t$ at periods when the consumer is attentive. It could for example be a truncated sequence of signals or a filtration of these signals. The only restriction I impose is that at any period $t$, the consumer's information $\overline{\mathcal{I}}_t$ may be retrieve from the latent information set $\mathcal{I}_t$. That is, $\sigma(\{\bar{z}_k\}_{k=0}^t) \subseteq \sigma(\{z_k\}_{k=0}^t)$ where $\sigma(.)$ denotes a $\sigma$-algebra. Following Molin and Hirche (2010), I will refer to this property as the nestedness of the information structure. Intuitively, this condition implies that the signals $z_t$ are the only source of information for the consumer. Moreover, the consumer may catch up with any information she previously ignored.

The above-described consumer problem may be expressed as a discounted linear-quadratic Gaussian problem in the following form:

$$\begin{aligned} \min_{\{c_t,\tau_t\}_{t=0}^{T-1}\in\mathbb{R}^T\times\{0,1\}^T} \quad & E_0\left(\sum_{t=0}^{T-1}\beta^t\Big((c_t-\bar{c})^2+\lambda\tau_t\Big)+\beta^T q_T s_T^2\middle|\Big\{\mathcal{I}_t,\overline{\mathcal{I}}_t\Big\}\right) \qquad (2.3) \\ \text{s.t.} \quad & s_{t+1}=(1+r)s_t-(c_t-\bar{c})+\zeta_{t+1} \\ & c_t=f_t(\overline{\mathcal{I}}_t);\tau_t=g_t(\mathcal{I}_t) \\ & s_0|\overline{\mathcal{I}}_0\sim\mathcal{N}(\bar{s}_0,\sigma_{s_0}^2) \\ & \bar{s}_0=a_0+\mu_0-\frac{1-(1+r)^{-T}}{1-(1+r)^{-1}}\bar{c} \end{aligned}$$

Problem (2.3) states that the consumer maximizes her intertemporal utility given the aforementioned period budget constraint and information structure. Her instantaneous utility is quadratic $u(c)=-(c-\bar{c})^2$ with $\bar{c}\in\mathbb{R}^+$ the bliss point. The control variables are consumption $c_t$ and attention $\tau_t\in\{0,1\}$ and $\lambda$ is a fixed utility cost from being attentive. In the term $\beta^T q_T s_T^2$, $q_T$ is an arbitrary large constant which is used to impose the terminal condition $s_T=0$.[36] In the last condition, I normalize the initial state $\bar{s}_0$ by substracting an intertemporal consumption stream equal to $\bar{c}$ at each period to $s_0$.[37] The period budget constraint is amended accordingly (hence the term $(c_t-\bar{c})$). Finally, the third constraint imposes that the initial uncertainty in the state variable is Gaussian.

Following from the information structure, 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 $\mathcal{I}_t$. Hence, the consumption choice $c_t$ depends on the consumer's information at period $t$. However, attention strategies may be more sophisticated. Here, the consumer bases her attention strategy on the information contained in the latent

[36]This condition rules out Ponzi-schemes $\beta^{T-1}s_T\geq 0$ and ensures that the transversality condition holds. This constraint could as well be included in the list of constraints as is usually done in the consumption literature. Instead, we multiply it by $\beta$ and introduce it as a terminal condition in the loss function. This approach is common in the optimal control literature (Åström, 2012) and is mathematically convenient.

[37]This novel state variable represents the consumer's net permanent income given that she will consume $\bar{c}$ at each period. More specifically, the permanent income at period zero is $s_0\equiv a_0+\mu_0$, while the permanent income net of a constant consumption stream equal to $\bar{c}$ is $\bar{s}_0\equiv s_0-\sum_{t=0}^{T-1}(1+r)^{-t}\bar{c}$. This simple transformation allows to take $c_t-\bar{c}$ as the control variable, i.e., the consumer chooses her deviation from the consumption bliss point. To account for the new normalization of the state variable, the period budget constraint is amended to account for the deviation between actual consumption $c_t$ and the consumption bliss point $\bar{c}$. The budget constraint now also refers to the evolution of permanent income net of the constant consumption stream and, therefore, refers to $\bar{s}_{t+1}$ and $\bar{s}_t$. I omit the bar notation for notational convenience.

information set $\mathcal{I}_t$. That is, the consumer has the opportunity to select the type of information she will be attentive to. A real-life illustration of this strategy could for example to be attentive to economic news only when they contain keywords e.g. crises, recession or low interest rate.

The above formulation encompasses some well-known models of inattentive consumers as limiting cases. When the attention cost $\lambda$ is nil, the consumer is always attentive to the Gaussian signals. Hence, the consumer's information set coincides with the latent information set and problem (2.3) collapses to the textbook linear-quadratic Gaussian problem with incomplete state information.

Closer in spirit is the sticky information model of Reis (2006a). He also considers the problem of a rationally inattentive consumer who must pay a fixed cost to observe information. In his model, the consumer perfectly observes the state variable $s_t$ when she is attentive. Transposed to problem (2.3), it implies that the signals are noiseless $z_t = s_t$. Further, he restricts the attention strategies to depend only on the consumer's information. Using the above notation, he focuses on solutions in the form $\tau_t = g_t(\overline{\mathcal{I}}_t)$. In particular, this assumption implies that the consumer's attention strategy must be independent of the economic conditions between information updates. As a result, Reis (2006a) finds that the optimal updating behavior is purely time-dependent.[38]

### 2.2.2 Consumption and Information

This section highlights important intermediary results. In particular, it characterizes the optimal consumption policy and treatment of signals.

Problems related to (2.3) have recently been studied in engineering. The closest paper is Molin and Hirche (2010) who study an undiscounted discrete-time LQG setup with a similar information structure. In particular, they show that the certainty equivalence holds in their setup (Lemma 2, Molin and Hirche (2010)). Appendix 2.8 shows that the

[38] Reis (2006a) also develops a behavioral extension for extreme events where it is assumed that the latter are fully observable and instantaneously internalized by consumers. By construction, problem (2.3) directly accounts for these extreme events and allows to study their implications within a fully microfounded framework.

introduction of discounting does not affect this conclusion. Consequently, the consumption policy function coincides with the one we would have obtained under full-information rational expectations and is recalled in Lemma 1.

**Lemma 1** (Certainty equivalence)**.** *The optimal consumption is*

$$c_t = L_t E[s_t|\overline{\mathcal{I}}_t] + \bar{c} \qquad \forall t \in 0, \ldots, T-1 \tag{2.4}$$

*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* $p_T = q_T$*.*

*Proof.* See Appendix 2.8. □

Because the certainty equivalence holds here, the consumption function is not affected by the information structure. We therefore retrieve the well-established conclusions that the consumption function is linear and there are no precautionary savings when the consumer's utility is quadratic. More specifically, the consumption deviation from the bliss-point $\bar{c}$ depends on the discount rate $\beta$, the interest rate $r$, the terminal condition $q_T$ and the perceived permanent income given the information that has been processed at period $t$. The constant $L_t$ measures the change in consumption following a marginal increase in expected permanent income.

Equation (2.4) in Lemma 1 states that the consumption choice depends on the expected permanent income given the consumer's information set $\overline{\mathcal{I}}_t$. Characterizing the latter expectation $E[s_t|\overline{\mathcal{I}}_t]$ is not trivial. It first requires to determine the optimal estimator $E[s_t|\mathcal{I}_t]$ with respect to the latent information set $\mathcal{I}_t$. Following common practice in the literature (e.g. Sims (2003), Luo et al. (2017) and Maćkowiak et al. (2018)), I assume that the initial uncertainty surrounding the state variable $\sigma^2_{s_0}$ is at its steady state value. Consequently, $E[s_t|\mathcal{I}_t]$ is the linear least squares estimator given by the Kalman filter in Lemma 2.

**Lemma 2** (Latent Kalman filter)**.** *The optimal estimate of* $s_t$ *given the latent information* *set is*

$$E[s_t|\mathcal{I}_t] = (1+r)E[s_{t-1}|\mathcal{I}_{t-1}] - c_{t-1} + \bar{c} + K(z_t - (1+r)E[s_{t-1}|\mathcal{I}_{t-1}] + c_{t-1} - \bar{c}) \quad (2.5)$$

*where $K$ is the steady state Kalman gain defined in Appendix 2.9*

*Proof.* See Appendix 2.9. □

Now that we have identified $E[s_t|\mathcal{I}_t]$ in Lemma 2, we can characterize $E[s_t|\overline{\mathcal{I}}_t]$. When the consumer is attentive ($\tau_t = 1$), she may access the information contained in $\mathcal{I}_t$ to form an updated estimate $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1]$. Therefore, $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 1] = E[s_t|\mathcal{I}_t]$ since the latter estimator is optimal given $\mathcal{I}_t$ and the nestedness property of the information structure implies that there is no other source of information.

Deriving the optimal estimator $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0]$ at non-updating periods ($\tau_t = 0$) is more complex because it depends on the attention strategy of the consumer. To illustrate this dependance, realize that $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0] = E\Big[E[s_t|\mathcal{I}_t]\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big]$. Then, using equation (2.5) we have

$$\underbrace{E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0]}_{\text{estimate when inattentive}} = \underbrace{E[s_t|\overline{\mathcal{I}}_{t-1}]}_{\text{update}} + \underbrace{E\Big[(1+r)e_{t-1} + K(z_t - E[s_t|\mathcal{I}_{t-1}])\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big]}_{\text{corrective term accounting for inattention}} \quad (2.6)$$

where

$$e_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t] \quad (2.7)$$

is the perceived forecast error given the information in $\mathcal{I}_t$. The last term in the right-hand side of equation (2.6) represents a correction that the consumer may be willing to integrate in her consumption path when she decides to remain inattentive. Intuitively, if the choice to be attentive results from the occurence of a predetermined event, then being inattentive indicates that this event did not occur. For example, suppose that the consumer is extremely loss averse and willing to be attentive to permanent income drops only. Implicitly, this attention strategy would imply that no permanent income drops occur when she is inattentive. She would therefore infer that, on average, her permanent income is higher than what would be implied by a mechanical update $E[s_t|\overline{\mathcal{I}}_{t-1}] = (1+r)E[s_{t-1}|\overline{\mathcal{I}}_{t-1}] - c_t + \bar{c}$ at periods when she is not attentive. Consequently, she would revise her estimate $E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0]$ upward and, following from Lemma 1, increase her consumption during these periods.[39]

Appendix 2.10 shows that the corrective term accounting for inattention in equation (2.6) is nil at the optimum. We can therefore characterize the optimal permanent income estimate of the consumer in Lemma 3.

**Lemma 3** (Perceived permanent income)**.** *The optimal estimate of $s_t$ given the consumer's information set is*

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

*Proof.* See Appendix 2.10 . □

Lemmas 1-3 together characterize the consumption behavior of the consumer. At times when she is attentive, she chooses the consumption path which maximizes her expected intertemporal utility. The dynamics of this path is similar to the one she would have selected in the absence of information frictions. Therefore and unsurprisingly, it depends on the interest rate $r$, the discounting factor $\beta$, time $t$ and the horizon $T$. Then, at times when she is inattentive, she does not revise her consumption path and behaves as if no shock to permanent income occurred since the period she was last attentive. However, at times when she is attentive, she catches up with the information she previously ignored, revises her permanent income estimate and selects a novel consumption path.

[39] This mechanism, referred to as negative information, is central to event-based state estimation (see the book of Shi et al. (2016)). It may hold implications for consumption dynamics when one relaxes the assumption of Gaussian shocks to permanent income and/or quadratic utility. Extrapolating from Molin and Hirche (2010), it might result in a predetermined adjustment in consumption that would depend on time and the duration of inattention. See Nimark (2014) for an application of negative information in macroeconomics.

### 2.2.3 Attention strategy

This section focuses on the consumer's attention strategy. Building on the previous results, it demonstrates that the complex task of identifying the optimal attention strategy directly from the full consumer's problem (2.3) collapses to a much simpler discrete choice control problem that can be solved using standard dynamic programming tools.

The perceived forecast error $e_t$ defined in equation (2.7) measures the expected permanent income discrepancy between the latent information – i.e. the information the consumer would access if attentive – and the consumer's information when she remains inattentive. Using Lemmas 2 and 3, this error follows a Markov process given by

$$e_{t+1} \quad = \quad (1-\tau_t)(1+r)e_t + K\Big(z_{t+1} - E[s_{t+1}|\mathcal{I}_t]\Big) \tag{2.8}$$

That is, the error is incremented at each period by the innovation from the latent Kalman filter and previous errors are augmented by the gross interest rate $1+r$ whilst the consumer remains inattentive. Because $e_t$ depends only on the signals, consumption and attention choices, it is observable given the latent information $\mathcal{I}_t$. Moreover, Appendix 2.11 indicates that $e_t$ is a sufficient statistics to apprehend the consumer's disutility from being inattentive at time $t$. As a result, the difficult task of characterizing the optimal attention policy in problem (2.3) collapses to the much simpler task of computing the solution to the following discrete choice optimal control problem with perfect state observation $e_t$.

**Lemma 4** (Attention problem)**.** *The optimal attention strategy of the consumer is the solution to the following Bellman equation*

$$\begin{aligned} J_t(e_t) \quad &= \quad \min_{\tau_t\in\{0,1\}} (1-\tau_t)L_t^2(1+\beta p_{t+1})e_t^2 + \tau_t\lambda + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t] \\ \text{s.t.} \qquad & e_{t+1} = (1-\tau_t)(1+r)e_t + K\Big(z_{t+1} - (1+r)E[s_t|\mathcal{I}_t] + c_t - \bar{c}\Big) \end{aligned} \tag{2.9}$$

*Proof.* See Appendix 2.11 . □

The loss function in equation (2.9) represents the expected intertemporal utility costs associated to the attention choice once we account for Lemmas 1-3. When the consumer is inattentive ($\tau_t = 0$), her consumption choice is suboptimal given $\mathcal{I}_t$. In terms of intertemporal utility, this translates into an instantaneous misoptimization cost $L_t^2(1+\beta p_{t+1})e_t^2$ and a discounted future misoptimization cost $\beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t]$. When the consumer is attentive today ($\tau_t = 1$), she observes the perceived forecast error $e_t$ (Lemma 3) and adjusts her consumption plan accordingly (Lemma 1). Therefore, the consumer does not suffer from the aforementioned misoptimization costs but must pay a utility cost $\lambda$ to be attentive.

The problem in Lemma 4 is standard and can be solved using a DP algorithm. More specifically, the following Proposition holds.

**Proposition 1** (Attention strategy)**.** *The optimal attention policy $g_t(e_t)$ is symmetric and such that $g_t(e_t) = 1 \iff |e_t| \geq \pi_t$ and $0$ otherwise. The threshold $\pi_t \in \mathbb{R}^+$ follows from Lemma 4 and solves $\forall t \in \{0, \ldots, T-1\}$*

$$\lambda + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = 0] \quad = \quad L_t^2(1+\beta p_{t+1})\pi_t^2 + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = \pi_t] \tag{2.10}$$

*Proof.* See Appendix 2.11. □

There is therefore a symetric inattention region, such that $|e_t| < \pi_t$, where the consumer disregards new information and adopts a wait-and-see consumption strategy. Sporadically, the absolute value of the perceived forecast error gets larger than the threshold $\pi_t$. The occurence of this event triggers the consumer's attention.

### 2.2.4 Stationary policies

As a final step in characterizing the model solution, I consider the infinite horizon limit of problem (2.3). Appendix 2.12 demonstrates that when the horizon $T$ is infinite, problem (2.3) converges to stationary policies $f(.)$ and $g(.)$. These stationary policies are reported in Proposition 2.

**Proposition 2.** *When the horizon is infinite, the policy functions converges to stationary policies $f(.)$ and $g(.)$. Consequently, and assuming it exists, the consumption function is*

$$c_t = \frac{\beta(1+r)^2 - 1}{\beta(1+r)} E[s_t|\overline{\mathcal{I}}_t] + \bar{c} \tag{2.11}$$

*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*

$$\pi = \frac{\sqrt{\beta(1+r)\Big(\lambda + \beta(E[J(e_{t+1})|\mathcal{I}_t, e_t = 0] - E[J(e_{t+1})|\mathcal{I}_t, e_t = \pi]\Big)}}{\beta(1+r)^2 - 1} \tag{2.12}$$

$J(.)$ *is the functional fixed-point solution to the infinite horizon reformulation of the Bellman equation (2.9).*

*Proof.* See Appendix 2.12. □

The stationary consumption policy (2.11) is standard and we retrieve the well-known result that the consumption path is constant over time when $\beta^{-1} = (1 + r)$. Further, the inattention region becomes time-independent as well when the horizon tends to infinity.

Unless stated otherwise, I consider the infinite horizon formulation in the rest of the paper to avoid the unnecessary burden of indexing each variable with a time $t$ index. Extending the results to the finite horizon formulation nevertheless directly follows from using Proposition 1.

## 2.3 Consumer's inattention

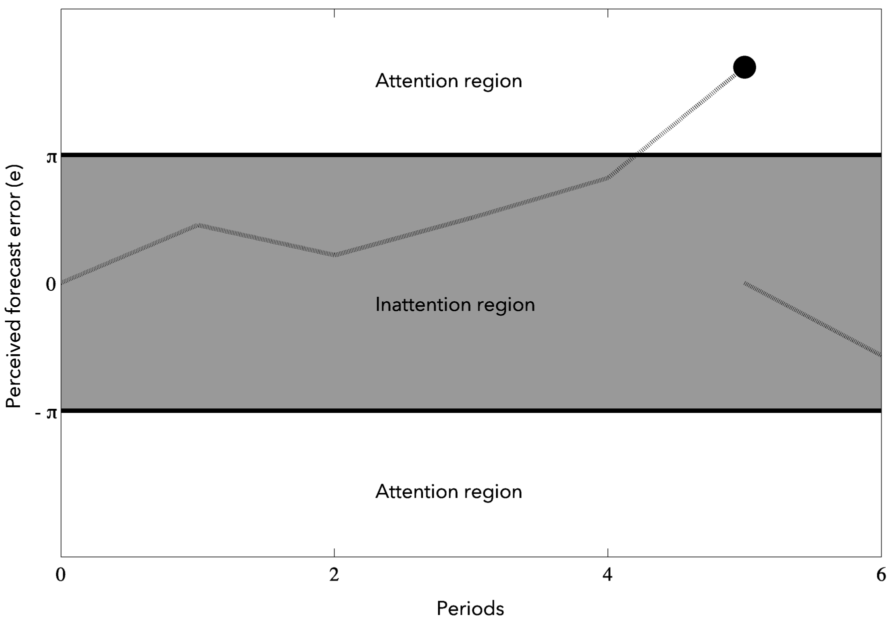
Following from Proposition 1, there exists an inattention region that depends on the discrepancy between the consumer's information and the latent information stemming from the continuous signals. This section further analyses how this inattention region affects the joint dynamics of the latent perceived forecast error and consumers' attention. It then derives implications for the distribution of inattention lengths.

### 2.3.1 Attention dynamics

The consumer's attention dynamics is driven by the latent perceived forecast error $e_t$ whose law of motion is given in equation (2.8). Further noticing that $z_{t+1} - E[s_{t+1}|\mathcal{I}_t]$ is the latent Kalman filter innovation, this law of motion equivalently writes $e_{t+1} = (1 -$

$\tau_t)(1+r)e_t + \omega_{t+1}$ where $\omega_{t+1}$ is a Gaussian white noise with variance $\sigma_\omega^2$.[40] Consequently, the perceived forecast error follows an AR(1) process with a reseting at 0 when the consumer is attentive ($\tau_t = 1$).

Figure 2.1: Inattention region




NOTE: An illustration of the joint dynamics of the latent perceived forecast error ($e_t$) and attention. The grey line represents the evolution of the latent perceived forecast error over time. The latter is unknown to the consumer. The only information available to her corresponds to black lines (i.e. the threshold $\pi$) and the black dot in the upper attention region. The threshold $\pi$ follows from the infinite horizon problem whose solution is time-invariant and reported in proposition 2.

Figure 2.1 illustrates the joint dynamics of the perceived forecast error and attention. Starting from a period 0 when the consumer was attentive, the perceived forecast error smoothly incorporates the continuous information arising from the Gaussian signals. This evolution is represented by the grey line on the graph. As long as the latent forecast error remains in the inattention region, the consumer does not observe it. However, when it exceeds the lower ($-\pi$) or upper ($\pi$) threshold, the consumer becomes attentive and observes $e_t$. In the illustration from Figure 2.1, this event occurs at the fifth period and is pictured with the black dot in the upper attention region. Because the consumer observes $e_t$ when being attentive, she catches up with the latent information and the dynamics of

[40]We have $\sigma_\omega^2 = K^2(\bar{p}_+ + \sigma_\vartheta^2)$ with $\bar{p}_+$ the steady state posterior variance of the latent Kalman gain defined in Appendix 2.9.

the perceived forecast error restarts from zero.

The attention dynamics depicted in Figure 2.1 is coherent with the observation that consumers sometimes rationally prefer to ignore new information. Using a controlled information experiment, Armantier et al. (2016) investigates how US consumers' inflation expectation are affected by the provision of novel information. They show that while the provision of information – about the average inflation forecast in the Survey of Professional Forecasters – increases the probability that a consumer is attentive to this information, the probability to remain inattentive to it remains large (42% in their study). Further, they find a nonlinear relation between the perception gap – namely the difference between a consumer's prior and the information she is provided with (i.e. the latent Kalman filter innovation in the framework of this paper) – and consumers' average revision, thus suggesting that consumers are more likely to be attentive to the information they are provided with when it contradicts their prior. Abe and Ueno (2015) run a similar information experiment on Japanese consumers and reach similar conclusions. Moreover, they provide direct evidence of an inattention region with respect to the perception gap.

### 2.3.2 Inattention lengths

*How long will a consumer remain inattentive?* A consumer's 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.

The attention dynamics discussed in Section 2.3.1 may be apprehended as resulting from a first time passage problem (with reseting).[41] That is, how long will it take for the latent perceived forecast error $e_t$ to reach one of the attention regions? Formally, let

[41] There are, at least, two complementary views to analyze the stochasticity of the attention behavior. One may either consider tracking the evolution of $e_t$ and analyze an update as a situation such that $|e_t| \geq \pi$. In this case, it is best to apprehend the updating as resulting from a first time passage problem (in discrete time) with reseting. On the other hand, one may only value the time dimension without regard for the specifics of the dynamics of $e_t$. In this case, one may use tools from survival analysis. In the rest of the paper, I will rely on tools from both approaches depending on which is the most convenient. See Aalen et al. (2001) for a discussion.

$l_t \equiv \sup\{i : \tau_i = 1, i \leq t\}$ 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

$$q(k) \equiv P(d = k) = P(\tau_{l_t+k} = 1 | \cap_{i=1}^{k-1} \tau_{l_t+i} = 0) \qquad \forall k \in \mathbb{N} \tag{2.13}$$

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 as well.

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

$$\Lambda(k) \equiv 1 - \int_{\Xi} f(e|k) de \qquad \forall k \in \mathbb{N} \tag{2.14}$$

where $\Xi \equiv [-\pi, \pi]$ and $f(e|k)$ is the distribution of the latent perceived forecast error $e_t$ given that the consumer was inattentive for $k$ consecutive periods. Equation (2.14) 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 2.3.1 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

$$f(e|k) \propto \int_{\Xi} f_{ar}(e|\bar{e}) f(\bar{e}|k-1) d\bar{e} \qquad \forall k \in \mathbb{N} \tag{2.15}$$

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. Equation (2.15) illustrates how the uncertainty from not using the extra information acquired through the latent information set $\mathcal{I}_t$ evolves whilst the consumer is inattentive. On the one hand, if the consumer was not attentive at the previous period, she did not observe $e_{t-1}$. Consequently, she did not adjust her consumption path accordingly and the uncertainty surrounding her current permanent income increases mechanically by a factor indexed on the gross interest rate $(1+r)$. This mechanical increase in uncertainty is captured through $f_{ar}(e|\bar{e})$ in equation (2.15). On the other hand, she knows that being inattentive at the previous period implies that $\tau_{t-1} = g(e_{t-1})$ was equal to zero. Since, at the optimum, she chooses a triggering law such that $g(e_t) = 1 \iff |e_t| \geq \pi$ and zero otherwise, she knows that $e_{t-1}$ belonged to $\Xi$. Hence, this negative information leads to truncate the integration of the distribution $f(\bar{e}|k-1)d\bar{e}$ in equation (2.15).

In order to provide quantitative predictions, it is necessary to compute the distribution $f(e|k)$ from equation (2.15). The latter distribution is not standard and explicitly iterating on equation (2.15) may lead to large numerical errors (Shi et al., 2016). I therefore 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 Appendix 2.15.2. □

In conclusion, 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 $\pi$ which characterizes the shape of the inattention region and the variance $\sigma_\omega^2$ reproducing the volatility of the perceived forecast error.

## 2.4 Application to ARMA income

This section calibrates the model parameters with an ARMA(1,1) income change process and derives quantitative implications for optimal inattentiveness.

### 2.4.1 Income process and calibration

The setup introduced in Section 2.2 requires that the income process follows from a multivariate linear state space model with Gaussian white innovations.[42] Following Friedman (1957), and more recently Reis (2006a) and Luo (2008), I assume that income is the sum of two independent components $y_t^P$ and $y_t^T$. The first component is the permanent part of income and follows a random walk with variance $\sigma_P^2$. It captures permanent variations in income that may arise for instance from changes in employment status, experience, education or severe health shocks. The second component is transitory income and follows an AR(1) with parameter $\rho$ and variance $\sigma_T^2$. Shocks to transitory income have a temporary effect on income and the larger $\rho$ is the less persistent their effects are. These transitory shocks may represent for instance fluctuations in overtime labor supply, bonuses, lottery prizes and bequests. MaCurdy (1982) finds that such income process fits the US data well.

Following the methodology presented in Section 2.2, I reformulate this income process in terms of permanent income $s_t$. Shocks to permanent income are thus equal to

$$\zeta_t \quad = \quad \frac{1+r}{r}\varepsilon_t^P + \frac{1+r}{1+r-\rho}\varepsilon_t^T \tag{2.16}$$

where $\varepsilon_t^P$ is the shock to the permanent part of income and $\varepsilon_t^P$ the shock to the tran-

[42]It is worth mentioning that these linear state space models are unable to capture important nonlinearities in the income process that have recently been identified (Meghir and Pistaferri, 2011; Arellano et al., 2017). Examining how these nonlinearities interact with consumers' inattention and, ultimately, consumption dynamics is left for future work.

sitory part. In the following, I directly focus on the impact of shocks to permanent income. It allows me to characterize the impact of income shocks independently of their type. Equation (2.16) nevertheless permits to retrieve the impact of each shock separately.

The income process is calibrated following Pischke (1995) and such that $\sigma_P = \$45$, $\bar{y} = \$6,926$, $\rho = 0.487$, $\sigma_T = \$1,962$ and $r = 0.015$. The time period is a quarter and the observational unit a household. The discount rate is $\beta = 0.99$. Regarding the updating behavior, estimates at the macro-level indicate that individuals update once a year on average (Carroll, 2003; Mankiw et al., 2003; Reis, 2006a). I set the attention cost $\lambda$ accordingly.

The remaining parameter $\sigma_\vartheta^2$ stands for the signal informativeness. Everything else being equal, $\sigma_\vartheta^2$ determines the latent Kalman filter gain from Lemma 2 and hence the rate at which income shocks are incorporated. I therefore calibrate $\sigma_\vartheta^2$ to match the impulse-response function of US consumption estimated in Reis (2006a). He finds that about 40% of the consumption response to an income shock arises on impact. Given the steady state dynamics of the model, this implies that the latent Kalman filter gain is equal to 55%.

### 2.4.2 Optimal inattentiveness

Table 2.1 reports the threshold $\pi$ normalized by the permanent income standard deviation. At the benchmark calibration, households update whenever their perceived forecast error $e_t$ is larger than 1.40 $\sigma_\varsigma$. Using the approximation procedure from Lemma 5 it implies that consumers update their expectations once a year on average.

Table 2.1: Optimal inattentiveness

| | Benchmark | Impact of a 5% decrease | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $r$ | $\beta$ | $\rho$ | $\sigma_T$ | $\sigma_P$ | $\lambda$ | $\sigma_\vartheta$ |
| $\bar{\pi}$ | 1.40 | 1.45 | 0.82 | 1.43 | 1.43 | 1.42 | 1.37 | 1.40 |
| $\bar{d}$ | 4.00 | 4.19 | 2.25 | 4.09 | 4.10 | 4.07 | 3.90 | 4.01 |

NOTE: Optimal normalized threshold $\bar{\pi} = \pi/\sigma_\varsigma$ 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 10% while keeping others constant.

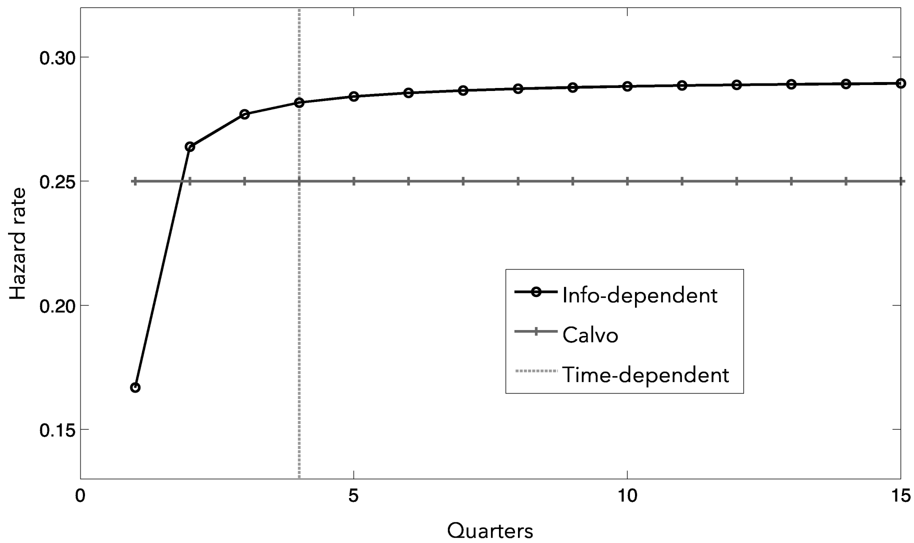
In order to asses the sensitivity of consumers' optimal inattentiveness to the model parameters, Table 2.1 also displays the implied change in the normalized threshold and average duration between updates when one parameter decreases by 5% while others remain at the benchmark calibration. The information provided in Table 2.1 could thus be used to compute the updating threshold and average duration elasticities with respect to each of these parameters.[43] The updating threshold and average duration are decreasing in the persistence of transitory shocks $\rho$, the standard deviation of permanent and transitory shocks (resp. $\sigma_P$ and $\sigma_T$). These results are not surprising as an increase in any of the latter 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. Our simulations indicate that this second effect dominates. The updating threshold and average duration increase with the discount rate $\beta$. This is because an individual smoothes consumption more when she values tomorrow more. Therefore, when $\beta$ 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 inattentiveness length increase when the attention cost $\lambda$ increases. Finally, the attention behavior is relatively unaffected by the signal informativeness $\sigma_\vartheta$. 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 2.1, 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

[43] For example, the elasticity of the average duration with respect to the interest rate $r$ is $-\frac{4.19-4}{0.05\times 4} = -0.95$. These elasticities are only valid to locally asses the impact of a five percent decrease in the parameter values.

the optimal attention strategy of the consumer.[44]

Figure 2.2 reports the hazard rates for the benchmark calibration. As can be seen, they increase with the inattention length. In particular, the probability that a consumer updates after only one quarter is small, less than 17%. These probabilities, conditionally on not updating in between, then smoothly converge to a value of 29% as time passes.

Figure 2.2: Inattention in sticky expectation models




NOTE: The figure reports the hazard rates for different models of sticky expectations. For the information-dependent model considered in this paper, the distributions is computed for the benchmark calibration and relies on the approximation procedure from Lemma 5. The Calvo model assumes that there is a constant probability to update at each period. It corresponds to the sticky expectation model in e.g. Mankiw and Reis (2002), Carroll (2003) and Carroll et al. (2020). The time-dependent model assumes that consumers update on a purely time dependent basis. It corresponds to the sticky expectation model in e.g. Reis (2006a). For the latter, the vertical line indicates that the hazard rate jumps from zero to one, and return back to zero afterward. All models are calibrated so that the average inattention length is equal to a year.

The literature offers different models of sticky expectations that lead to incompatible predictions regarding agents' inattention. The first generation of sticky information models (e.g. Mankiw and Reis (2002), Carroll (2003) and recently Carroll et al. (2020)) assumes that agents have a constant probability to update their expectations, independently of their attention history. I label these models as Calvo sticky expectations models in analogy to the Calvo price setting model (Calvo, 1983). As is reported in Figure 2.2

[44]The relative independence of the attention strategy with respect to the signal precision holds locally given our calibration. Simulations and straightforward intuitions indicate that this property is not global.

these models imply that the hazard rates are constant and equal to 25% when the average inattention length is of a year. Following these initiatives, Reis (2006a) develops a microfounded model of sticky expectations and reach the opposite conclusion that the updating behavior depends exclusively on the attention history. That is, assuming away any form of ex ante heterogeneity, each consumer should update her expectations each year (for the average inattention length to be of a year). In Reis' [2006a] model the attention strategy is therefore purely time-dependent and the associated hazard rates – reported in Figure 2.2 as well – are systematically nil excepted at the fourth quarter where it is equal to 1. The model considered in this paper follows Reis' [2006a] attempt to provide microfoundations to sticky expectations. By allowing for more general attention strategies, it generates an information-dependent attention[45] with increasing hazard rates. In this regard, it is more flexible as it allows the probability that a consumer becomes attentive to evolve with her inattention history but does not impose that inattention histories are the sole determinant of attention.

### 2.4.3 Welfare cost

Appendix 2.14 decomposes the overall welfare cost from costly information processing. It may be apprehended as the sum of three independent terms: the utility cost from paying $\lambda$ 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 2.2 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. When the CRRA is equal to one, the overall welfare cost represents 0.04% of a household consumption (at period 0). That is, less than \$1.3 per quarter. These costs remain negligible even when one considers extremely risk averse consumers. When the CRRA is 10, the welfare cost increases to only 0.29% of consumption.

[45] In analogy to the pricing literature, one may be willing to label the present model as one of state-dependent attention. I nevertheless believe that this terminology might be misleading as the state of interest to the consumer here is permanent income $s_t$. However, when the information source is imperfect, this state is never observable and the consumer bases her attention behavior according to the information content of the information she will observe (i.e. the latent perceived forecast error $e_t$).

Table 2.2: Welfare Cost

| Coeff. of relative risk aversion | 1 | 2 | 4 | 10 |
|---|---|---|---|---|
| Bliss point $\bar{c}$ | $\$13,852$ | $\$10,389$ | $\$8,658$ | $\$7,619$ |
| Overall welfare cost (% consumption) | 0.04% | 0.07% | 0.12% | 0.29% |
| Welfare decomposition ($/quarter) | 1.28 | 2.57 | 5.13 | 12.83 |
| *Updating cost* | *0.70* | *1.40* | *2.81* | *7.02* |
| *Latent information cost* | *0.41* | *0.81* | *1.62* | *4.05* |
| *Noisy information cost* | *0.18* | *0.35* | *0.71* | *1.76* |
| Consumption change to update | 2.92% | 2.32% | 2.10% | 1.99% |

NOTE: The welfare cost refers to misoptimization cost induced by costly information processing. The coefficient of relative risk aversion is equal to $\mu/(\bar{c}-\mu)$. 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 problem.

Decomposing the welfare cost, we find that more than half of it is attributable to the fixed utility cost $\lambda$ that the consumer pays when attentive. When the CRRA is equal to one (resp. 10), the monetary equivalent from paying $\lambda$ when being attentive is ¢70 (resp. $7.02) per quarter. Given that she will update once a year on average, the monetary equivalent for $\lambda$ is $2.1 (resp. $21.06).

Consumption models with fixed costs of adjustment (e.g. Caballero (1995) and Chetty and Szeidl (2016)) predict that we should observe (i) large consumption jumps and (ii) long lasting periods when a household consumption remains constant.[46] The model presented in this paper does not share these unappealing predictions. First, the predicted consumption jumps are modest. Given the model calibration, the consumption change which prompts the consumer to update ($L \times \pi$) is relatively small, equal to $133. When normalized by period 0 consumption, the last row in Table 2.2 reports an order of magnitude of a few percentage points. In contrast Caballero (1995) estimates that in order to capture the stickiness of US aggregate consumption data, the implied jump in an adjustment consumption model would be of almost 6% – namely twice as large than what we find here. Second, and as is further discussed in the subsequent section, consumption is not constant here whilst the consumer is inattentive. Intuitively, this is because the consumer is inattentive around a consumption path and not a consumption level.

[46]See Reis (2006a) and Carroll et al. (2020) for a critical discussion.

## 2.5 Households' consumption dynamics

This section 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.

### 2.5.1 Consumption changes

At the micro 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 equation (1) that

$$\Delta c_{t+1}|(\tau_{t+1} = 0) = (r - L)c_t \tag{2.17}$$

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

$$\Delta c_{t+1}|(\tau_{t+1} = 1) = Le_{t+1} + \Delta c_{t+1}|(\tau_{t+1} = 0) \tag{2.18}$$

is a Borel-measurable function with associated $\sigma$-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 (2006a) 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 (2.8), (2.17) and (2.18) together imply that $\Delta c_t$ 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 $\Delta c_t$ in that case. $\square$

Many studies have tested wether household level consumption growth is serially correlated. The recent meta-analysis of Havranek et al. (2017) considers 190 estimates from these studies on micro data and reports a median (resp. mean) estimate of 0.0 (resp. 0.1). These findings are in line with Proposition 3. They are however hardly reconcilable with the use of consumption habits as a mechanism to generate smoothness in aggregate consumption dynamics: "If habits are a true structural characteristic of people's utility functions, we should see their effects in microeconomic data as well as macroeconomic aggregates. But empirical studies using household-level data strongly reject the existence of habits of the magnitude necessary to explain aggregate consumption dynamics." (Carroll et al., 2020)

### 2.5.2 Asymmetry and magnitude

Now, consider a translation in the distribution of permanent income shocks at time $t+1$ such that $\zeta_{t+1} \sim \mathcal{N}(\bar{\zeta}, \sigma_\zeta^2)$. Let $\text{MPC}(e, \bar{\zeta}) \equiv \partial E[c_{t+1}|e = e_t]/\partial\bar{\zeta}$ be the expected instantaneous marginal propensity to consume for a consumer whose latent perceived forecast error is $e_t$. Then,

$$\text{MPC}(e_t, \bar{\zeta}) = L(1+r)\left[\frac{\partial q_1(e_t + \nu\bar{\zeta})}{\partial\bar{\zeta}}(e_t + \nu\bar{\zeta}) + \nu q_1(e_t + \nu\bar{\zeta})\right] \tag{2.19}$$

where $\nu = K/(1+r)$ and $q_1(x) = 1 + \Phi(-(\pi + (1+r)x)/\sigma_\omega) - \Phi((\pi - (1+r)x)/\sigma_\omega)$ is the probability to update in one period given an initial latent forecast error $x$. Equation (2.19) underlines the two margins affecting the expected consumption response to an income change. On the one hand, and taking the probability to update as given, the consumer will internalize a proportion $K$ of the shock when updating. Accordingly, the expected marginal propensity to consume increases by $LK$ times the probability of an update. On the other hand, the shock $\bar{\zeta}$ also affects the consumer's probability to update. The latter probability is $q_1(e_t + \nu\bar{\zeta})$ where the term $\nu$ is used to account for the fact that the consumer will only internalize a fraction $K$ of the shock on average and to express the impact of the shock on the probability from a change in the initial condition at period $t$.

Equation (2.19) further reveals that the expected marginal propensity to consume is both history- and shock-dependent. To highlight the mechanisms behind these two dependences, first realize that the function $q_1(x)$ is symmetric around its minimum at 0, attains its maximum 1 at $\pm\infty$ and is monotonically decreasing on $[-\infty, 0]$ and monotonically increasing on $[0, \infty]$. As a consequence, the marginal change in the probability to update at the next period given a history leading to $e_t$ and a shock $\bar{\zeta}$ is positive if and only if $e_t + \nu\bar{\zeta} \geq 0$ and negative otherwise. When the latter term is positive, the consumer is more likely to internalize the impact of the income shock and its expected marginal propensity to consume out of this income news is larger. More specifically, equation (2.19) implies the following behavior

**Proposition 4.** *The one-period ahead expected marginal propensity to consume out of an income shock is such that*

- $\mathrm{MPC}(e_t, \bar{\zeta})$ *increases with respect to* $|e_t + \nu\bar{\zeta}|$.
- *Given that* $e_t$ *belongs to* $\Xi_t$*, there always exists a finite and large enough income shock* $\varkappa$ *such that*

$$\mathrm{MPC}(e_t, \varkappa) > \mathrm{MPC}(e_t, 0) \quad \text{and} \quad \mathrm{MPC}(e_t, -\varkappa) > \mathrm{MPC}(e_t, 0)\ \forall e_t \in \Xi_t$$

- *Let* $\bar{\zeta} > 0$*, then*

$$\begin{aligned}
\mathrm{MPC}(e_t, \bar{\zeta}) &= \mathrm{MPC}(e_t, -\bar{\zeta}) \iff e_t = 0 \\
\mathrm{MPC}(e_t, \bar{\zeta}) &> \mathrm{MPC}(e_t, -\bar{\zeta}) \iff e_t > 0
\end{aligned}$$

According to the magnitude hypothesis, the consumption response to an income shock depends on its size. Such relation is expected to hold here as is apparent from the first two bullets in Proposition 4. The literature review in Section 2.1 mentions studies providing evidence to support the magnitude hypothesis. A few authors postulated that rational inattention may offer an explanation since the costs from not smoothing consumption increase with the size of the shock. For example, Hsieh (2003) states that "households will not bother to change their consumption paths when the computational costs involved are large relative to the utility gains". A similar argument is made in Browning and Collado (2001): households "do not bother to adjust optimally to small income changes since the utility cost [...] is small". The consumption model developed in this paper offers a microfounded framework confirming this guess. Accordingly, the magnitude hypothesis arises as a consequence of the joint dynamics of information rigidities and consumption. Large income shocks are more likely to prompt consumers to revise their expectations and, consequently, to adjust their consumption path to account for this shock. It it is worth mentioning that this conclusion holds independently of the sign of the income shock and therefore differs from any explanation based on the presence of credit constraints or risk aversion.[47] Proposition 4 however reveals that the size of an income shock is not a sufficient statistics to apprehend the magnitude hypothesis because other perceived income shocks since the last update matter as well. Alternatively saying, the consumption response to an income shock is both history (through $e_t$) and shock-dependent (through $\bar{\zeta}$). One shall therefore simultaneously account for both dependences to derive implications for the consumption response.

The first and last bullets in Proposition 4 indicate that the consumption response to an income shock is asymmetric with respect to the sign of the shock. This asymmetry results from two complementary forces, the history-dependence and the shock-dependence. The consumption model predicts that the expected marginal propensity to consume at the household level is large for positive shocks when the perceived change in permanent income that has not been internalized yet is positive (and large) and for negative shocks when the perceived change in permanent income that has not been internalized yet is negative (and large in absolute value). On the other hand, the expected MPC decreases after a positive shock when the perceived change in permanent income that has not been internalized yet is negative (and large in absolute value) and after a negative shock when the perceived change in permanent income that has not been internalized yet is positive (and large). These asymmetries, which cannot be explained by standard extensions of

[47] Following an income decline, a credit constraint does not affect the marginal propensity to consume. See Jappelli and Pistaferri (2010) for a discussion. Regarding risk aversion, Tutino (2013) shows that the consumption response to negative income shocks is higher in a framework with a CRRA utility and inattention à la Sims (2003).

the permanent income model, such as credit constraints or habits, have been identified by Caballero (1995) using data on US aggregate consumption. Up to the best of my knowledge, no study has analyzed the potential asymmetric reaction to negative and positive shocks during periods of income increases and declines at the household level.

Finally, equation (2.19) also indicates that the expected marginal propensity to consume out of an income shock is bounded by the rate at which the estimator with respect to $\mathcal{I}_t$ incorporates new information. It is easily seen from computing $\lim_{\bar{\zeta}\mapsto\pm\infty} \mathrm{MPC}(e_t, \bar{\zeta}) = L(1+r)K$ where $0 < K < 1$ is the Kalman gain from equation (2.5).

## 2.6 Aggregate consumption dynamics

This section focuses on the implications for aggregate consumption. It shows that costly information processing generates impulse response functions for consumption which depend on both the state of the economy and the size of the shock. Moreover, aggregate consumption growth is found to be highly persistent at the steady state. This persistence however depends on the endogenously time-varying share of attentive consumers.

### 2.6.1 Cross-sectional distribution

To asses the aggregate dynamics of the economy, I assume it is composed of a unit mass continuum of infinitely-lived consumers who are ex ante identical and whose problem is given by (2.3). For simplicity, I further assume that the discount rate is equal to $\beta = (1+r)^{-1}$ in the following. Because they have different idiosyncratic shocks, signals and updating histories, households differ in terms of their latent perceived forecast error. Analyzing the aggregate dynamics of information frictions and consumption therefore requires to keep track of the distribution of households. To this end, let $a_t(e)$ denote the cross-sectional distribution of consumers at the end of period $t$ and before the resetting. The share of updating consumers in the economy at period $t$ is therefore $\lambda_t \equiv \int_{e\notin\Xi} a_t(e)de$.

In the absence of aggregate shocks, the cross-sectional distribution dynamics is

$$a_t(e) \propto \frac{1}{\sigma_\omega}\Bigg[\underbrace{\int_{\tilde{e}\in\Xi}\phi\left(\frac{e-(1+r)\tilde{e}}{\sigma_\omega}\right)a_{t-1}(\tilde{e})d\tilde{e}}_{\text{Non updaters at } t-1}+\underbrace{\phi\left(\frac{e}{\sigma_\omega}\right)\int_{\tilde{e}\notin\Xi}a_{t-1}(\tilde{e})d\tilde{e}}_{\text{Updaters at } t-1}\Bigg] \tag{2.20}$$

The dynamics of the cross-sectional distribution directly follows from the dynamics of the latent perceive forecast error in equation (2.8). Proposition 5 characterizes the steady state cross-sectional distribution in the absence of aggregate shocks.

**Proposition 5.** *Equation (2.20) admits a stationary distribution*

$$a^\star(e) = \sum_{k=1}^{\infty}\lambda^\star(k)f^\star(e|k) \tag{2.21}$$

*where*

$$\lambda^\star(k) = \lambda^\star(1)S^\star(k-1) \quad \forall k \geq 2 \tag{2.22}$$

*and* $\lambda^\star(1) = 1/\sum_{k=0}^{\infty}S^\star(k)$ *with* $S^\star(k)$ *the unconditional survival function and* $\lambda^\star(k)$ *the steady-state share of consumers whose last update was k periods ago.*

*Proof.* Appendix 2.13. □

I now introduce aggregate shocks in the economy and make the following assumption.

**Assumption 1.** *Shocks to permanent income* $\zeta_t$ *are the sum of an aggregate shock* $\chi_t$ *common to each consumer and an independent idiosyncratic shock* $\iota_{i,t}$*. Both shocks are i.i.d. Gaussians with zero mean and respective variance* $\sigma_\chi^2$ *and* $\sigma_\iota^2$*.*

Under assumption 1, the consumption and triggering policies from Proposition 2 are unaffected. As a result, the only impact of aggregate shocks is to continuously and persistently translate the cross-sectional distribution of consumers. Indeed, when a shock occurs, it is only observable through the noisy information channel $\mathcal{I}_t$ and is therefore gradually perceived by consumers. Therefore, an aggregate shock does not only disturb the cross-sectional distribution when it occurs, but does so persistently. Given that consumers rely on a latent Kalman filter to incorporate new information (Lemma 2), the translation in the cross-sectional distribution at period $t$ is given by a weighted sum of past aggregate income shocks $\{\bar{\chi}_i\}_{i=1}^{t}$

$$S_t = \sum_{s=0}^{t-1} \upsilon_s \bar{\chi}_{t-s} \tag{2.23}$$

where

$$\upsilon_s = K(1+r)^s(1-K)^s \tag{2.24}$$

is the share of the aggregate shock – augmented by its returns – that is internalized on average at period $t+s$. Accounting for this new state variable, it is possible to derive the impact of an aggregate shock on consumption dynamics for any history of aggregate shocks.

**Proposition 6.** *The dynamics of the economy in the presence of aggregate shocks is characterized by the following system of dynamic equations*

$$a_{t+1}(e) \propto \frac{1}{\sigma_\omega}\left[\int_{\tilde{e}\in\Xi} \phi\left(\frac{e - S_{t+1} - (1+r)\tilde{e}}{\sigma_\omega}\right) a_t(\tilde{e})d\tilde{e} + \int_{\tilde{e}\notin\Xi} \phi\left(\frac{e - S_{t+1}}{\sigma_\omega}\right) a_t(\tilde{e})d\tilde{e}\right] \tag{2.25}$$

$$S_{t+1} = (1-K)(1+r)S_t + K\bar{\chi}_{t+1} \tag{2.26}$$

*along with initial conditions* $a_0(e)$ *and* $S_0 = \int_{\mathbb{R}} e a_0(e) de$.

Proposition 6 indicates that $S_t$ follows a markovian process. In general, the dynamics of $S_t$ is stationary since $r$ should be small in comparison to $K$. This state variable characterizes the business cycle in this economy: a positive $S_t$ coincides with an expansionary cycle while a negative $S_t$ results in a contraction.

### 2.6.2 Consumption dynamics at the steady state

To illustrate the dynamics of the economy starting from the steady-state cross-sectional distribution, I first analyze a one time only aggregate shock. Formally, consider a shift $\bar{\chi}$ in the distribution $\chi_t \sim \mathcal{N}(\bar{\chi}, \sigma_\chi^2)$ at time $t$. Because the economy is initially at its steady-state and there is a one time only aggregate shock, Proposition 6 implies that the impact of the aggregate shock on the cross-sectional distribution follows from iterating on

$$a_{t+s}(e)|\bar{\chi} \propto \frac{1}{\sigma_\omega}\left[\int_{\tilde{e}\in\Xi}\phi\left(\frac{e-\upsilon_s\bar{\chi}-(1+r)\tilde{e}}{\sigma_\omega}\right)a_{t-1+s}(\tilde{e})d\tilde{e}+\phi\left(\frac{e-\upsilon_s\bar{\chi}}{\sigma_\omega}\right)\int_{\tilde{e}\notin\Xi}a_{t-1+s}(\tilde{e})d\tilde{e}\right] \tag{2.27}$$

which has for initial condition $a_{t-1}(e) = a^\star(e)$ so that $S_{t-1} = 0$. Therefore, the impact of an aggregate shock is first to shift the cross-sectional distribution of consumers. The stationary distribution $a^\star(e)$ being symmetric, unimodal and centered around zero, an aggregate shock increases the share of agents who update and the magnitude of the average consumption change of those who would have updated in the absence of the shock.[48]

Not everyone will update (a.s.) after the aggregate shock. Hence, the cross-sectional distribution at the next period must account for the impact of the initial shock for those who have not updated (whence the term $\int_{\tilde{e}\in\Xi}\phi\left(\frac{e-(1+r)\tilde{e}}{\sigma_\omega}\right)a_{t-1+s}(\tilde{e})d\tilde{e}$ in the right-hand side of equation (2.27)). Moreover, because information is imperfect, the cross-section of consumers continues to be translated at each period in order to account for new pieces of information regarding the initial shock after it occurred. This new information depends on the rate at which agents rely on their signals $z_t$ to form $E[s_t|\mathcal{I}_t]$. Consequently, the cross-sectional distribution is translated a second time at period $t+1$ by a factor $\upsilon_1\bar{\chi}$. As time passes, $\upsilon_s$ tends to zero and the cross-sectional distribution is not disturbed anymore. It then smoothly converges back to the stationary distribution $a^\star(e)$ as consumers update to account for their idiosyncratic shocks.

Aggregating equations (2.17) and (2.18) over consumers when $\beta = (1+r)^{-1}$, the impulse response function (IRF) following an aggregate shock is thus

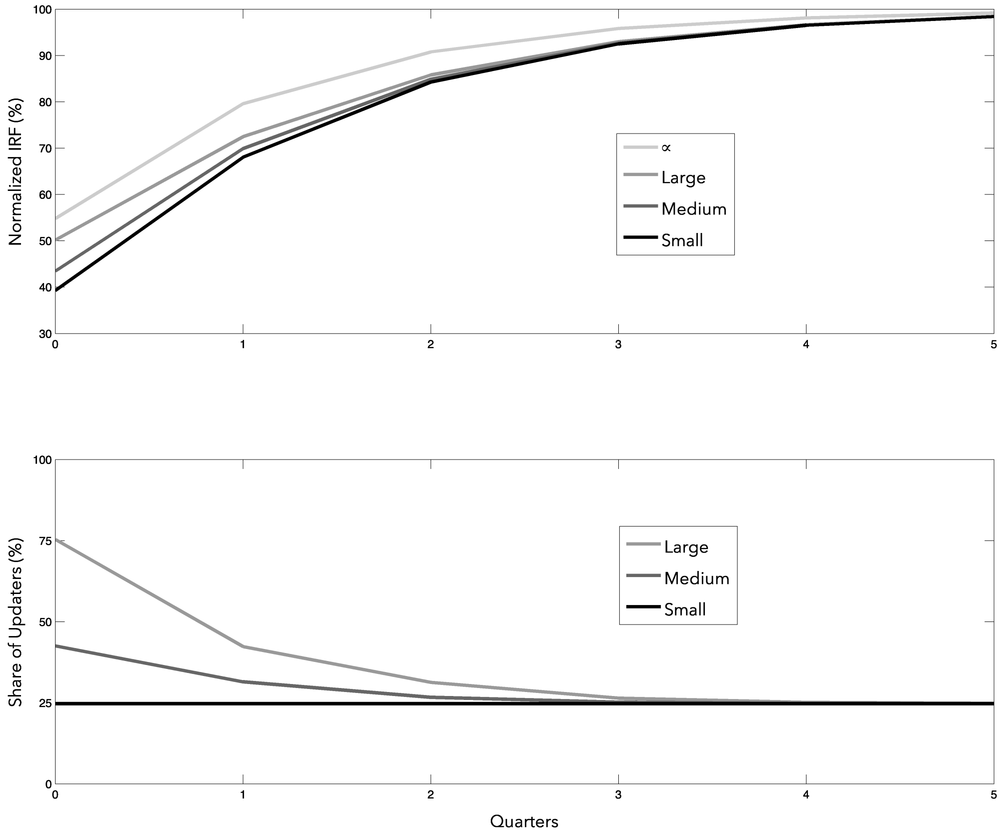
$$\Delta C_{t+s}|_{\chi=\bar{\chi}} - \Delta C_{t+s}|_{\chi=0} = L\int_{e\notin\Xi} e\ \left(a_{t+s}(e)|\bar{\chi} - a^\star(e)\right)\,de \tag{2.28}$$

Equation (2.28) highlights the dependence between consumption and information rigidities dynamics. Indeed, $(a_{t+s}(e)|\bar{\chi} - a^\star(e))$ stands for the change in the perceived forecast

[48] In the state-dependent pricing literature, these responses are respectively labeled the extensive and intensive margins (Caballero and Engel (2007)).

error distribution stemming from the aggregate shock. It thus captures the change in the share of updating consumers – as well as the change in their average $e$.

Figure 2.3: IRF at the steady state




NOTE: The figure reports the impulse response function of aggregate consumption (top panel) and shares of updates (middle panel) for different values of aggregate income shocks. The small shock is a one dollar shock, the medium is a one standard deviation $\sigma_\zeta$ shock and the large shock a two standard deviations shock. The bottom panel plots the evolution of the cross-sectional distribution of consumers following the medium shock. The large plain line is the steady-state distribution.

Figure 2.3 plots the impulse response function for different aggregate shocks. As is apparent from the top panel, aggregate information rigidities and consumption dynamics are shock-dependent. When the shock size is small, it barely affects the updating behavior which remains at its steady state dynamics. The shock is then smoothly accounted for by consumers as they update to internalize the impact of idiosyncratic shocks. However, when the shock is large, the share of agents updating jumps on impact and persistently remains above its steady state level. As a consequence, the short run response of aggregate consumption is much sharper following a large shock. Irrespectively of the size of the initial shock, the normalized IRFs then converge as consumers update to internalize the impact of idiosyncratic shocks.[49] Figure 2.3 also reports the dynamics at the limit when $\bar{\chi} \mapsto \infty$. After such shock, the share of updaters jumps to 100%. The instantaneous normalized response is thus equal[50] to $1-(1-K)(1+r)$. At the second period, the new piece of information prompts each consumer to update again a.s. and the normalized response is equal to $(v_0+v_1)(1-(1-K)(1+r))/K$.

### 2.6.3 Consumption dynamics during booms and busts

The previous section studies the consumption dynamics assuming the economy was initially at its steady state and concludes that it depends only on the size of the aggregate shock. However, and similarly to what we found at the household level, aggregate consumption dynamics is both history-dependent and shock-dependent. The present section thus considers the role of previous aggregate shocks for consumption dynamics. Importantly, it shows that the dynamics of aggregate consumption depends on the state of the economy when the shock occurs.

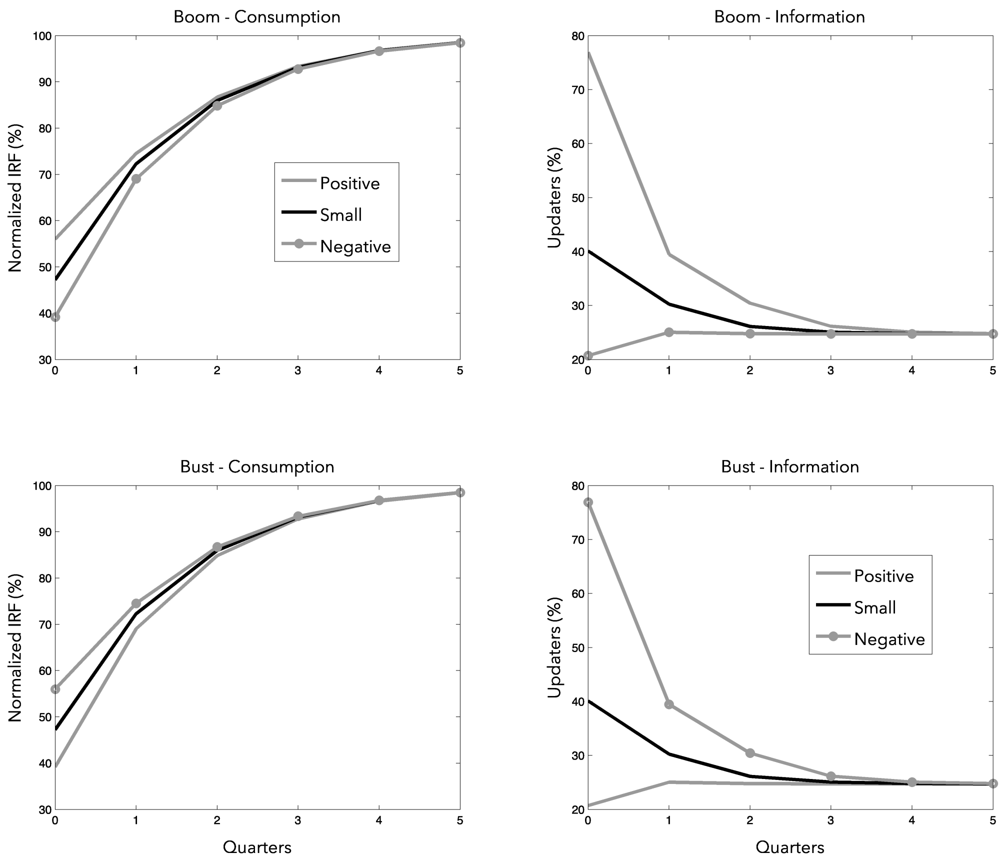
The situation studied in section 2.6.2 was such that $a_0(e)$ was equal to $a^\star(e)$. The mean of the latter being nil, we had $S_0 = 0$. It was thus a special case of proposition 6. The impulse response function follows from equation (2.28) where the counterfactual distribution is now $a_{t+s}(e)|\bar{\chi}_t = 0$, that is the cross-sectional distribution in the absence of aggregate shock at time $t$. Figure 2.4 reports the implied impulse response functions when the economy is initially either in a booming period ($S_0 > 0$) or in a busting period ($S_0 < 0$). The implied dynamics are fundamentally different. During a boom (resp. bust), a positive (resp. negative) aggregate shock increases the share of updates and consumers rapidly revise their consumption plan accordingly. In this case, the consumption response

[49] Note that the long run impact of the shocks are not identical because income news that haven't been processed grow at a constant rate $r$. See Luo (2008) for a related discussion. However, the normalized IRF being defined as the response at time $t+s$ divided by the long run response, it must converge to one.

[50] When $(1-K)(1+r) < 1$. This value is computed as follows. The long run response is $K/(1-(1-K)(1+r))$, and the response on impact is $v_0 = K$ as everyone update. The normalized response being the ratio of the latter over the former, it is equal to $1-(1-K)(1+r)$.

is brisker when the shock size is larger. However, a positive (resp. negative) shock may have an ambiguous effect on the updating behavior during recessions (resp. expansions). If the size of the shock is small relatively to $|S_0|$, then the share of updates decreases after the shock and consumption dynamics are more sluggish. However, and similarly to proposition 4, there always exists a large enough income shock such that consumers update more and consumption dynamics are abrupter. This latter observation may easily be understood by realizing that the impact of an infinitely large shock (computed in section 2.6.2) is independent of the initial state of the economy.

Figure 2.4: IRF During Booms and Busts




NOTE: The figure reports the impulse response functions of aggregate consumption (left panels) and the shares of updates (right panels) following a one and two standard deviations positive permanent income shocks and a one and two standard deviations negative shocks when the economy is initially in a boom (top panels) or a bust (bottom panels).

Caballero (1995) finds that US aggregate consumption dynamics display asymmetries during booms and busts as those described above: "in good times, consumers respond more promptly to positive than to negative wealth shocks, while the opposite is true in bad times". Similarly, Ocal and Osborn (2000) recently concluded that the dynamics of aggregate consumption in the UK depends on the the state of the economy and the sign of the shock.

### 2.6.4 Implications for the persistence of aggregate consumption

The degree of consumers' information frictions drive the persistence of aggregate consumption (Carroll, 2003; Mankiw and Reis, 2006; Reis, 2006a; Carroll et al., 2020). This section focuses on aggregate consumption growth and provides a mapping between the persistence of aggregate consumption and the share of inattentive consumers. It demonstrates that in periods when information frictions are low, such as recessions, consumption persistence drops as well.

To apprehend the time-varying persistence of aggregate consumption, I simulate the model 15 times for 4,000 periods[51] and estimate the following equation

$$\Delta C_t = \alpha + \gamma_t \Delta C_{t-1} + \beta X_t + error_t \tag{2.29}$$

that relates the aggregate consumption change to its lag $\Delta C_{t-1}$ and a set of controls $X_t$ including aggregate income shocks and a further lag for the change in aggregate consumption. Similar specifications are used in the literature in order to estimate the persistence of consumption (Havranek et al., 2017; Carroll et al., 2020). Generally, the persistence parameter $\gamma$ is assumed to be time-invariant. Regressing equation (2.29) on the simulated data, the estimated time-invariant persistence is equal to 0.70. This value corresponds to the (weighted) average persistence estimate for aggregate consumption reported in Havranek et al. (2017) meta-analysis for the U.S. I match this value by setting the relative variance of aggregate income shocks accordingly. The R-squared is equal to 99.83%, thus indicating that the specification in equation (2.29) is a good approximation of the

[51] I start the simulation from the steady-state cross-section of consumers, simulate the model for 5,000 periods and drop the first 1,000 periods.

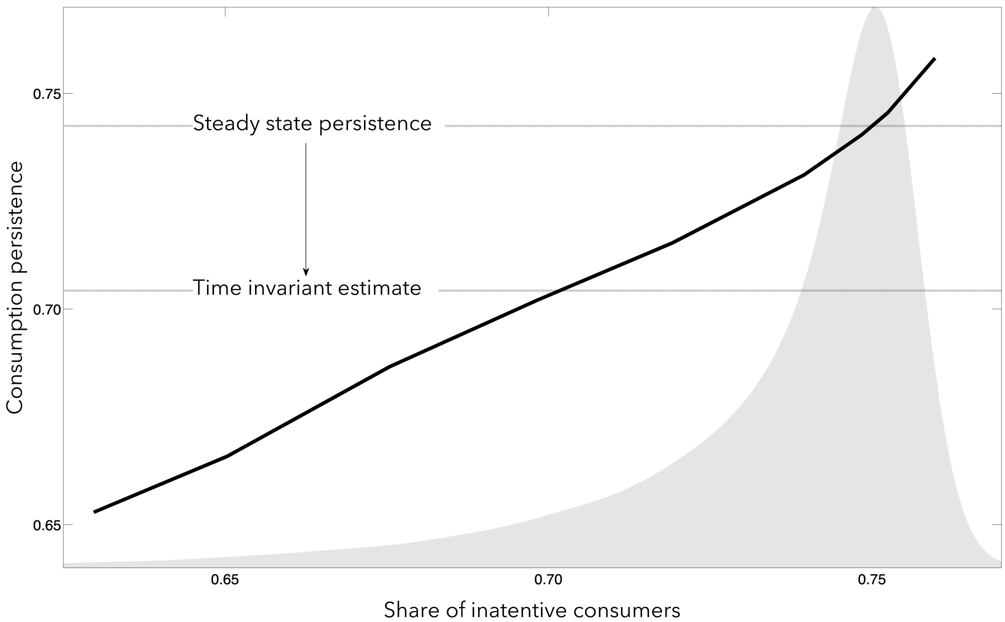
aggregate dynamics predicted by the consumption model developed in this paper.

In models with information frictions on the household's side, the degree of information frictions drives the persistence parameter $\gamma$. In the present paper, the degree of information frictions is endogenous and varies over time. Consequently, I analyze how the persistence of consumption growth depends on the endogenously time-varying share of attentive consumers. More specifically, one may want to derive a mapping between the persistence parameter and the share of attentive consumers. That is, to estimate the function $\gamma_t = \gamma(\lambda_t)$. To this end, I use interaction terms to estimate the parameters $\gamma_t$ in equation (2.29) at different percentiles of the distribution of the share of attentive consumers. This method does not impose a parametric assumption on $\gamma(.)$.

Figure 2.5 displays the predicted mapping between consumption persistence and consumers' information rigidities. We observe an increasing mapping between the persistence of consumption and the share of inattentive consumers. Accounting for this variation in the persistence parameter significantly improves the model goodness-of-fit and the R-squared increases to 99.92%. Interestingly, it also indicates that the time invariant persistence parameter estimated in the literature underestimates the steady state persistence of aggregate consumption. This underestimation is a consequence of the skewness of the distribution of information frictions and non-linearities in the mapping $\gamma(.)$.

Figure 2.5 indicates that in normal times information rigidities are near their steady state level (75%) and consumption persistence is high and relatively constant (around 0.74). However, during unusual times such as recessions, information rigidities decrease a lot and so does aggregate consumption persistence. The decrease in consumption persistence may be large (more than 10 percentage points given the model calibration). These findings are in line with Kumar and Jia (2019) who report systematic decreases in consumption growth persistence during recessionary periods. Figure 2.5 relates this change in aggregate consumption dynamics to Dräger and Lamla (2012) and Coibion and Gorodnichenko (2015) findings that information rigidities also decrease during these periods.

Figure 2.5: Aggregate consumption persistence and attention




NOTE: Model predictions for the relation between consumption persistence and information frictions. Aggregate consumption change is approximatively equal to $\Delta C_t = \gamma_t \Delta C_{t-1} + \beta X_t$ where $\gamma_t$ is a measure of the time-varying persistence of consumption growth and $X_t$ a set of controls discussed in Section 2.6. The black line relates the evolution of the estimated time-varying persistence $\gamma_t$ to the observed share of inattentive consumers from simulating the model 15 times over 4,000 periods. The grey area is the distribution of the share of inattentive consumers. The steady state persistence is the persistence when the share of inattentive consumers is at its steady state level (i.e. 75%) and the time-variant estimate is obtained from an OLS estimation when the persistence is assumed to be constant. The model is calibrated at the household level (see Section 2.4).

### 2.6.5 Related discussions

*A - Inattention, excess smoothness and serial correlation:* Consumers' inattention explains both the serial correlation (Carroll et al., 2020) and the excess smoothness of aggregate consumption (Reis, 2006a). The former is generally found to be about 0.70 - 0.75 (Havranek et al., 2017) and the excess smoothness ratio about 0.65 to 0.75 (Galí, 1993) for nondurables in the U.S. If both result from consumers' inattention, then the predictions from these models should be able to match these two characteristics of aggregate consumption. In the Sticky information model where a constant share of consumers are randomly selected, Proposition 4 in Reis (2006a) predicts an excess smoothness ratio of about 0.42 (resp. 0.38) when the share of non-updaters is 0.70 (resp. 0.75). The share of non-updaters coincides with the serial correlation (see Carroll et al. (2020)). The model presented in this paper, however, predicts a lower excess smoothness as evidenced by the data. When the steady state share of updaters is 0.75, the average serial correlation is 0.71 and the excess smoothness ratio is 0.55. While the model continues to predict a too large excess smoothness in comparison to the data, it improves the predictions from related models.[52]

*B - Attention to the interest rate:* In the model, consumers behave as if the gross real interest was constant. This raises the important question of how to think about monetary policy in this framework. Or, equivalently when there are no nominal rigidities, how would consumption respond to a change at time $t$ in the equilibrium path of present and future real interest rates $\{r_{t+s}\}_{s=0}^{\infty}$ when $\exists s : r_{t+s} \neq r$.[53] Appendix 2.15.4 considers two extensions with a time varying real interest rate. It keeps on maintaining the assumption that consumers are fully inattentive to variations in the real interest rate. It shows that policy interventions on the gross real interest rate enters through consumers' innovations in permanent income (that capture the substitution and income effects). As a result, unexpected changes in the real interest rate shift the cross-sectional distribution $a_t(e)$. The impact on aggregate consumption dynamics then follows from Proposition 6. Abstracting from the potential heterogeneity induced by asset holdings, this shock would be equivalent to an aggregate shock $\chi_t$.[54]

*C - Attention to GE effects:* The scope of this paper is limited to partial equilibrium. It therefore does not consider any potential behavioral misperception of general equilibrium (GE) effects. For instance, the Ricardian equivalence holds and forward guidance remains a puzzle here. The reason is that we maintain the permanent income (PI) hypothesis and considers that consumers are inattentive to variations in their PI. That is, they

[52]The increased excess smoothness is the consequence of the inattention region. Here, consumers do not adjust to small income shocks so that the variance of aggregate consumption change increases when the selection of attentive consumers is not random. Noisy information is known to generate excess smoothness as well (Luo, 2008), and we find that if one was to select a constant share of randomly selected attentive consumers at each period in our model with both Noisy and Sticky information, the excess smoothness ratio would be very small (0.27). The inattention region therefore drastically increases the excess smoothness ratio.

[53]It must be emphasized that the assumption of constant real interest is here for mathematical convenience and limits the scope to a partial equilibrium analysis. In Appendix 2.15.4, I propose two simple extensions that allow to think about the potential implications of the inattention region for monetary policies. Explicitly introducing a monetary authority in this framework is left for future work.

[54]Also potentially at all $t+s$ such that $r_{t+s} \neq r$ with the first extension in Appendix 2.15.4.

do not perfectly observe the innovations $\zeta_t$, but these innovations perfectly account for all the factors affecting the consumer's consumption-saving problem. Nevertheless, one may easily extend the definition of $\zeta_t$ to account for misperception of GE effects, e.g., k-level thinking (Farhi and Werning, 2019), variable-specific attention and/or discounted attention (Gabaix, 2020). The predictions of the present paper are therefore orthogonal to these papers. They focus on how people compute innovations about PI whereas, here, we are concerned about how and when people incorporate these perceived innovations. Using Gabaix's (2020) terminology, we look at the attention intercept and show that it is correlated with the perceived innovations. The predictions discussed in the present paper then follows from this novel property.

*D - Stabilizing policies:* The endogeneity of attention acts as a stabilizing force: the effect of large shocks dissipate more rapidly, thus limiting the length of large deviations from steady states dynamics. This stabilizing force is welfare improving: information-dependent attention strategies are preferable to purely time-dependent attention strategies. However, analyzing how optimal stabilizing policies, e.g., Taylor rule, interact with the time-varying and state-dependence information frictions remains an open question. Although essential, analyzing this, and related questions, is well beyond the scope of the present paper.

*E - Stimulating demand during recessions:* The paper sheds new lights on the effectiveness of discretionary fiscal and monetary policies during recessions. In particular, it demonstrates that expansionary policies are less effective in the very short run *when the economy is already in a recession.* The reason is that consumers are more likely to be attentive to further negative destabilizing shocks than to positive ones during these periods. As a consequence, feeble expansionary policies may actually result in longer recoveries than would have happened otherwise. This short run ineffectiveness vanishes for large policy interventions that are deemed attention worthy by consumers. Therefore, fixed budget fiscal policies will, generally, be more effective in boosting short run aggregate demand if they target a subpopulation of consumers. Indeed, the very short run consumption response to a \$1 transfer to every households will be much smaller than the budget equivalent policy of implementing a \$1,000 transfer to a thousandth of households.[55]

[55]These transfers should not be paid back in the future – see the remark on the Ricardian equivalence

## 2.7 Conclusions

This paper proposes a novel model to explain the state-dependence of information rigidities. Consumers must pay a fixed cost to observe noisy signals on the state of the economy. They face an inattention region where they temporarily ignore new signal releases and do not update their expectations. Accounting for the interaction between variations in information rigidities and consumption choices allows to reproduce non-linear features of consumption dynamics that have been observed in the data.

This paper stresses potentially important and novel policy implications that would be interesting to investigate. At the aggregate level, the dynamics of consumption is governed by the dynamics of information rigidities. As a consequence, households tend to pay more attention to negative income shocks than to positive income shocks during recessions. Although there is no explicit fiscal or monetary authorities in the present model, this result may hold significant consequences for economic policies aiming to boost demand during recessions. By leaning against the wind, these policies may reduce households' incentive to update their expectations. As a result, feeble expansionary policy interventions may in fact increase the persistence of consumption growth in the aftermath of a recession, and thereby result in longer recoveries. It would therefore be interesting to enrich the model with fiscal and monetary authorities in future work. Such work would be useful in any attempt to understand the interaction between economic policies, information frictions, and macroeconomic outcomes.

The optimal attention strategy of the consumer is found to be purely information-dependent. It should be emphasized that this result relies on some assumptions made along the paper. Beside the LQG framework, these include: the infinite horizon and the fact that when attentive the consumer may catch up with all information that she previously disregarded. Proposition 1 shows that when the horizon is finite, then the inattention region evolves as the consumer ages. Similarly, if one was to assume instead that the consumer was unable to catch up with the information she previously ignored,

and GE effects above. Pushing this argument, it indicates that financing the targeted \$1,000 policy by raising a \$1 lump-sum tax paid by everyone may also stimulate demand in the very short run. Targeted policies would be even more efficient in the very short run if the government was able to condition transfers on observables correlated to $e_{i,t}$.

then the time since the last update would also alter the shape of the inattention region.[56] While these extensions are unlikely to be qualitatively relevant for consumption dynamics, they may be worthy to investigate in other applications.

[56]The intuition being that not updating today would also imply that the posterior uncertainty will increase persistently.

## 2.8 Proof of Lemma 1

Start from the straightforward identity (used in, e.g., Åström (2012), Chapter 8 - Proof of Lemma 6.1)

$$\beta^T q_T s_T^2 \quad = \quad p_0 s_0^2 + \sum_{t=0}^{T-1} \left(\beta^{t+1} p_{t+1} s_{t+1}^2 - \beta^t p_t s_t^2\right) \tag{2.30}$$

since $p_T = q_T$. Moreover,

$$p_{t+1} s_{t+1}^2 = ((1+r)s_t - u_t + \zeta_{t+1})^2 p_{t+1} \tag{2.31}$$

where $u_t \equiv c_t - \bar{c}$. Further noticing from the definition of $L_t$ and the Riccati equation for $p_t$ presented on page 61 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

$$\begin{aligned} E[\beta^{t+1} p_{t+1} s_{t+1}^2 | \mathcal{I}_0] \quad = \quad & E\Big[\beta^t (u_t - L_t s_t)^2 (1 + \beta p_{t+1}) + \beta^{t+1} p_{t+1} \zeta_{t+1}^2 \\ & + \beta^{t+1} (1+r)^2 p_{t+1} s_t^2 - \beta^t (1+\beta p_{t+1}) L_t^2 s_t^2 - \beta^t u_t^2 \Big| \mathcal{I}_0\Big] \end{aligned} \tag{2.32}$$

because $\zeta_{t+1}$ is independent with respect to $u_t$ and $s_t$. Moreover,

$$\beta^t p_t s_t^2 \quad = \quad \beta^t (1+r) L_t s_t^2 \tag{2.33}$$

so that equation (2.30) writes in expectation

$$E[\beta^T q_T s_T^2 | \mathcal{I}_0] \quad = \quad E\Big[p_0 s_0^2 + \sum_{t=0}^{T-1} \beta^t (u_t - L_t s_t)^2 (1 + \beta p_{t+1}) + \beta^{t+1} p_{t+1} \zeta_{t+1}^2 - \beta^t u_t^2 \Big| \mathcal{I}_0\Big] \tag{2.34}$$

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

$$V_0 \quad = \quad E\Big[p_0 s_0^2 + \sum_{t=0}^{T-1} \beta^t \lambda \tau_t + \beta^{t+1} p_{t+1} \zeta_{t+1}^2 + \beta^t (u_t - L_t s_t)^2 (1 + \beta p_{t+1}) \Big| \mathcal{I}_0\Big] \tag{2.35}$$

When the triggering choices $\tau_t$ are predetermined (or, equivalently here, independent of the control law), it is optimal to set $u_t = L_t E[s_t | \overline{\mathcal{I}}_t]$. Lemma 2 in Molin and Hirche

(2010) proves that this a dominating strategy, i.e., that it is also the solution to the initial problem when the triggering choices are not predetermined. (Q.E.D. Lemma 1).

## 2.9 Proof of Lemma 2

Using this result, the last term in equation (2.35) writes $\beta^t L_t^2(E[s_t|\overline{\mathcal{I}}_t] - s_t)^2(1+\beta p_{t+1})$ where $s_t - E[s_t|\overline{\mathcal{I}}_t] = s_t - E[s_t|\mathcal{I}_t] + e_t$ and $e_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t]$. Thus,

$$E[(s_t - E[s_t|\overline{\mathcal{I}}_t])^2|\mathcal{I}_0] = E[(s_t - E[s_t|\mathcal{I}_t])^2|\mathcal{I}_0] + E[e_t^2|\mathcal{I}_0] \tag{2.36}$$

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

$$V_0 = E\Big[p_0 s_0^2 + \sum_{t=0}^{T-1} \beta^t \lambda \tau_t + \beta^{t+1} p_{t+1} \zeta_{t+1}^2 + \beta^t L_t^2 \Big((s_t - E[s_t|\mathcal{I}_t])^2 + e_t^2\Big)(1+\beta p_{t+1})\Big|\mathcal{I}_0\Big] \tag{2.37}$$

Thanks to the additivity of the above equation, we can now characterize the optimal estimator $E[s_t|\mathcal{I}_t]$. Make the educated guess that it is a Kalman filter which admits a steady state variance. Then, the steady state posterior 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 $K = p_+(p_+ + \sigma_\eta^2)^{-1}$ and the prior steady state variance is $p_- = (1-K)p_+$.

We focus on situations such that the initial uncertainty surrounding the state variable is initially at its steady state value. That is, we impose $\sigma_{s_0}^2 = p_-$. Then, $E[(s_t - E[s_t|\mathcal{I}_t])^2|\mathcal{I}_0] = p_-$ is at its steady state. Therefore, the optimal estimator must minimize this steady state variance given the linear law of motion for $s_t$. By definition, this estimator is the Kalman filter, thus confirming our guess when it admits a steady state variance (Q.E.D. Lemma 2).

## 2.10 Proof of Lemma 3

The corrective term in equation (2.6) is a predetermined bias that depends on the information in the hands of the consumer when inattentive ($\overline{\mathcal{I}}_t, \tau_t = 0$). It therefore depends on time $t$, the inattention length and the triggering law $g_t(.)$ (see Lemma 4 in Molin and

Hirche (2010)). Let $l_t \equiv \sup\{k : \tau_k = 1, k \leq t\}$ be the most recent period when the consumer was attentive. Equation (2.6) then writes

$$E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0] \quad = \quad E[s_t|\overline{\mathcal{I}}_{t-1}] + \alpha(t, l_t) \tag{2.38}$$

where $\alpha(t, l_t) \equiv E\Big[(1+r)e_{t-1} + K(z_t - E[s_t|\mathcal{I}_{t-1}])\Big|\overline{\mathcal{I}}_t, \tau_t = 0\Big]$. Using the definition for $e_t$ in (2.7) we have

$$e_{t+1} = (1-\tau_t)(1+r)e_t - \alpha(t, l_t) + K\Big(z_{t+1} - (1+r)E[s_t|\mathcal{I}_t] + c_t - \bar{c}\Big) \tag{2.39}$$

Moreover, note that only the second and fourth terms in (2.37) depend on the triggering law $g_t(.)$. Hence, the triggering law solves the following problem

$$\begin{aligned} \min_{g_.(.),\alpha(.)} \quad & E\Big[\sum_{t=0}^{T-1} \beta^t \lambda \tau_t + \beta^t L_t^2 e_t^2 (1 + \beta p_{t+1})\Big|\mathcal{I}_0\Big] \\ \text{s.t.} \quad & e_{t+1} = (1-\tau_t)(1+r)e_t - \alpha(t, l_t) + \omega_{t+1} \end{aligned} \tag{2.40}$$

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 $\sigma_\omega^2 = K^2(\bar{p}_+ + \sigma_\vartheta^2)$. The difficulty in solving problem (2.40) is that $\alpha(.)$ depends on $g_.(.)$ and vice versa.

Molin and Hirche (2012) develop an iterative algorithm to solve a similar problem when $\beta = 1$. They show that when the distributions of $e_0$ and $\{\omega_t\}$ are symmetric and unimodal, $\alpha(.) = 0$ is a globally asymptotically stable fixed-point of the algorithm (Theorem 5, Molin and Hirche (2012)). The Online Appendix 2.15.1 shows that this theorem still holds for (2.40). (Q.E.D. Lemma 3)

## 2.11 Optimal triggering rule

Note that only the second and fourth terms in (2.37) depend on the triggering law $g(.)$. Hence, the updating behavior solves the following problem

$$\min_{\{\tau_t\}_{t=0}^{T-1}\in\{0,1\}^T} E\Big[\sum_{t=0}^{T-1}\beta^t\lambda\tau_t + \beta^t L_t^2 e_t^2(1+\beta p_{t+1})\Big|\mathcal{I}_0\Big] \tag{2.41}$$

along with a transitory dynamics for $e_{t+1}$. Following from Lemma 3, the law of motion for $e_{t+1}$ is at the optimum

$$e_{t+1} = (1-\tau_t)(1+r)e_t + K\Big(z_{t+1} - (1+r)E[s_t|\mathcal{I}_t] + c_t - \bar{c}\Big) \tag{2.42}$$

Problem (2.41) along with the law of motion (2.42) is a standard optimal control problem with perfect state observation $e_t$ and could therefore be solved using a DP algorithm.

From problem (2.41) the choice to update depends on the state variable $e_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t]$. Let the cost associated to the terminal condition $q_T$ be arbitrarily large. Then, at period $T-1$, the consumer updates almost surely as $\lim_{q_T\mapsto\infty} L_{T-1}^2(1+\beta q_T) = \infty$. Therefore, $g_{T-1}(e_{T-1}) = 1 \iff |e_{T-1}| > 0$. Let $\pi_t^+$ and $\pi_t^-$ denote the thresholds such that the consumer updates at period $t$ if and only if $e_t \leq \pi_t^-$ or $e_t \geq \pi_t^+$. Then, $\pi_{T-1}^+ = -\pi_{T-1}^- = 0$ and the triggering law is indeed symmetric at period $T-1$. Writing problem (2.41) in its Bellman form, we have

$$\begin{aligned} J_t(e_t) \quad &= \quad \min_{\tau_t\in\{0,1\}} L_t^2(1+\beta p_{t+1})e_t^2 + \tau_t\lambda + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t] \\ \text{s.t.} \quad & \quad e_{t+1} = (1-\tau_t)(1+r)e_t + K\Big(z_{t+1} - (1+r)E[s_t|\mathcal{I}_t] + c_t - \bar{c}\Big) \end{aligned} \tag{2.43}$$

We thus have $E[J_{T-1}(e_{T-1})|\mathcal{I}_{T-2}] = \lambda$ and the consumer updates if and only if

$$|e_{T-2}| \geq \frac{1}{L_{T-2}}\sqrt{\frac{\lambda}{1+\beta p_{T-1}}} \tag{2.44}$$

Again, the thresholds are symmetric $-\pi_{T-2}^- = \pi_{T-2}^+$. This symmetry arises because $J_{T-1}(e_{T-1}) = J_{T-1}(-e_{T-1})$ is symmetric, the expectation is taken over an unimodal and symmetric distribution so that $E[J_{T-1}(e_{T-1})|\mathcal{I}_{T-2}, e_t] = E[J_{T-1}(e_{T-1})|\mathcal{I}_{T-2}, -e_t]$ and

$L_t^2(1+\beta p_{t+1})e_t^2$ is symmetric as well. Therefore, iterating backward and using the same argument, $J_t(e_t) = J_t(-e_t)\ \forall t \in \{0,\ldots,T-1\}$ so that $-\pi_t^- = \pi_t^+\ \forall t \in \{0,\ldots,T-1\}$ thus confirming the guess that the optimal triggering law $g_t(e_t)$ is symmetric. Moreover, at any time $t \in \{0,\ldots,T-1\}$, $\pi_t \in \mathbb{R}^+$ solves

$$\lambda + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = 0] \;=\; L_t^2(1+\beta p_{t+1})\pi_t^2 + \beta E[J_{t+1}(e_{t+1})|\mathcal{I}_t, e_t = \pi_t] \quad (2.45)$$

(Q.E.D. Lemma 4 and Proposition 1)

## 2.12 Stationary policies

I now demonstrate that the triggering law converges to a stationary policy when $T$ tends to $\infty$. The first step is to show that $p_t$ converges to a stationary solution. To this end, consider the following infinite horizon deterministic linear quadratic control problem:

$$\begin{aligned} \min_{\{c_t\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t (c_t - \bar{c})^2 \\ \text{s.t.} \quad x_{t+1} &= (1+r)x_t - c_t + \bar{c} \end{aligned} \quad (2.46)$$

where $\beta \in (0,1)$. Assuming it exists,[57] 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

$$\bar{p} \;=\; (1+r)^2 \frac{\beta\bar{p}}{1+\beta\bar{p}} \quad (2.47)$$

that is, $\bar{p} = \frac{\beta(1+r)^2-1}{\beta}$.[58] Thanks to the certainty equivalence of the orignal problem, the consumption policy (2.4) admits a stationary solution $L = (1+r)\frac{\beta\bar{p}}{1+\beta\bar{p}} = \frac{\beta(1+r)^2-1}{\beta(1+r)}$. As a result, $L_t^2(1+\beta p_{t+1})$ converges to $\frac{[\beta(1+r)^2-1]^2}{\beta(1+r)}$ so that the reward function in the Bellman equation (2.9) is stationary. The latent Kalman filter being at its steady state, $z_{t+1} - E[s_{t+1}|\mathcal{I}_t]$ follows a gaussian distribution with mean zero and variance $p_+ + \sigma_\eta^2$ which is time-invariante. Hence, problem (2.9) is an infinite horizon discrete time Markov decision

[57] 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.

[58] These results also follow directly from Lemma 1 since the certainty equivalence holds.

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

$$\begin{aligned} J(e_t) &= \min_{\tau_t \in \{0,1\}} \frac{\Big(\beta(1+r)^2-1\Big)^2}{\beta(1+r)} e_t^2 + \tau_t \lambda + \beta E[J(e_{t+1})|\mathcal{I}_t] \qquad (2.48) \\ \text{s.t.} \quad & e_{t+1} = (1-\tau_t)e_t + k\Big(z_{t+1} - (1+r)E[s_{t+1}|\mathcal{I}_t] + c_t - \bar{c}\Big) \end{aligned}$$

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

$$\lambda + \beta E[J(e_{t+1})|\mathcal{I}_t, e_t = 0] = \frac{\Big(\beta(1+r)^2-1\Big)^2}{\beta(1+r)} \pi^2 + \beta E[J(e_{t+1})|\mathcal{I}_t, e_t = \pi] \qquad (2.49)$$

and $J(.)$ is the stationary value function from (2.48).

## 2.13 Cross-sectional stationary distribution

Take the limit when $T \mapsto \infty$ and let $\lambda_t(k)$ be the share of consumers who did not update for $k \in \{1, 2, \dots, \infty\}$ periods. Then,

$$a_t(e) = \sum_{k=2}^{\infty} \lambda_{t-1}(k-1)(1-\Lambda_t(k-1)) f_t(e|k) + f_t(e|1)\Big(\sum_{k=1}^{\infty} \lambda_{t-1}(k)\Lambda_t(k)\Big) \qquad (2.50)$$

The stationary distribution being time independent, the $\lambda^\star(k)$ must solve

$$\begin{aligned} \lambda^\star(1) &= \sum_{k=1}^{\infty} \lambda^\star(k)\Lambda^\star(k) && (2.51) \\ \lambda^\star(k) &= \lambda^\star(k-1)(1-\Lambda^\star(k-1)) \quad \forall k \geq 2 && (2.52) \\ 1 &= \sum_{k=1}^{\infty} \lambda^\star(k) && (2.53) \end{aligned}$$

where $\Lambda^\star(k) = 1 - \int_{\Xi} f(e|k)de$ and $\Xi$ is the time invariant non-updating set following from corollary 1 and $f(e|k)$ the corresponding time invariant[59] distribution obtained from

[59] It is straightforward to see that the latter distribution is time invariant when the non-updating set is time invariante.

iterating on equation (2.15). Iterating backward, equation (2.52) writes

$$\lambda^\star(k) \;=\; \lambda^\star(1)\prod_{i=1}^{k-1}(1-\Lambda^\star(i)) \quad \forall k \geq 2 \tag{2.54}$$

Noting that $S^\star(k-1) = \prod_{i=1}^{k-1}(1-\Lambda^\star(i))$ where $S^\star(k)$ is the time invariante survival function and $S^\star(0)=1$, we may introduce this expression in (2.53) to get

$$\lambda^\star(1) \;=\; \frac{1}{\sum_{k=1}^{\infty} S^\star(k-1)} \tag{2.55}$$

Further noticing that equation (2.51) holds independently of $\lambda^\star(1)$ because $q_k^\star \equiv S(k-1)\Lambda^\star(k)$ and $\sum_{k=1}^{\infty} q_k^\star = 1$, equations (2.55), (2.54) and (2.15) fully characterize the stationary cross-sectional distribution

$$a^\star(e) \;=\; \sum_{k=1}^{\infty}\lambda^\star(k) f^\star(e|k) \tag{2.56}$$

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. In the simulations, the (evenly discretized) stationary distribution is first computed using lemma 5 as a first guess and then iterated on a few times following (2.20) to achieve convergence.

## 2.14 Welfare decomposition

Taking the infinite horizon version of equation (2.37), the value function at period 0 writes

$$V_0 \;=\; E\Big[ps_0^2 + \frac{p}{1-\beta}\sigma_\zeta^2 + \frac{L^2(1+\beta p)}{1-\beta}p_- + \sum_{t=0}^{\infty}\beta^t\Big(L^2(1+\beta p)e_t^2 + \lambda\tau_t\Big)\Big|\mathcal{I}_0\Big] \tag{2.57}$$

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 processing information.

Equation (2.57) 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. Further, 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 5. 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,

$$E_{a^\star(.)}[V_0(e_0)] \quad = \quad p(\bar{s}_0^2 + p_-) + \frac{p}{1-\beta}\sigma_\zeta^2 + \frac{L^2(1+\beta p)}{1-\beta}(p_- + \sigma_a^2) + \frac{\lambda\bar{\lambda}^\star}{1-\beta} \tag{2.58}$$

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:

$$\text{WC} \quad = \quad \frac{r(1-\beta)^{-1}}{2(\bar{c}-\mu)(1+r)}\Big[\frac{[\beta(1+r)^2-1]^2}{\beta(1+r)}(p_- + \sigma_a^2) + \lambda\bar{\lambda}^\star\Big] \tag{2.59}$$

## 2.15 Complementary results

### 2.15.1 On the optimality of $\alpha(.) = 0$

In this appendix, I show that Theorem 5 in Molin and Hirche (2012) applies to the problem considered in Appendix 2.10. 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, I recast the problem in Appendix 2.10 using the notation used in their proof.[60] Let

$$\hat{e}_t \equiv E[s_t|\mathcal{I}_t] - E[s_t|\overline{\mathcal{I}}_t, \tau_t = 0] + \alpha(t, l_t) \tag{2.60}$$

Accordingly, problem (2.40) can be written as

$$\begin{aligned} \min_{g.(.),\alpha(.)} \quad & E\Big[\sum_{t=0}^{T-1} \beta^t \Big((1-\tau_t)\Gamma_t(\hat{e}_t - \alpha(t, l_t))^2 + \lambda\tau_t\Big)\Big|\mathcal{I}_0\Big] \\ \text{s.t.} \quad & \hat{e}_{t+1} = (1-\tau_t)(1+r)\hat{e}_t + \omega_{t+1} \end{aligned} \tag{2.61}$$

where $\Gamma_t \equiv L_t(1 + \beta p_{t+1})$. Moreover, let

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

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

$$\hat{c}_t^{\varrho_t}(\hat{y}_t, l_t, \tau_t) = \beta^t \Big((1-\tau_t)R^{2t}\Gamma_t(\hat{y}_t - \varrho_{t,l_t})^2 + \lambda\tau_t\Big) \tag{2.62}$$

The optimization problem for Molin and Hirche (2012) iterative procedure is thus given by

$$\min_{\hat{g},\varrho} \hat{J} \tag{2.63}$$

[60] The original proof of Theorem 5 in Molin and Hirche (2012) can be found here: https://arxiv.org/pdf/1203.4980.pdf .

with

$$\hat{J}(\hat{g},\varrho) = E_{\hat{g}}\left[\sum_{t=0}^{N-1} \hat{c}_t^{\varrho_t}(\hat{y}_t, l_t, \tau_t)\right] \tag{2.64}$$

where the subscript $\hat{g}$ emphases that the expectation is taken with respect to the triggering policy. The proof in Molin and Hirche (2012) requires that, for a fixed vector $\varrho^i$ of all combinaisons $\varrho_{t,l_t}$, the following symmetry and monotonicity properties hold for the running cost:

$$\begin{aligned} \hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i + \Delta, l_t, \tau) &= \hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i - \Delta, l_t, \tau) \\ &\quad \forall \Delta \in \mathbb{R}, l_t \in \{0,\dots,t-1\}, \tau \in \{0,1\} \end{aligned} \tag{2.65}$$

and

$$\begin{aligned} 0 \le \Delta_1 \le \Delta_2 \Longrightarrow \quad \hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i + \Delta_1, l_t, \tau) &\le \hat{c}_t^{\varrho_t^i}(\varrho_{t,l_t}^i + \Delta_2, l_t, \tau) \\ &\quad \forall l_t \in \{0,\dots,t-1\}, \tau \in \{0,1\} \end{aligned} \tag{2.66}$$

It is straightforward to see that these properties hold here given (2.62). As a result, the subsequent results in the proof in Molin and Hirche (2012) are valid and their Theorem 5 applies.

### 2.15.2 Approximating the distribution $f_t(e|k, e_{t-k})$

To avoid confusion in the following, let the realization $e_{t-k} = a$ and recall that $\sigma_\omega^2 = k^2(\bar{p}_+ + \sigma_\eta^2)$ where $k$ and $p_+$ are respectively the gain and the one period ahead error variance at the Kalman filter steady state. Define $f_t(e|0,a) = \delta(e-a)$. Then from iterating on (2.15), we have for k=1:

$$f_t(e|1,a) = \frac{1}{\sigma_\omega}\phi\left(\frac{e-(1+r)a}{\sigma_\omega}\right) \tag{2.67}$$

When $k = 2$,

$$\begin{aligned} f_t(e|2,a) &\propto \int_{\Xi_{t-1}} \frac{1}{\sigma_\omega^2}\phi\left(\frac{e-(1+r)\bar{e}}{\sigma_\omega}\right)\phi\left(\frac{\bar{e}-(1+r)a}{\sigma_\omega}\right)d\bar{e} \\ &\propto \int_{\Xi_{t-1}} \frac{1}{2\pi\sigma_\omega^2}\exp\left\{-\frac{e^2 - 2(1+r)\bar{e}e + (1+r)^2\bar{e}^2 + \bar{e}^2 - 2(1+r)\bar{e}a + (1+r)^2a^2}{2\sigma_\omega^2}\right\}d\bar{e} \end{aligned} \tag{2.68}$$

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

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

Where

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

Therefore, (2.68) writes

$$
\begin{aligned}
f_t(e|2,a) \;\propto\;& \int_{\Xi_{t-1}} \frac{1}{2\pi\sigma_\omega^2}\exp\Big\{-\frac{(1+R^2)}{2\sigma_\omega^2}\Big[\Big(\bar{e} - \frac{R}{1+R^2}(e+a)\Big)^2 + \Big(\frac{1}{1+R^2}e - \frac{R^2}{1+R^2}a\Big)^2\Big]\Big\}d\bar{e} \\
\propto\;& \int_{\Xi_{t-1}} \frac{\sqrt{1+R^2}}{\sqrt{2\pi}\sigma_\omega}\exp\Big\{-\frac{(\bar{e}-\frac{R}{1+R^2}(e+a))^2}{2\frac{\sigma_\omega^2}{1+R^2}}\Big\}\frac{1}{\sqrt{2\pi}\sqrt{1+R^2}\sigma_\omega}\exp\Big\{-\frac{(e-R^2a)^2}{2(1+R^2)\sigma_\omega^2}\Big\}d\bar{e} \\
\propto\;& \int_{\Xi_{t-1}} \frac{\sqrt{1+R^2}}{\sigma_\omega}\phi\Big(\frac{\bar{e}-\frac{R}{1+R^2}(e+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\Big)\frac{1}{\sqrt{1+R^2}\sigma_\omega}\phi\Big(\frac{e-R^2a}{\sqrt{1+(1+r)^2}\sigma_\omega}\Big)d\bar{e} \\
\propto\;& \frac{1}{\sqrt{1+R^2}\sigma_\omega}\Big[\Phi\Big(\frac{\pi_{t-1}-\frac{R}{1+R^2}(e+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\Big) - \Phi\Big(-\frac{\pi_{t-1}+\frac{R}{1+R^2}(e+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\Big)\Big]\phi\Big(\frac{e-R^2a}{\sqrt{1+R^2}\sigma_\omega}\Big)
\end{aligned}
\tag{2.69}
$$

When $k = 3$,

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

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 (2.70), I obtain

$$\begin{aligned}
f_t(e|3,a) \;\propto\; & \frac{1}{\sqrt{1+R^2}\sigma_\omega^2}\phi\left(\frac{e-R^3a}{\sqrt{1+R^2+R^4}\sigma_\omega}\right)\times \qquad (2.70) \\
& \int_{\Xi_{t-1}}\phi\left(\frac{\bar{e}-\frac{R(1+R^2)e+R^2a}{1+R^2+R^4}}{\sqrt{\frac{1+R^2}{1+R^2+R^4}}\sigma_\omega}\right)\left[\Phi\left(\frac{\pi_{t-2}-\frac{R}{1+R^2}(\bar{e}+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\right)-\Phi\left(-\frac{\pi_{t-2}+\frac{R}{1+R^2}(\bar{e}+a)}{\frac{\sigma_\omega}{\sqrt{1+R^2}}}\right)\right]d\bar{e}
\end{aligned}$$

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

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

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

The Online Appendix discusses the implications to the proposed approximation procedure. It 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 updates is however negligible as the proportion of agents who encounters a large $k$ is small.

### 2.15.3 Excess smoothness ratio

When $\beta(1+r) = 1$, the full information rational expectation (FIRE) version of the model implies (Hall, 1978)

$$\Delta C^\star_{t+1} = r\chi_{t+1}$$

That is, the innovation of aggregate consumption to a one dollar aggregate shock to permanent income is $r$ in the FIRE model. Hence, the standard deviation of aggregate consumption changes writes

$$\sigma^\star(\Delta C) = r\sigma_\chi$$

Following Galí (1993), the excess smoothness ratio is $\Psi = \sigma(\Delta C)/\sigma^\star(\Delta C)$ where $\sigma(\Delta C)$ is the standard deviation of aggregate consumption changes predicted by the model. When this ratio is lower than one, consumption is said to be excessively smooth.

### 2.15.4 Inattention to the real interest rate

In the model, consumers are perfectly inattentive to the real interest rate. They anchor their perceived rate to its steady state value and never realize that it may slightly deviate from this value. However, these deviations in the real interest rate will transpire through their true budget constraint and innovations to permanent income. In this appendix, I briefly present two extensions that illustrate how shocks to the real interest may be introduced in the framework presented in the main paper. These extensions unsure that the consumer's budget constraint for permanent income is always correct. They differ in that the first extension assumes that the instantaneous budget constraint is always correct, while the second assumes that the intertemporal budget constraint is always correct. Both shows that consumers will learn about the variations in the interest rate through the signals she receives about innovations to permanent income.

The effect of time varying real interest on human wealth is trivial and follows from the definition of the innovation to permanent income. Recall that $\zeta_{t+1} \equiv \sum_{k=t+1}^{T-1}(1+r)^{t-k}\Big(E_{t+1} - E_t\Big)[y_k]$ in the model with a constant real interest rate. Introducing a time varying real interest rate $r_t$, it follows that

$$\tilde{\zeta}_{t+1} \equiv \sum_{k=t+1}^{T-1} \Big(E_{t+1} - E_t\Big)\Big[\frac{y_k}{\prod_{i=t+1}^{k}(1+r_i)}\Big]$$

*Correct instantaneous budget constraint:* The simplest way to introduce time varying real interest rates is to impose that the consumer's budget constraint is binding at each period. The consumer's true budget constraint is

$$\tilde{s}_{t+1} = (1 + r_{t+1})s_t - (c_t - \bar{c}) + \tilde{\zeta}_{t+1}$$

However, the consumer's perceived budget is slightly different since she is inattentive to variations in the real interest rate. Her perceived budget constraint writes

$$s_{t+1} = (1+r)s_t - (c_t - \bar{c}) + \hat{\zeta}_{t+1}$$

where the shock $\hat{\zeta}_{t+1}$ must be defined. The true and perceived budget constraints coincide at each period whenever $\hat{\zeta}_{t+1} \equiv \hat{r}_{t+1}s_t + \tilde{\zeta}_{t+1}$ with $\hat{r}_{t+1} \equiv r_{t+1} - r$ the deviation of the real interest rate with respect to its steady state value. Consumers then use their perceived model with constant real interest rate to derive the distribution of $\hat{\zeta}_{t+1}$ and they consequently interpret it as an innovation to permanent income $\sim \mathcal{N}(0, \sigma^2_{\zeta})$.

*Correct intertemporal budget constraint:* When we impose that only the instantaneous budget constraint is correct, the perceived and true intertemporal budget constraints may differ. For instance, consider the impact of an expected future change in the real interest rate in one period. Then, with the above extension, a consumer who perfectly observes the shocks $\hat{\zeta}_t$ at each period will adjust twice to the announcement. First, when the announcement is made through the impact that it has on the expected present value of her human wealth $\tilde{\zeta}_t$, and then when the policy change occurs through $\hat{r}_{t+1}s_t$. However, this latter change was predictable at time $t$ from the forward looking intertemporal budget constraint. Its expected present value at time $t$ is $E_t[\hat{r}_{t+1}s_t/(1+r_{t+1})]$. Iterating forward, the equality between the perceived and true forward looking intertemporal budget constraints implies that

$$\hat{\zeta}_{t+1} \equiv \nu_{t+1} + \tilde{\zeta}_{t+1}$$

with

$$\nu_{t+1} \equiv \sum_{k=t+1}^{T-1} \Big(E_{t+1} - E_t\Big)\Big[\frac{\hat{r}_k s_k}{\prod_{i=t+1}^{k}(1+r_i)}\Big]$$

Both approaches predict that the consumer behaves optimally given her perceived model and that the perceived prior distribution of innovations $\zeta_t$ differs from the actual distribution of $\hat{\zeta}_t$. An inattentive consumer, nevertheless, never observes the innovations $\hat{\zeta}_t$ but noisy signals $z_t = \hat{\zeta}_t + \vartheta_t$. She could then try to infer from these signals. She could, for instance, use bayes rule to update her prior about the distribution of $\zeta$ in her perceived model.[61] A weaker restriction could be that the consumer should not observe any salient pattern in, say, the correlation of signals over time. In this case, one would favor the second extension as it predicts that the $\hat{\zeta}_t$ truly are innovations, that is, they are uncorrelated over time even when the dynamics of $\hat{r}_t$ are serially correlated.

[61] If the posterior remains Gaussian, then she will adjust her attention accordingly. This bayesian updating might nonetheless result in a non Gaussian posterior in the general case. Analyzing the optimal inattentiveness of the consumer in this situation is beyond the scope of this paper which focuses on the LQG setup.

# Chapter 3

# Inattention and the Taxation Bias

*This chapter is a joint work with Antoine Ferey (École Polytechnique – CREST).*

**Abstract:** This paper studies how information frictions in agents' tax perceptions affect the design of actual tax policy. Developing a positive theory of tax policy, we show that agents' inattention interacts with policymaking and induces the government to implement inefficiently high tax rates: this is the taxation bias. We quantify the magnitude of this policy distortion for the US economy. Overall, our findings suggest that existing information frictions – and thereby tax complexity – lead to undesirable, large and regressive tax increases.[62]

Keywords: Optimal taxation; inattention.
JEL classification code: H21; D90.

[62]We are grateful to Pierre Boyer, Philippe Choné, Allan Drazen, Xavier Gabaix, Bas Jacobs, Laurence Jacquet, Etienne Lehmann, Benjamin Lockwood, Jean-Baptiste Michau, Xavier Ragot, Anasuya Raj, Daniel Reck, Alessandro Riboni, Emmanuel Saez, Stefanie Stantcheva, Dmitry Taubinsky, Clémence Tricaud, Nicolas Werquin and Mirko Wiederholt for their helpful and constructive comments in improving the paper. We also wish to thank audiences at AFSE 2018 Conference, CREST, CRED Taxation Group, EEA 2019 Conference, IIPF 2019 Conference, LAGV 2018 Conference, NTA 2018 Conference, PET 2017 Conference and UC Berkeley for their discussion and feedback. Boccanfuso gratefully acknowledges the financial support of Labex OSE and Ferey that of Labex ECODEC (ANR-11-LABX-0047). Ferey is also grateful for the hospitality of UC Berkeley. This paper was awarded the *ITAX Best PhD Paper Award* at the 2019 IIPF Conference.

## 3.1 Introduction

A growing body of evidence documents substantial information frictions in agents' tax perceptions (Chetty, 2015; Bernheim and Taubinsky, 2018; Stantcheva, 2019). In particular, taxpayers tend to partially ignore non-salient taxes and transfers (Chetty et al., 2009; Miller et al., 2015; Taubinsky and Rees-Jones, 2017), to rely on linearizing heuristics (Liebman and Zeckhauser, 2004; Rees-Jones and Taubinsky, 2020) and to misunderstand some characteristics of income tax schedules (Saez, 2010; Aghion et al., 2017). Taken together, these findings indicate that agents' tax perceptions are shaped by their attention to taxes and potential behavioral biases. f

In light of this evidence, a burgeoning normative literature analyzes the design of optimal tax policy in the presence of information frictions in agents' tax perceptions.[63] This literature characterizes optimal tax policies in terms of sufficient statistics that capture agents' earnings responses to tax changes and perception biases at the optimum. Doing so, it generally sidesteps the issue that agents' tax perceptions may adjust to changes in tax policy and remains agnostic about the mechanisms behind these adjustments.[64] In their general treatment of optimal taxation with behavioral agents, Farhi and Gabaix (2020, p. 13) emphasize that "a difficulty confronting all behavioral policy approaches is a form of Lucas critique: how do the underlying biases change with policy?".

In practice, tax policy is likely influenced by the way agents' perceptions adjust to tax changes. Policymakers may for instance be tempted to increase taxes if agents are inattentive and only perceive a fraction of tax increase. In contrast to their normative counterparts, such positive policy implications remain surprisingly unexplored. This paper aims at filling this gap by studying how information frictions in agents' tax perceptions affect the design of actual tax policy.

We develop a positive theory of tax policy in a setting where agents' labor supply is determined by their tax perceptions. We show that the adjustment of agents' tax

[63]For instance, Goldin (2015) shows that a government may implement non-salient taxes to reduce the deadweight loss of taxation. Gerritsen (2016) highlights that tax misperceptions introduce a new corrective motive for taxation and derives adjusted optimal tax formulas. Integrating both insights Allcott et al. (2018) revisit the Atkinson and Stiglitz (1976) result when commodity taxes are not salient.

[64]This is at the essence of the sufficient statistics approach. See Chetty (2009) or Kleven (2020) for a general discussion and Reck (2016) for a discussion in a behavioral context.

perceptions interacts with policymaking and generates a distortion in actual tax policy. Specifically, we show that inattention leads the government to implement inefficiently high tax rates: this is the taxation bias. The key insight is that inattention creates the illusion that earnings responses to tax reforms are lower than they actually are, thereby inducing a commitment problem in the choice of tax policy.

We then quantify the magnitude of this policy distortion through a simple sufficient statistics formula that we bring to the data. We further illustrate our theoretical results using numerical simulations which shed a new light on the implications of inattention and misperceptions. Overall, our findings suggest that existing information frictions lead to undesirable, large and regressive tax increases.

Our theoretical framework considers a population of heterogeneous and rationally inattentive agents who choose their earnings and consumption given their tax perceptions.[65] We model agents' tax perceptions as resulting from a Bayesian learning model with a choice of information (Mackowiak et al., 2018; Gabaix, 2019). That is, taxpayers are endowed with a prior (or belief) about tax policy and can collect additional, but costly, information in the form of a signal. The precision of this signal is endogenous: the more attentive a taxpayer is, the more precise her signal and the more accurate her posterior (or perception). As a result, agents' perceived tax rate is in expectation given by a weighted average between their prior and the actual tax rate where the weight on the latter captures agents' attention to tax policy. Importantly, we allow the prior to be systematically biased to capture potential perception biases thereby building a bridge between behavioral models with ad-hoc misperceptions and standard rational inattention frameworks. This model thus captures the use of biased rule-of-thumbs as default while allowing taxpayers to improve their tax perceptions if they find optimal to do so (Morrison and Taubinsky, 2019).

[65] The taxation bias follows from the presence of inattention, be it endogenous or exogenous. We nonetheless adopt a rational inattention model given the strong empirical support for the endogeneity of attention and this model in particular. For instance, Hoopes et al. (2015) find that rational inattention motives and shocks to tax salience drive taxpayers' online information search. Taubinsky and Rees-Jones (2017) show in a shopping experiment that tripling the tax rate nearly doubles agents' attention to taxes. Morrison and Taubinsky (2019) provide further compelling evidence that observed attention patterns are consistent with theoretical predictions of rational inattention models.

Building on this general tax perception model, we develop a positive theory of tax policy that we formalize as a simultaneous game between rationally inattentive agents and a welfarist government.[66] Agents endogenously choose their attention to taxes and the government sets tax policy to maximize social welfare taking attention choices into account. In equilibrium, (i) neither taxpayers nor the government has an incentive to deviate, and (ii) taxpayers' actions and perceptions are mutually consistent with the government's choice of tax policy. Our main result is that – irrespective of potential perception biases – inattention leads to the implementation of inefficiently high tax rates.

Central to this result is a dichotomy between direct and indirect adjustments in perceptions upon changes in tax policy. Indeed, as agents' tax perceptions are determined by a combination of the actual tax rate and their prior, there are two margins through which perceptions may adjust: a *direct* margin capturing the attention agents devote to observing taxes and thus changes in tax policies, and an *indirect* margin capturing variations in the prior. For a given prior, inattentive agents only perceive a fraction of the change in tax policy which dampens their earnings responses. The government thus targets a higher tax rate than if agents were perfectly attentive. In equilibrium, agents' priors must however be consistent with the government's choice of tax policy. As a result, ex post earnings responses are larger than what anticipated ex ante. The government implements inefficiently high tax rates because it fails to internalize the indirect adjustment of the prior (arising as an equilibrium mechanism) in its choice of tax policy. In a nutshell, taxpayers' inattention to taxes creates the illusion that tax reforms induce lower efficiency costs than they actually do and ultimately prompts the government to misbehave from a normative perspective.

Fundamentally, this reflects a commitment problem. By implicitly restricting the set of tax policies to precommited policy rules, the aforementioned normative literature characterizes tax policy under commitment. This commitment tax policy is by definition the welfare-maximizing tax policy in the presence of information frictions in agents' tax perceptions. However, a side effect of information frictions is that actual policymakers cannot credibly commit to implement this optimal policy. Indeed, given agents' inattention, a

[66]None of our theoretical result hinges on a particular objective function for the government; it could as well be reflecting political economy forces or wider fairness concerns (Saez and Stantcheva, 2016).

discretionary government cannot resist the temptation to increase tax rates beyond their optimal levels, thereby introducing a taxation bias. We formally define the taxation bias as the difference between equilibrium tax rates under discretion and commitment and establish the existence of a positive taxation bias under a mild general requirement.

The demonstration in the paper relies on the assumption of a full-adjustment of the prior in equilibrium (as is required in rational equilibria). We adopt this approach because a full adjustment of the prior in equilibrium is often seen as a strong restriction that should naturally rules out the type of inefficiencies that we highlight in this paper. It should however be clear that it is not a requirement for the taxation bias to hold as is apparent from our main Proposition 9 establishing the existence of a taxation bias. In particular, Appendix 3.9.2 discusses alternative behavioral biases and shows that the taxation bias also arises in setups with persistent misperceptions in equilibrium (e.g. imperfect adjustment of prior beliefs, salience and sparsity).

We then seek to illustrate the implications of this policy distortion and to quantify its magnitude. To do so, we parametrize our tax perception model with Gaussian distributions to provide further theoretical results and numerical simulations. They indicate that even small information frictions induce significant deviations in tax policy. Moreover, they allow to disentangle the implications of inattention to taxes from that of potential behavioral biases reflected in agents' priors. The optimal (or commitment) tax policy is mostly driven by equilibrium perception biases reflected in agents' posteriors. That is, the deviation in the optimal tax policy (Farhi and Gabaix, 2020) from a benchmark without information frictions (Saez, 2001) increases with the bias in the prior and decreases with attention. The actual (or discretionary) tax policy is similarly impacted by equilibrium perception biases but it also depends on taxpayers' attention through a second channel: the policy distortion induced by inattention. The taxation bias, which measures the difference between the two, is thus primarily shaped by attention to taxes and relatively less by potential behavioral biases.

This result transpires in the simple sufficient statistics formula we derive for the taxation bias when income taxes are linear. Indeed, beyond the elasticity of earnings with respect to changes in the perceived marginal net-of-tax rate, it shows that the key sufficient statistic to estimate is the income-weighted average attention in the population. This statistic captures the fact that richer agents are more attentive to tax policy as documented by Taubinsky and Rees-Jones (2017) and as emerges in our endogenous attention model. While income taxes are nonlinear in the US economy, a linear tax model provides a reasonable first-order approximation (Piketty and Saez, 2013). Fitting a linear tax model to US tax data, we find a tax rate of 29.5 percentage points. Further relying on the existing empirical literature to calibrate our sufficient statistics, we estimate that the taxation bias is approximately equal to 3.7 percentage points. This means that the linearized US income tax rate is more than 12% higher than what would be optimal holding the government's objective constant: the taxation bias is large.[67]

We then show that our findings hold important and counterintuitive implications. Situations in which behavioral biases were previously thought to be welfare improving may actually turn out to be welfare decreasing. To illustrate this point, we carry out a welfare analysis in an economy where taxpayers' priors systematically underestimate tax rates (e.g. salience bias). As this downwards bias reduces the efficiency cost of taxation for any given tax rate, we unsurprisingly find that information frictions induce a welfare gain if the government was to implement the optimal tax policy. However, the optimal tax policy cannot be credibly implemented and the actual tax policy features an additional welfare loss due to the taxation bias. As inattention grows, the welfare loss from this policy distortion increases faster than the welfare gain from tax underestimation. Therefore, even if agents systematically underestimate tax rates, information frictions can be detrimental to welfare when agents are not sufficiently attentive to tax policy.

Last, we extend our analysis to nonlinear tax schedules. The government's incentive to increase the marginal tax rate at a given earnings level then depends on agents' attention at (or close to) this earnings level. As a consequence, the positive correlation between income and attention results in an income-specific taxation bias that is globally decreasing with income: the taxation bias is large at low income levels and virtually nonexistent at top income levels. The taxation bias thus attenuates the U-shape pattern of marginal tax

[67]While the existence of a taxation bias does not depend on the objective of the policymaker, the magnitude of the taxation bias does. We use the social welfare weights that can be inferred from actual tax policy given agents' inattention as our baseline and provide a sensitivity analysis.

rates (Saez, 2001) and reduces the progressivity of actual income tax schedules.

This paper contributes to the behavioral public economics literature recently reviewed by Bernheim and Taubinsky (2018). It is the first to analyze the implications of information frictions in agents' tax perceptions for actual tax policy. We present a novel positive theory of tax policy that we link to the existing normative theory to show that inattention generates important policy distortions.[68,69]

Central to this result is the dichotomy between direct and indirect perception adjustments induced by inattention and highlighted in recent empirical evidence. For instance, Sausgruber and Tyran (2005) and Fochmann and Weimann (2013) show that, with time and experience, taxpayers tend to internalize the impact of non-salient taxes they initially ignored.[70] Moreover, if taxpayers act upon their perceptions this dichotomy should also be reflected in earnings choices. Chetty (2012) documents a systematic difference between micro (capturing direct adjustments) and macro (capturing total adjustments) estimates of the elasticity of taxable income and rationalizes this difference by the existence of adjustment rigidities such as information frictions at the micro level.

By showing that information frictions induce a commitment problem leading to inefficient policy outcomes, this paper builds a perhaps unexpected bridge to an earlier literature on the inconsistency of policymaking (Kydland and Prescott, 1977). This justifies our use of the term *taxation bias* in analogy to the *inflation bias* (Barro and Gordon, 1983). A large body of evidence documents the existence of information frictions affecting consumers, firms or even professional forecasters (see e.g. Coibion and Gorodnichenko, 2015). Consequently, our analysis suggests that policy distortions may arise in a wide variety of settings in which the portable framework developed in this paper could be fruitfully applied.

[68] While the term 'positive' sometimes refers to settings in which tax policy is determined as the outcome of a political economy process, it here refers to the discretionary nature of policymaking. See Matějka and Tabellini (2017) for the implications of (rational) inattention in a political economy process.

[69] This finding has potentially important implications for the inverse-optimum approach which aims at inferring the government's objective function from the shape of actual tax schedules (Bourguignon and Spadaro, 2012; Lockwood and Weinzierl, 2016; Jacobs et al., 2017).

[70] Indirect adjustments may also help explain why surveys of taxpayers' perceptions do not provide clear cut evidence of systematic tax rate underestimation (Fochmann et al., 2010; Gideon, 2017); a finding that is *prima facie* hard to reconcile with e.g. ironing behaviors or salience biases.

Policymaking is, at least to some extent, discretionary. In the realm of taxation, discretion is usually discussed in the context of capital levies in which there is indisputable historical evidence of discretionary policies (e.g. Japan post WWII, Italy in 1992, Cyprus in 2013). While less salient for income taxes, discretionary behaviors are likely reflected in the obnoxious complexity of existing tax systems. The French constitutional court has for instance repealed specific items of tax bills for their "excessive complexity" arguing they would not be understood by taxpayers (Conseil Constitutionnel, 2005, 2012). As a result, it should not come as a surprise that individuals strongly oppose tax complexity, even after acknowledging the potential advantages of differential tax treatments (Blesse et al., 2019). Indeed, our findings suggest that by inducing information frictions, tax complexity prompts the government to misbehave from a normative perspective.

The rest of the paper is organized as follows. To build up the intuition, we first characterize in Section 3.2 the taxation bias in a stylized representative agent model with exogenous attention. Section 3.3 microfounds the behavior of heterogeneous and rationally inattentive taxpayers. In Section 3.4, we formalize our positive theory of tax policy and establish the existence of a taxation bias. We then derive a simple sufficient statistics formula for the taxation bias that we take to the data and we illustrate our theoretical results with numerical simulations in Section 3.5. Section 3.6 turns to the welfare implications of information frictions in tax perceptions and Section 3.7 provides an extension to nonlinear tax schedules. The last section concludes. Unless stated in the text all proofs are relegated to the Appendix.

## 3.2 Taxation bias in a stylized model

Consider a canonical labor income taxation model where the government sets a linear tax rate $\tau$ to maximize tax revenue. Let $Y(1-\tau)$ be the aggregate earnings function. The tax revenue function $\tau Y(1-\tau)$ has an inverted U-shape and is nil when $\tau$ is equal to 0 or 100%. As is well-known (e.g. Piketty and Saez (2013)), the revenue maximizing tax rate follows an inverse elasticity rule and is equal to

$$\tau^r = \frac{1}{1+e} \tag{3.1}$$

where $e$ is the elasticity of aggregate earnings with respect to the net-of-tax rate.

Assume now that because of information frictions taxpayers are unable to perfectly observe the tax rate. They must nonetheless form an estimate of the latter to decide how much to work. Call this estimate the perceived tax rate $\tilde{\tau}$ and suppose it is determined by a convex combination of a common prior $\hat{\tau}$ and the actual tax rate $\tau$

$$\tilde{\tau} = \xi\tau + (1-\xi)\hat{\tau}, \tag{3.2}$$

where the weight $\xi \in [0,1]$ can be interpreted as a measure of taxpayers' attention to the actual tax rate $\tau$.[71] Indeed, when $\xi = 1$, taxpayers perfectly observe changes in the tax rate whereas, when $\xi = 0$, they are completely inattentive to tax changes and fully anchor their perception on their prior. Since individual earnings choices depend on their perceived tax rate, aggregate earnings now write $Y(1-\tilde{\tau})$. The tax revenue function becomes $\tau Y(1-\tilde{\tau})$ which remains concave with respect to the actual tax rate $\tau$ given the tax perception model in equation (3.2).

*How do information frictions interact with the design of actual tax policy?* Consider a situation in which taxpayers expect the government to implement the optimal tax rate in the absence of information frictions, that is $\hat{\tau} = \tau^r$. Now, suppose the government sets $\tau = \tau^r$ and consider the effect of a policy deviation that consists in a small increase in the tax rate $d\tau$. This mechanically increases tax revenues by $M = Y(1-\tilde{\tau})\big|_{\tilde{\tau}=\tau^r} d\tau$ while it generates a behavioral response $dY = -\frac{\partial Y(1-\tilde{\tau})}{\partial 1-\tilde{\tau}}\Big|_{\tilde{\tau}=\tau^r} \xi d\tau$ as inattentive taxpayers only observe a fraction $\xi$ of the increase in the tax rate $d\tau$. By definition of $\tau^r$, the mechanical effect $M$ outweighs the fiscal externality $FE = \tau^r dY$ induced by the behavioral response when agents are not fully attentive ($\xi < 1$). As a result, the government systematically deviates from tax policy $\tau^r$ and ends up choosing a higher tax rate.

[71] Gabaix (2019) argues this is a unifying framework to modeling various behavioral biases and attention theories.

Conceptually, an important consequence of inattention is to anchor taxpayers' perceptions on their prior. Because of this anchoring, the government has an incentive to implement policy deviations that taxpayers are going to partially ignore. This is a form of discretionary policy which arises as a side-effect of information frictions in tax perceptions. The government thus chooses its tax policy taking agents' priors and attention into account. Specifically, tax policy $\tau(\xi, \hat{\tau})$ is decreasing in taxpayers' prior $\hat{\tau}$ and attention parameter $\xi$

$$\tau(\xi, \hat{\tau}) \quad = \quad \begin{cases} \frac{1-(1-\xi)\hat{\tau}}{\xi(1+e)} & \text{if } \hat{\tau} \geq 1 - \frac{\xi}{1-\xi} e \\ 1 & \text{otherwise} \end{cases} \tag{3.3}$$

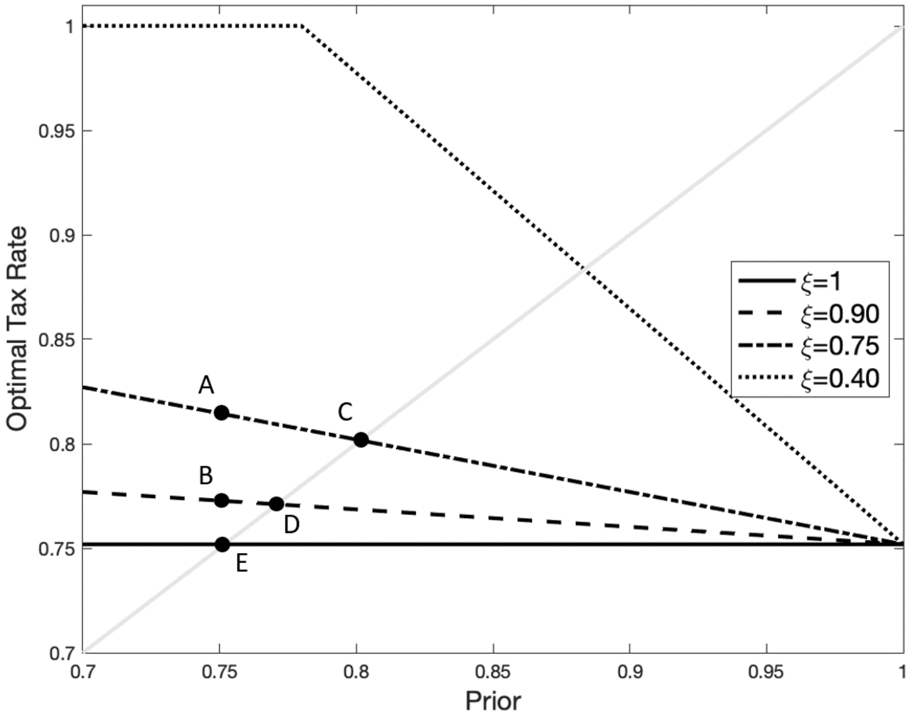
where the elasticity of aggregate earnings is defined with respect to the perceived net-of-tax rate $e \equiv \frac{1-\tilde{\tau}}{Y} \frac{\partial Y}{\partial(1-\tilde{\tau})} \geq 0$. The solution is interior whenever attention $\xi$ or the prior $\hat{\tau}$ are high enough – otherwise the government finds it optimal to impose a 100% tax – and coincides with the inverse elasticity rule when agents are perfectly attentive ($\xi = 1$) and thus fully informed.

Figure 3.1 plots tax policy $\tau(\xi, \hat{\tau})$ as a function of agents' prior for different attention levels. It shows that small information frictions generate notable deviations in the government behavior. If agents' prior is that the government implements the inverse elasticity rule – $\tau^r = 75\%$ assuming $e = 0.33$ – the government chooses a tax rate of 82% (resp. 77%) when the attention parameter $\xi$ is equal to 0.75 (resp. 0.90).[72] This corresponds to point A (resp. B).

An equilibrium is as a situation in which (i) neither taxpayers nor the government has an incentive to deviate and (ii) taxpayers' actions and perceptions are mutually consistent with the government's choice of tax policy. We here focus on rational equilibria in which agents correctly anticipate the equilibrium tax policy $\hat{\tau} = \tau^{\text{eq}}$. We defer the introduction of biased equilibria to Section 3.4 and Appendix 3.9.2 illustrates that our conclusions are also expected to hold in models with persistent misperceptions in equilibrium (e.g.

[72] As an element of comparison, Gabaix (2019) states (p. 5) that "on average, the attention parameter estimated in the literature is 0.44, roughly halfway between no attention and full attention" while adding that "attention is higher when the incentives to pay attention are stronger" which should likely be the case when it comes to taxing 75% of one's income.

Figure 3.1: Optimal tax policy and equilibrium outcomes




NOTE: Optimal policy as a function of the prior $\hat{\tau}$ for different values of the attention parameter $\xi$. The elasticity of aggregate earnings with respect to the perceived net-of-tax rate is set to 0.33.

partial prior adjustment, salience and sparsity). Plugging this equilibrium condition in the government's choice of tax policy (3.3) we obtain that the equilibrium tax policy is

$$\tau^{\text{eq}} = \frac{1}{1 + \xi e} \tag{3.4}$$

Graphically, the rational equilibrium is represented by the point where the 45-degree line ($\hat{\tau} = \tau^{\text{eq}}$) intersects the government policy function. Again, small information rigidities lead to large deviations in equilibrium. In a rational equilibrium, the tax rate is 80% (resp. 77%) when attention is such that 75% (resp. 90%) of a tax reform is directly internalized by taxpayers. This corresponds to point C (resp. D).

The government is unable to reach the top of the Laffer curve: the equilibrium tax rate is inefficiently high. We refer to this phenomenon as the *taxation bias* in analogy to the *inflation bias* (Barro and Gordon, 1983). While the government internalizes the direct impact of its choice on agents' perceptions (proportional to attention $\xi$), it does not internalize the equilibrium impact associated with the adjustment of the prior (proportional to inattention $1-\xi$). In Barro and Gordon's (1983) words, "the equality of policy expectations and realizations is a characteristic of equilibrium – not a prior constraint" (p. 591). Hence, inattention resurges the possibility that discretionary policies lead to inefficient outcomes (Kydland and Prescott, 1977).

To formalize this result, we characterize the optimal policy under commitment $\tau^\star$. This is the optimal policy of a government who can credibly commit to implement a tax level and thereby has to take into account the equilibrium effect of its choice of tax policy on perceptions. By definition, this is the optimal policy in the presence of information frictions and it coincides in this stylized rational equilibrium framework with the inverse elasticity rule $\tau^\star = \tau^r$ (point E). However and as shown above, it cannot be an equilibrium policy under discretion in the presence of inattention.

Defining the taxation bias as the difference between the tax rates under discretion and commitment we have

$$\tau^{\text{eq}} - \tau^\star \quad = \quad \frac{(1-\xi)e}{(1+\xi e)(1+e)} \geq 0 \tag{3.5}$$

Therefore, the taxation bias is strictly positive when agents are not fully attentive to taxes ($\xi < 1$). Moreover, the (absolute) size of the taxation bias increases with agents inattention $1-\xi$ and with the elasticity $e$ as they intuitively both make policy deviations relatively more attractive.

To highlight the mechanisms that lead to a taxation bias, we have analyzed in this section a stylized representative agent model in which agents' behavior and attention are exogenously given and in which the government implements a linear tax policy to maximize tax revenue. In the remainder of the paper we broaden the scope of the analysis by studying the problem of a welfare-maximizing government facing a heterogeneous population of agents whose individual behavior is fully micro-founded and whose attention is endogenous. We also extend the equilibrium concept to allow for perception biases in equilibrium – while maintaining the assumption of full adjustment of the prior in equilibrium, see Appendix 3.9.2 for other equilibrium assumptions – and examine how the taxation bias affects (the progressivity of) non-linear tax schedules. These extensions provide valuable insights on the magnitude and implications of the taxation bias in a policy relevant environment.

## 3.3 Agents' behavior, perceptions and attention

This section describes the behavior of taxpayers in the economy. Because of information frictions, taxpayers may not freely observe the tax rate implemented by the government. They rely on a Bayesian learning model with costly information acquisition to form their perceptions about the tax schedule in order to decide how much to earn and consume.

### 3.3.1 Primitives and assumptions

We consider a population of agents with heterogeneous productivities $w$ which are private information and distributed from a well-defined probability distribution function $f_w(w)$. We assume taxpayers have a utility function $U(c, y; w)$ where $c$ is consumption and $y$ earnings and where we impose $U(.)$ to be continuously differentiable, increasing in consumption ($U_c > 0$), decreasing in effort ($U_y < 0$ and $U_w > 0$) and such that the Spence-Mirrlees condition holds ($MRS_{yc}$ decreases with skill $w$). For simplicity, we consider a separable and quasi-linear utility $U(c, y; w) = c - v(y; w)$ in the body of the paper and show in the Online Appendix how we can extend the analysis to more general utility functions.[73]

Agents choose their consumption $c$ and earnings $y$ subject to an income tax $T(y)$. Because of information frictions, we assume that taxpayers are unable to freely observe $T(y)$ and instead rely on individual-specific perceived income tax schedules denoted $\tilde{T}(y)$.

**Assumption 2** (linear representation)**.** *Individuals use a linear representation of the tax schedule* $\tilde{T}(y) = \tilde{\tau} y - \tilde{R}$

To make their consumption and earnings choices, individuals rely on their perceptions of the tax liability at each earnings level. Assumption 2 imposes that taxpayers use a

[73] Separability between earnings and consumption preferences combined with quasi-linearity guarantees the absence of income effects in labor supply decisions and considerably simplifies the analysis.

linear representation of the tax schedule. Hence, agents only need to form estimates of the marginal tax rate $\tilde{\tau}$ and the intercept $\tilde{R}$ thereby reducing the dimensionality of the perceptions formation problem to two parameters.[74]

In most of the paper, we consider that the actual tax schedule is also linear and denote by $(\tau_0, R_0)$ its slope and intercept. Consequently, we define $(\tau, R)$ as the associated random variables from the point of view of the agents. In Section 3.7 we extend the analysis to non-linear tax schedules.

### 3.3.2 Individual problem

Individuals jointly choose an allocation $(c, y)$ and how much information to collect about the tax schedule. This one-step problem is equivalent to a two-step problem that we characterize. The first step identifies the optimal allocation choice given a perceived tax schedule $\tilde{T}(y)$ while the second step determines the optimal information acquisition taking into account how perceptions affect allocations.

**Allocation choice** Agents choose consumption $c$ and earnings $y$ to maximize their utility subject to their perceived budget constraint which depends on their perceptions of the tax schedule. This problem writes

$$\begin{aligned} \max_{c,y} \quad & \int_{\tau} U(c, y; w)\ \tilde{q}(\tau) d\tau \\ \text{s.t.} \quad & \quad c \leq R + (1-\tau)y \end{aligned} \tag{3.6}$$

where $\tilde{q}(\tau)$ is the perceived probability distribution of the marginal tax rate $\tau$. With a separable and quasi-linear utility function, the first-order condition determining earnings writes

$$\frac{\partial v(y; w)}{\partial y} = 1 - \tilde{\tau} \tag{3.7}$$

[74] Beyond the fact that a linear approximation is usually a good approximation of existing tax schedules (Piketty and Saez, 2013), recent empirical evidence suggests that in practice taxpayers tend to use linear representations of tax schedules (Rees-Jones and Taubinsky, 2020).

with $\tilde{\tau} \equiv E_{\tilde{q}(\tau)}[\tau]$ the average perceived marginal tax rate. Consequently, the average perceived marginal tax rate $\tilde{\tau}$ is a sufficient statistics for labor supply and uniquely pins down optimal earnings $y^\star(\tilde{\tau}; w)$. Hence, a direct implication of quasi-linear separable preferences is that tax liability, and in particular the perceived value of the demogrant $\tilde{R}$, is irrelevant for labor supply and only matters to determine agents' consumption levels.

**Assumption 3** (slack budget)**.** *Consumption adjusts such that agents exhausts their true budget i.e.* $c^\star(\tilde{\tau}; w) = R_0 + (1 - \tau_0) y^\star(\tilde{\tau}; w)$

We assume consumption adjusts to ensure that the true budget constraint holds ex post.[75] The only parameter of interest for agents' allocation choice is thus the perceived marginal tax rate $\tilde{\tau}$.

Given this allocation choice, an agent's indirect utility is

$$V(\tilde{\tau}, \tau_0, R_0; w) = R_0 + (1 - \tau_0) y^\star(\tilde{\tau}; w) - v\Big(y^\star(\tilde{\tau}; w); w\Big) \tag{3.8}$$

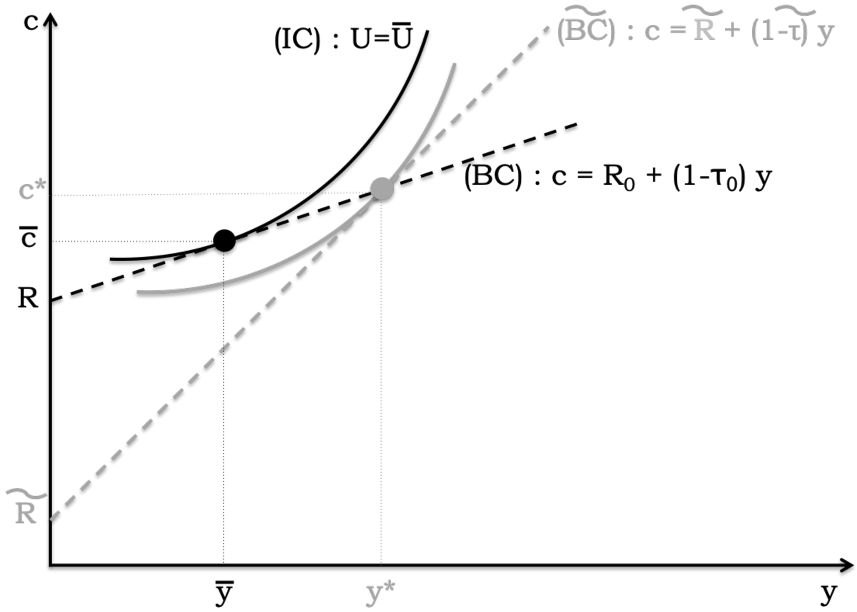
Figure 3.2 summarizes the allocation choice in a y-c diagram. Perceptions of the tax schedule determine earnings (tangency condition with perceived budget line) while consumption adjusts to the true budget constraint (intersection with true budget line).

A natural observation from Figure 3.2 is that misperceptions induce utility misoptimization costs: the utility level $\bar{U}$ associated with the choice under accurate perceptions (black dot) is higher than the utility level $U^\star$ associated with the choice under misperceptions (grey dot). However, when agents underestimate tax rates ($\tilde{\tau} < \tau_0$) misperceptions also induce efficiency gains: earnings $y^\star$ chosen under misperceptions are larger than earnings $\bar{y}$ at the optimal allocation. As a result, tax underestimation may increase social welfare if efficiency gains dominate utility misoptimization costs.

**Perceptions formation** Tax perceptions here follow from a Bayesian learning model with a choice of information (in Gabaix's (2019) terminology). They result from the combination of an exogenous and free prior (also referred to as a belief or an anchor) and

[75]This assumption is used throughout the behavioral tax literature as emphasized in Reck (2016) who discusses different budget adjustment rules in misperception models. See also Farhi and Gabaix (2020) for a related discussion.

Figure 3.2: Allocation choice in a y-c diagram




NOTE: The figure displays an exemple of allocation choice when the agent underestimates the marginal tax rate $\tilde{\tau} < \tau_0$. The grey (resp. black) dot represents the allocation choice of an individual who misperceives (resp. correctly perceives) the tax schedule. The plain lines are the indifference curves and the dashed lines the budget contraints.

an endogenous and costly information acquisition process. We choose this model for its wide use in economics, its well-understood micro-foundations and the fact that – as we show – it generates predictions that are consistent with the empirical evidence.

Let $\hat{q}(\tau)$ be the prior probability distribution about the tax rate and $\hat{\tau} \equiv E_{\hat{q}}[\tau]$ the expected tax rate derived from the prior. This probability distribution accounts for sources of structural and subjective uncertainty which may be related to policy primitives (e.g. hidden tastes for redistribution), economic fundamentals (e.g. shocks to the government expenditure requirements), institutions (e.g. inability to implementation a chosen policy), heuristic decision rules (e.g. ironing), etc.

In the following, we voluntarily remain agnostic about the origin of the prior and the sources of uncertainty it may capture for two reasons. First, the assumed ex ante uncertainty essentially represents a motive for taxpayers to learn in our setup and the main results of the paper will hold for a wide variety of well-defined smooth priors. Second, while the empirical literature clearly indicates that taxpayers tend to misperceive tax rates, there is yet no consensus on the exact rationale – or rationales – behind these misperceptions. Hence, we consider diverse situations ranging from priors that are correct on average to priors that are systematically biased due to cognitive or perception biases.

Information about the actual tax rate $\tau_0$ takes the form of an unbiased Gaussian signal with precision $1/\sigma^2$. For a realization $s$ of the signal, the posterior belief follows from Bayes law

$$\tilde{q}(\tau|s;\sigma) \quad \propto \quad \phi(s;\tau,\sigma^2)\hat{q}(\tau) \tag{3.9}$$

where $\phi(s;\tau,\sigma^2)$ is the Gaussian pdf with mean $\tau$ and variance $\sigma^2$. Building on the rational inattention literature (Sims, 2003), the information content transmitted through the signal is measured from the entropy reduction between the prior and the posterior

$$\mathcal{I}(\sigma) \equiv H(\hat{q}(\tau)) - E_{p(s)}\Big[H(\tilde{q}(\tau|s;\sigma))\Big] \tag{3.10}$$

where $H(q(\tau)) \equiv -\int q(\tau)\log_2(q(\tau))d\tau$ is the differential entropy (in bits) of the probability distribution $q(\tau)$ and $E_{p(s)}[.]$ the expectation taken over the marginal distribution of signals $p(s) \equiv \int \phi(s;\tau,\sigma)\hat{q}(\tau)d\tau$. Intuitively, $\mathcal{I}(\sigma)$ is a measure of the expected amount of information transmitted through the signal. To account for the energy and time devoted to acquiring and processing information, taxpayers suffer a utility cost $\kappa$ per unit (bit) of processed information.[76] The attention strategy of a taxpayer with productivity $w$ thus results from an arbitrage between improved private decisions thanks to more accurate information and the cost to acquire this information. More specifically, she chooses the signal's precision – or equivalently its standard error $\sigma^\star(\hat{q}(\tau),\kappa,w)$ – to maximize her expected indirect utility

$$\max_{\sigma} \qquad \iint V\Big(\tilde{\tau}(s,\sigma),\tau,R;w\Big)\ \phi(s;\tau,\sigma)\ \hat{q}(\tau)\ dsd\tau - \kappa\mathcal{I}(\sigma) \tag{3.11}$$

where $\tilde{\tau}(s,\sigma) \equiv E_{\tilde{q}(\tau|s;\sigma)}[\tau]$ is the expected perceived marginal tax rate once the signal is observed and henceforth referred to as the perceived tax rate. Note that the decision to acquire information is here only based on the information contained in the prior

[76] Our results naturally extend to more general information cost functions.

distribution $\hat{q}(\tau)$ which ensures the internal consistency of this learning model.[77]

In the following, we denote by $f_{\tilde{\tau}}(\tau|\tau_0, w)$ the posterior distribution of $\tilde{\tau}(s, \sigma^\star)$ for a taxpayer with productivity $w$ and signal $s$ drawn from the Gaussian distribution with mean $\tau_0$ and variance $\sigma^\star$. This function summarizes the distribution of agent $w$ perceptions in the economy. Moreover, for a given perceived tax rate $\tilde{\tau}$, agent $w$ indirect utility net of information costs writes

$$\mathcal{V}(\tilde{\tau}, \tau_0, R_0; w, \kappa) = V(\tilde{\tau}, \tau_0, R_0; w) - \kappa\mathcal{I}(\sigma^\star) \tag{3.12}$$

### 3.3.3 Tractable Gaussian model

The general Bayesian learning model presented above is generally intractable. We here focus on the Gaussian formulation in order to derive some predictions and implications of the model. As highlighted in the inattention literature (Maćkowiak and Wiederholt, 2015; Mackowiak et al., 2018), a closed form solution to problem (3.11) can be obtained under the following assumption.

**Assumption 4** (tractable Gaussian learning)**.** *Let the prior $\hat{q}(\tau)$ be the Gaussian distribution with mean $\hat{\tau}$ and variance $\hat{\sigma}^2$ and assume that agents use a quadratic approximation of their indirect utility to choose their attention strategies.*

Under the assumption that the prior is Gaussian, the posterior will also be Gaussian and the information measure $\mathcal{I}(\sigma^\star)$ takes a simple form.[78] Relying on a second-order approximation of indirect utility, the solution to this problem can then be derived.

**Lemma 6.** *In a tractable Gaussian learning model, the expected perceived marginal tax rate $\tilde{\tau}$ is given by*

$$\tilde{\tau}(s, \sigma^\star) = \xi(\sigma^\star)s + (1 - \xi(\sigma^\star))\hat{\tau} \tag{3.13}$$

[77] In other words, agents "don't know what they don't know".

[78] One may instead consider that the prior is a truncated Gaussian with support on $[0, 1]$ in order to ensure that perceived tax rates $\tilde{\tau}$ always remain between zero and one. Doing so, the problem remains tractable but formulas become a lot less transparent. In practice, our simulations suggest that when the prior is sufficiently informative (low variance) and the tax rate not too extreme, the posterior support belongs to $[0, 1]$ a.s. with a Gaussian prior.

*where* $\xi(\sigma^\star) \equiv \frac{\hat{\sigma}^2}{\hat{\sigma}^2+\sigma^{\star 2}} \in [0,1]$ *is a measure of attention strategies.*

*Proof.* See Appendix 3.9.3 □

The perceived tax rate $\tilde{\tau}$ is given by a convex combination of the prior $\hat{\tau}$ and the realization of the signal $s$ where the weight $\xi$ is a measure of attention. Indeed, the lower the attention parameter $\xi$ is, the more taxpayers rely on their prior $\hat{\tau}$ and the less attention they devote to acquiring information about the actual tax rate through the signal $s$. In other words, agents tend to choose to ignore their signal if they do not invest in information acquisition and the signal is hence relatively uninformative in comparison to the prior.

**Lemma 7.** *In a tractable Gaussian learning model, the optimal attention strategy* $\xi$ *is given by*

$$\xi = \max\left(0, 1 + \frac{\kappa}{\hat{\sigma}^2 \int \frac{\partial^2 y^\star}{\partial \tilde{\tau}^2}\Big|_{\tilde{\tau}=\tau} \phi(\tau;\hat{\tau},\hat{\sigma}) d\tau}\right). \tag{3.14}$$

*Proof.* See Appendix 3.9.3 □

Each taxpayer's attention strategy $\xi$ is characterized by equation (3.14). Attention decreases with the information cost $\kappa$ and increases with the uncertainty in the prior $\hat{\sigma}^2$. It also depends on the responsiveness of agents labor supply decisions to changes in perceived tax rates through $\frac{\partial^2 y^\star}{\partial \tilde{\tau}^2}$. Indeed, an agent's responsiveness to changes in perceptions determines the value of information acquisition. As a result, attention increases with earnings ability $w$ and with expected prior tax rates. Intuitively, agents who are more productive have a greater latitude in their earnings choice and will thus be more attentive – and hence responsive – to taxes. In a similar fashion, responsiveness to tax changes and thus attention increases when expected tax rates increase because it shifts the labor supply function to regions with a larger curvature.

**Predictions and implications** Agents choose their attention strategies through the maximization of their expected indirect utility which is based on their prior. Attention is hence unaffected by an unanticipated change in the realized tax rate $d\tau_0$. As a result, an unanticipated change in the tax rate $d\tau_0$ induces a change in the distribution of the posterior $f_{\tilde{\tau}}(\tilde{\tau}|w)$ only through a change in the signal $s$ which is now drawn from a novel distribution $\phi(s; \tau_0 + d\tau_0, \sigma^\star)$. Therefore, taxpayers' perceived tax rate slowly adjusts to news and perceptions are anchored on the prior. Anchoring is a widely documented bias in the behavioral literature (Gabaix, 2019) which has two major implications in this context.

First, if agents' prior is biased ($\hat{\tau} \neq \tau_0$) and agents are not fully attentive ($\xi < 1$), the posterior and hence agents perceptions of the tax schedule will also be biased (almost surely). Indeed, taking expectations over signal realizations we have that

$$\tau_0 - E_{\phi(s)}[\tilde{\tau}(s, \sigma^\star)] = \tau_0 - [\xi\tau_0 + (1-\xi)\hat{\tau}] = (1-\xi)(\tau_0 - \hat{\tau}). \tag{3.15}$$

Second, taxpayers labor supply responses to unanticipated changes in the tax rate are attenuated by anchoring. Indeed, taking the prior as given, responses to tax changes only transit through the variation of the signal which is weighted by attention $\xi$. Formally, this means $\frac{d\tilde{\tau}}{d\tau_0} = \xi$ such that

$$\frac{dy^\star}{d\tau_0} = \xi \frac{dy^\star}{d\tilde{\tau}} \tag{3.16}$$

Intuitively, $\frac{dy^\star}{d\tilde{\tau}}$ captures agents' preferences while $\xi$ is a dampening factor that captures the fraction of the tax change that agents perceive. As a result the elasticity of labor supply with respect to unanticipated changes in the tax rate decreases in the presence of inattention.

The predictions derived from a Bayesian learning model with a choice of information thus seem consistent with the bulk of the empirical evidence on tax perceptions and behavioral responses to taxes. Most importantly, it can account for the presence of systematic perception biases and implies that elasticities will be lower when inattention is at play. In addition the model also generates dispersion in perceptions – through the noisiness of the signal – and features an increase in overall attention upon tax increases which hold potentially important welfare implications (Taubinsky and Rees-Jones, 2017).

## 3.4 Discretion, commitment and the taxation bias

This section introduces the problem of the government and formalizes our positive theory of tax policy. It characterizes tax policy under discretion and commitment and provides a formal definition of the equilibrium (with full adjustment of the prior). A general result on the existence of a taxation bias concludes.

### 3.4.1 Government problem and discretion

We consider a welfarist government that maximizes a general social welfare function summing an increasing and weakly concave transformation $G(.)$ of taxpayers' indirect utilities net of information costs. It chooses a target tax schedule $(\tau_g, R_g)$, where $\tau_g$ is the marginal tax rate and $R_g$ the demogrant, taking the distribution of skills $f_w(w)$ in the population as given.

Following Matějka and Tabellini (2017), we introduce implementation shocks $\vartheta$ as an underlying source of uncertainty in the model. The target tax rate is implemented up to a realization of this implementation shock such that the actual tax rate is $\tau_0 = \tau_g + \vartheta$ where $\vartheta$ is a white noise drawn from an exogenous distribution $f_\Theta(\vartheta)$ known to both taxpayers and the government. We assume the actual demogrant $R_0$ adjusts to the realization of the implementation shock $\vartheta$ to ensure that the government budget constraint is always binding. Conceptually, these implementation shocks are introduced to ensure that Bayesian taxpayers have an incentive to learn in equilibrium. They allow to formally close the model but have an otherwise negligible impact on the optimal tax policy. Hence, we sometimes use small shocks approximations in which case we explicitly disregard the small effects they may induce.

The government problem writes

$$\max_{\tau_g, R_g} \quad E_\vartheta\left[\iint G\Big(\mathcal{V}(\tilde{\tau}, \tau_0, R; \kappa, w)\Big)\ f_{\tilde{\tau}}(\tau|\tau_0; w)\ f_w(w) d\tau dw\right] \tag{3.17}$$

$$\text{s.t.} \quad \iint \tau_0 y^\star(\tilde{\tau}; w)\ f_{\tilde{\tau}}(\tau|\tau_0; w)\ f_w(w) d\tau dw \geq R_0 + E \tag{3.18}$$

where $E$ is an exogenous expenditure requirement, the expectation is taken over the implementation shock $\vartheta$ and $f_{\tilde{\tau}}(\tau|\tau_0; w)$ is the posterior distribution of perceived rates for a taxpayer with productivity $w$ given the actual tax rate $\tau_0 = \tau_g + \vartheta$.

**Discretionary policy** The government's optimal tax policy solves problem (3.17). When doing so, it takes the prior distribution $\hat{q}(\tau)$ as given. This is a form of Nash conjecture used to compute the best response of the government. While the problem is fundamentally simultaneous, it can be equivalently described by the following sequence of events which we here layout for the sake of clarity:

0. Agents are endowed with a common prior $\hat{q}(\tau)$ and the distribution of skills is $f_w(w)$.

1. The government sets the target tax policy $(\tau_g, R_g)$ to maximize (3.17).

2. The actual tax rate $\tau_0 = \tau_g + \vartheta$ is implemented up to an implementation shock drawn from a known distribution $f_\Theta(\vartheta)$ and the actual demogrant $R_0$ adjusts to the resource constraint.

3. Taxpayers choose their attention strategies using their common prior $\hat{q}(\tau)$, observe a Gaussian signal $s$ about $\tau_0$ which precision depends on their attention and decide how much to consume and earn.

4. The government levies taxes and redistributes through the demogrant.

The government understands that taxpayers will gather information and adjust their decisions in reaction to its choice of tax policy, it therefore plays "first" in the above-described sequence of events. However, it (i) treats the prior distribution $\hat{q}(\tau)$ and the skill distribution $f_w(w)$ as predetermined state variables and (ii) cannot directly influence agents' attention strategies since they are based on agents' predetermined prior. As a result, the government does not have a particular strategic advantage from playing "first" – thus reflecting the simultaneous nature of the problem. Importantly, the government is as rational and informed as in the standard Mirrlees (1971) model and the novelty relates to information frictions on the agents' side.

The tax policy of the government follows from Proposition 7, where first-order conditions have to hold in expectation of the realization of the implementation shock.

**Proposition 7.** *The discretionary tax policy* $(\tau_g, R_g)$ *is characterized by*

$$(\tau_g): \quad E_\vartheta\Big[\int \Big\{ \underbrace{\int \Big[-\frac{G'(\mathcal{V})}{p} y^\star + y^\star\Big] f_{\tilde{\tau}}(\tau|\tau_0; w) d\tau}_{\text{mechanical \& welfare effects}}$$

$$+ \underbrace{\int \Big[\frac{G(\mathcal{V})}{p} + \tau_0 y^\star\Big] \frac{d f_{\tilde{\tau}}(\tau|\tau_0; w)}{d\tau_g}\Big|_{\hat{q}(.)} d\tau}_{\text{direct behavioral responses}} \Big\} f_w(w) dw\Big] = 0 \quad (3.19)$$

$$(R_g): \quad E_\vartheta\Big[\iint \Big[\frac{G'(\mathcal{V})}{p} - 1\Big] f_{\tilde{\tau}}(\tau|\tau_0; w) f_w(w) d\tau dw\Big] = 0 \quad (3.20)$$

*together with the resource constraint (3.18) and where p represents the social marginal cost of public funds.*

*Proof.* See appendix 3.10 □

The first order condition (3.19) captures the (expected) effects of a marginal increase in the target tax rate $d\tau_g$. The first line measures the impact of the reform on allocations when the distribution of perceptions remains fixed. It corresponds to the standard mechanical and welfare effects: a marginal increase in the tax rate mechanically increases tax revenue by $y^\star d\tau_g$ additional dollars but reduces taxpayers' consumption and thus welfare by $\frac{G'(\mathcal{V})}{p} y^\star \, d\tau_g$ dollars (Piketty and Saez, 2013).

The second line in condition (3.19) relates to the impact of the reform on the distribution of perceptions and thus captures behavioral responses to the reform. Indeed, behavioral responses transit through variations in the posterior distribution $f_{\tilde{\tau}}(\tau|\tau_0, w)$ of perceived tax rates $\tilde{\tau}$ which reflect changes in the actual tax rate $\tau_0$. A marginal increase in the tax rate increases, on average, the perceived tax rate by $d\tilde{\tau}$ and thus reduces tax revenue by $\tau_0 y^\star(\tilde{\tau}) d\tilde{\tau}$. This is a reformulation of the standard behavioral effect. Moreover, because agents misoptimize, the envelope theorem no longer applies and a marginal deviation from taxpayers' perceived rate induces a welfare cost equal to $\frac{G(\mathcal{V}(\tilde{\tau}))}{p} d\tilde{\tau}$. This new welfare effect introduces a corrective motive for taxation in the presence of misperceptions common to optimal tax models with behavioral agents (Gerritsen, 2016; Farhi and Gabaix, 2020).

Condition (3.20) states that in the absence of income effects, social marginal welfare weights $g \equiv \frac{G'(\mathcal{V})}{p}$ average to 1 at the optimum: the government is indifferent between having an additional dollar or redistributing an additional dollar (Saez, 2001).

### 3.4.2 Equilibrium definition

An equilibrium is a set of target tax policy, denoted $(\tau_g^{\text{eq}}, R_g^{\text{eq}})$, and a set of attention, consumption and earnings decisions such that neither the government nor taxpayers have an incentive to deviate. Moreover, in equilibrium agents' prior $\hat{q}(\tau)$ must be mutually consistent with the government's target tax rate and with the uncertainty induced by the implementation shock.

As discussed in the introduction, there is a large body of evidence suggesting the existence of systematic perception biases. Therefore, we allow for a potential perception bias $b$ in agents' common prior but remain agnostic on the origin of this potential bias. We henceforth call rational (resp. biased) an equilibrium in which agents correctly (resp. incorrectly) anticipate the target tax policy such that $b = 0$ (resp. $b \neq 0$).

Given the structure of the problem, the only free variables are the government's target tax rate $\tau_g$, agents' attention strategies and the equilibrium distribution of the common prior $\hat{q}(\tau)$. Hence, for the sake of simplicity our formal definition of the discretionary equilibrium only involves these variables. Once they are set, all remaining variables may be mechanically deduced.

**Definition 1** (equilibrium)**.** *Given the distribution of the implementation shock $f_\Theta(\vartheta)$, the equilibrium is a set of target tax rate $\tau_g^{\text{eq}}$ chosen by the government and attention strategies chosen by the agents such that*

*(a) The target tax rate $\tau_g^{\text{eq}} \in [0, 1]$ solves the government's problem (3.17) given the common prior distribution $\hat{q}(\tau)$.*

*(b) Attention strategies solve agents' problem (3.11) given the prior distribution $\hat{q}(\tau)$.*

*(c) The common prior distribution $\hat{q}(\tau)$ is the pdf of $\tau_g^{\text{eq}} + b + \vartheta$.*

Condition (a) and (b) guarantee that the government and the agents will not have an incentive to deviate while condition (c) ensures that agents' prior and actual tax policy are mutually consistent up to an arbitrary bias $b$. Indeed, condition (c) implies that the average prior is $\hat{\tau} = \tau_g^{\mathrm{eq}} + b$ in equilibrium. Consequently, taxpayers correctly anticipate the government policy in the rational equilibrium ($b = 0$) and their attention strategies then reflect their willingness to observe the implementation shock $\vartheta$ – which is indeed the only information conveyed through the signals. Hence, implementation shocks are here essentially introduced to ensure that Bayesian taxpayers have an incentive to learn in equilibrium but do not otherwise play an economically meaningful role.

### 3.4.3 Commitment and the taxation bias

The discretionary equilibrium is socially suboptimal. To formalize this point, we characterize the welfare-maximizing feasible tax policy. It corresponds to the optimal policy that would be chosen by the government if it could credibly commit to a tax policy. We thus refer to it as the commitment tax policy.

**Commitment policy** The commitment tax policy is the policy that would be chosen by a benevolent social planner who has the same information as the government but internalizes all equilibrium effects of tax policy. By implicitly restricting the set of tax policies to precommited policy rules, the normative literature (e.g. Farhi and Gabaix (2020)) characterizes this commitment tax policy which corresponds to the optimal tax policy in the presence of information frictions.

Formally, the commitment tax policy solves the government's problem (3.17) subject to the additional feasibility condition that agents' prior and actual tax policy realizations have to be mutually consistent in equilibrium (condition (c) in Definition 1). It is characterized by the following first order conditions.

**Proposition 8.** *The commitment tax policy* $(\tau_g^\star, R_g^\star)$ *is characterized by*

$$(\tau_g^\star) \quad : \quad E_\vartheta\Big[\int\Big\{\underbrace{\int\Big[-\frac{G'(\mathcal{V})}{p}y^\star + y^\star\Big]f_{\tilde{\tau}}(\tau|\tau_0;w)d\tau}_{\text{mechanical \& welfare effects}}$$

$$+ \underbrace{\int\Big[\frac{G(\mathcal{V})}{p} + \tau_0 y^\star\Big]\frac{df_{\tilde{\tau}}(\tau|\tau_0;w)}{d\tau_g}d\tau}_{\text{direct \& equilibrium behavioral responses}}\Big\}f_w(w)dw\Big] = 0 \tag{3.21}$$

$$(R_g^\star) \quad : \quad E_\vartheta\Big[\iint\Big[\frac{G'(\mathcal{V})}{p} - 1\Big]f_{\tilde{\tau}}(\tau|\tau_0;w)f_w(w)d\tau dw\Big] = 0 \tag{3.22}$$

*together with the resource constraint (3.18) and where p represents the social marginal cost of public funds. This is the policy implemented in a commitment equilibrium.*[79]

*Proof.* See Appendix 3.10 □

As before, conditions have to hold in expectation because of the implementation shock $\vartheta$. The main difference between Propositions 7 and 8 is that the derivative $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_0;w)}{d\tau_g}$ in equation (3.21) now reflects changes in the signal received (direct adjustment) as well as changes in the prior (equilibrium adjustment). Hence, equilibrium adjustments are here internalized in the choice of tax policy.

**Taxation bias** The discrepancy between the discretionary and commitment equilibria represents a taxation bias. It is a measure of the deviation from the welfare-maximizing feasible tax policy $\tau_g^\star$.

**Definition 2** (taxation bias)**.** *The taxation bias is the difference between the equilibrium tax rates under discretion* $\tau_g^{\text{eq}}$ *and commitment* $\tau_g^\star$*.*

The taxation bias arises as a consequence of the government's inability to internalize equilibrium adjustments in its choice of tax policy which induces a commitment problem. Proposition 9 relates the existence of a positive taxation bias to the associated aggregate equilibrium behavioral responses.

[79]The commitment tax policy is not an equilibrium policy in the sense of Definition 1 because tax policy solves a different problem under commitment. Hence, notions of equilibrium under commitment implicitly refer to the equilibrium of a game in which tax policy would solve the commitment problem. The equilibrium tax policy is then simply equal to the commitment tax policy since all equilibrium adjustments are internalized in the choice of tax policy through the feasibility constraint.

**Proposition 9.** *When both equilibria exist and are unique, there is a positive taxation bias if and only if*

$$E_{\vartheta}\left[\iint \left(\frac{G(\mathcal{V})}{p} + (\tau_g^\star + \vartheta)y^\star\right)\left(\frac{df_{\tilde{\tau}}(\tau|\tau_g^\star + \vartheta; w)}{d\tau_g} - \frac{df_{\tilde{\tau}}(\tau|\tau_g^\star + \vartheta; w)}{d\tau_g}\Big|_{\hat{q}(.)}\right) f(w) d\tau dw\right] \leq 0 \quad (3.23)$$

*Proof.* $\tau_g^\star$ solves equation (3.21). Then, condition (3.23) implies that the left hand-side of equation (3.19) is $\geq 0$ when evaluated at $\tau_g^\star$. Hence, it directly follows from the existence and uniqueness of the discretionary equilibrium that $\tau_g^{\text{eq}} \geq \tau_g^\star$. □

Equation (3.23) represents the expected change in welfare due to a marginal increase in the prior average. The term $G(\mathcal{V})/p$ stands for the welfare impact of the failure of the envelope condition. It is therefore of second order and can be overlooked when perception biases $b$ are small. Therefore, the above condition holds whenever the expected change in aggregate tax revenue following a marginal increase in the prior average is negative (and of first order). In other words, when perception biases are small there is a positive taxation bias as long as agents tend to work less when they anticipate higher taxes – a mild condition. This shows that information frictions lead to upward distortions in actual tax policy: a discretionary government implements inefficiently high tax rates in equilibrium.[80]

## 3.5 Gaussian illustration and sufficient statistics

This section presents an application to a setting with Gaussian implementation shocks. This allows us to derive simpler characterizations of the discretionary and commitment tax policies and to illustrate our findings with numerical simulations. We further provide

[80] It is worth noting that equation (3.23) indicates that the taxation bias arises from a discrepancy between the changes in the posterior distribution when the prior is constant or adjusted. There are at least two reasons for these terms to differ: an anchoring effect (agents anchor their perceptions on their prior) and a debiasing effect (the prior drives the choice of attention). In the model, the adjustment of the prior is the consequence of the equilibrium condition 1.c. This condition may appear to be quite restrictive as it requires a full adjustment of the prior (i.e. rational equilibrium up to a constant bias $b$). The condition in equation (3.23) does not directly rely on this equilibrium assumption and could also be interpreted as follows: if agents' prior beliefs are not irrelevant for their perceptions and somehow related to the equilibrium tax, then we should observe a taxation bias. Furthermore, under the mild condition discussed above, this taxation bias is likely positive if the mapping between the expected prior and the equilibrium tax is strictly positive.

a sufficient statistics formula for the taxation bias that we use to empirically assess its magnitude in the actual US economy.

### 3.5.1 Gaussian discretionary equilibrium

Let the implementation shocks be normally distributed, that is $f_\Theta(\vartheta)$ is the pdf of the Gaussian distribution $\mathcal{N}(0, \sigma_\vartheta^2)$. The common prior distribution $\hat{q}(\tau)$ is then also Gaussian in equilibrium to ensure that priors are consistent with actual tax policy realizations (condition (c) in Definition 1). Because the Gaussian family is self-conjugate with respect to a Gaussian likelihood, agent $w$ posterior distribution $f_{\tilde{\tau}}(\tau|\tau_0, w)$ is Gaussian as well with (type-specific) mean $\mu = \xi\tau_0 + (1-\xi)(\tau_g + b)$ in equilibrium. Introducing these equilibrium conditions into Proposition 7, we characterize the discretionary equilibrium tax policy.

**Proposition 7′.** *Up to a first order approximation of the integrands in Proposition 7, the Gaussian discretionary equilibrium tax policy* ($\tau_g^{\text{eq}}$,$R_g^{\text{eq}}$) *solves*

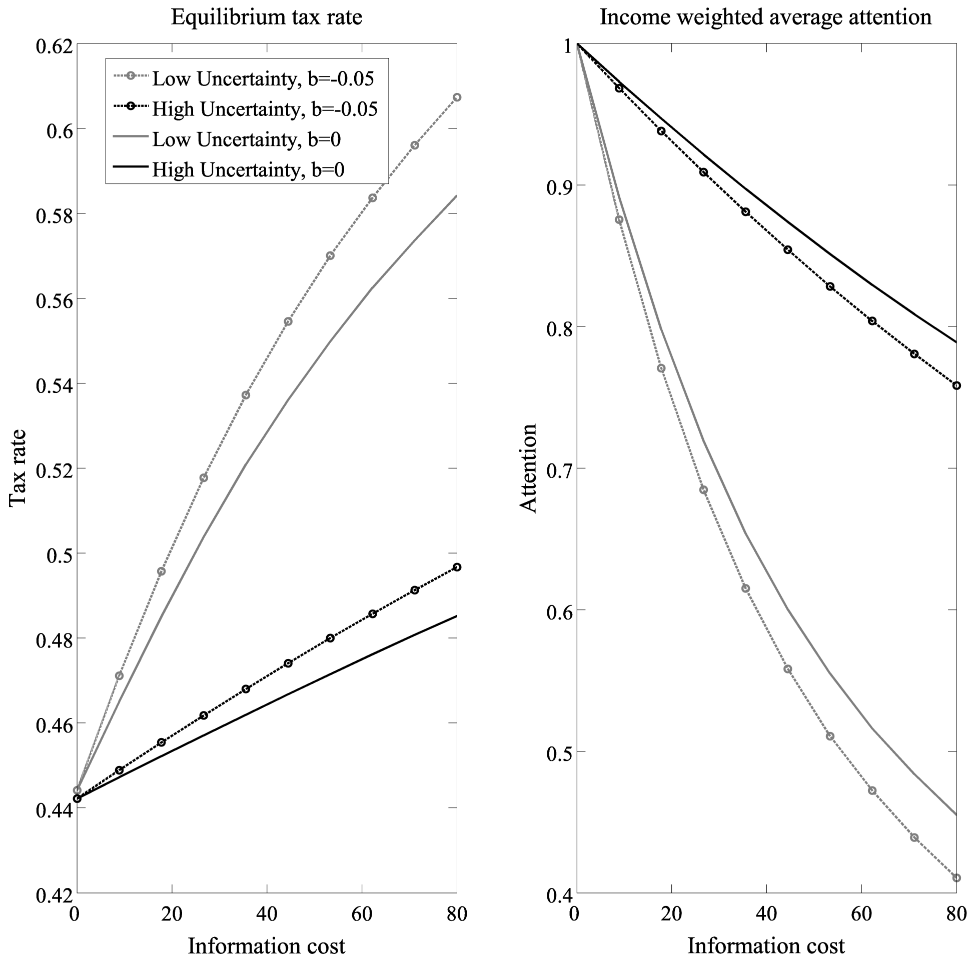
$$E_\vartheta\left[\int \left\{ \underbrace{(1-g)\, y^\star}_{\text{mech. \& wel. effects}} + \underbrace{\left(g(1-\xi)(b-\vartheta)+\tau_0\right)\frac{dy^\star}{d\tilde{\tau}}\xi}_{\text{direct behavioral responses}} \right\}\Big|_{\tilde{\tau}=\mu} dF_w(w)\right] = 0 \quad (3.24)$$

*together with* $E_\vartheta\left[\int g_{|\tilde{\tau}=\mu} dF(w)\right] = 1$*, the government resource constraint (3.18) and where we have introduced social marginal welfare weights* $g \equiv \frac{G'(\mathcal{V})}{p}$.

*Proof.* See Appendix 3.11. □

Equation (3.24) provides a simple expression of taxpayers' direct behavioral responses to tax changes and their impact for the discretionary equilibrium tax policy. Indeed, taxpayers adjust their earnings choice according to changes in their perceived tax rate. Behavioral responses $\frac{dy^\star}{d\tilde{\tau}}$ are thus attenuated by the type-specific attention parameter $\xi$ measuring the fraction of the change in taxes that agents observe. Moreover, the new welfare effect associated with the failure of the envelope theorem is directly proportional to the average size of the error in the posteriors $\mu - \tau_0 = (1-\xi)(b-\vartheta)$ multiplied by social welfare weights $g$. It again captures the corrective motive for taxation in the presence of perception biases.

Figure 3.3: Gaussian discretionary equilibrium




NOTE: The left panel reports the equilibrium target tax rates for different values of the information cost $\kappa$ expressed in \$ /bit /year. The right panel reports the average attention parameter $\xi$ weighted by incomes. Low (resp. high) uncertainty corresponds to Gaussian implementation shocks with a standard deviation equal to 0.05 (resp. 0.1). $b$ is the equilibrium perception bias in agents' prior. The government has a log social welfare function and its policy follows from Proposition 7. Taxpayers have an iso-elastic disutility to work $v(y, w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$. The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data and a Pareto tail for high incomes.

Figure 3.3 plots the tax rates (left panel) and income weighted average attention levels (right panel) in the Gaussian discretionary equilibria for different values of the information cost parameter $\kappa$. In these simulations, the distribution of skills is calibrated from the 2016 March CPS data and extended with a Pareto tail ($k = 2$) for incomes above \$200,000. We assume that the government has a log objective and agents have iso-elastic work disutility given by $v(y, w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ where we set $\epsilon = 1/e$ with the structural elasticity parameter $e = 0.33$ (Chetty, 2012). A detailed presentation of the simulation procedure and the calibration strategy is available in the Online appendix.[81]

These simulations highlight the importance of information rigidities for tax policy. Under discretion, the equilibrium tax rate (left panel) increases substantially when taxpayers are inattentive (right panel). As the information cost parameter $\kappa$ increases, attention decreases and the equilibrium tax rate increases. For example, when the average attention level (weighted by incomes) is equal to 0.8, the tax rate at the rational equilibrium is 47.5% in comparison to a 44% tax rate without information frictions. Introducing a systematic downward bias of 5 percentage points in agents priors further increases the equilibrium tax rate, for instance by 1 to 2 percentage points when $\kappa$ = \$60/bit/year. The influence of the systematic perception bias $b$ strengthens with taxpayers inattention because it is the equilibrium perception bias that matters for tax policy.

Finally, taxpayers are *ceteris paribus* less attentive when there is little prior uncertainty about the tax rate or, equivalently in equilibrium, when the variance of implementation shocks $\sigma_\vartheta^2$ is small. In this case, the government has higher incentives to increase taxes and the discretionary equilibrium tax rate is higher. It should however be noted that the main effect of the parameter $\sigma_\vartheta^2$ is to rescale the mapping between attention levels $\xi$ and the information cost parameter $\kappa$. Indeed equilibrium attention strategies depend on the ratio $\kappa/\sigma_\vartheta^2$ as can be seen from equation (3.14). Therefore, once we consider pairs of $(\kappa, \sigma_\vartheta^2)$ that induce the same (income-weighted) average attention, tax rates in the low and high uncertainty equilibria are similar.[82]

### 3.5.2 Commitment and the taxation bias

With Gaussian implementation shocks, the characterization of the commitment tax policy can be simplified to

**Proposition 8′.** *Up to a first order approximation of the integrands in Proposition 8, the*

[81] Our simulations indicate that the loss in accuracy due to the approximation in Proposition 7′ is very small (with our calibration). Comparing this tax rate to the one obtained directly from Proposition 7, the largest error is smaller than 1% in relative terms.

[82] For example, we find that when the (income-weighted) average attention level is 80%, the difference between the tax rates in the low and high uncertainty equilibria is only equal to 0.2 percentage points.

*Gaussian commitment (equilibrium) tax policy* $(\tau_g^\star, R_g^\star)$ *solves*

$$E_\vartheta\Big[\int\Big\{\underbrace{(1-g)\,y^\star}_{\text{mech. \& wel. effects}}+\underbrace{\Big(g(1-\xi)(b-\vartheta)+\tau_0\Big)\frac{dy^\star}{d\tilde{\tau}}\Big(1-\frac{d\xi}{d\tau_g}(b-\vartheta)\Big)}_{\text{direct \& equilibrium behavioral responses}}\Big\}\Big|_{\tilde{\tau}=\mu}dF_w(w)\Big]=0 \quad (3.25)$$

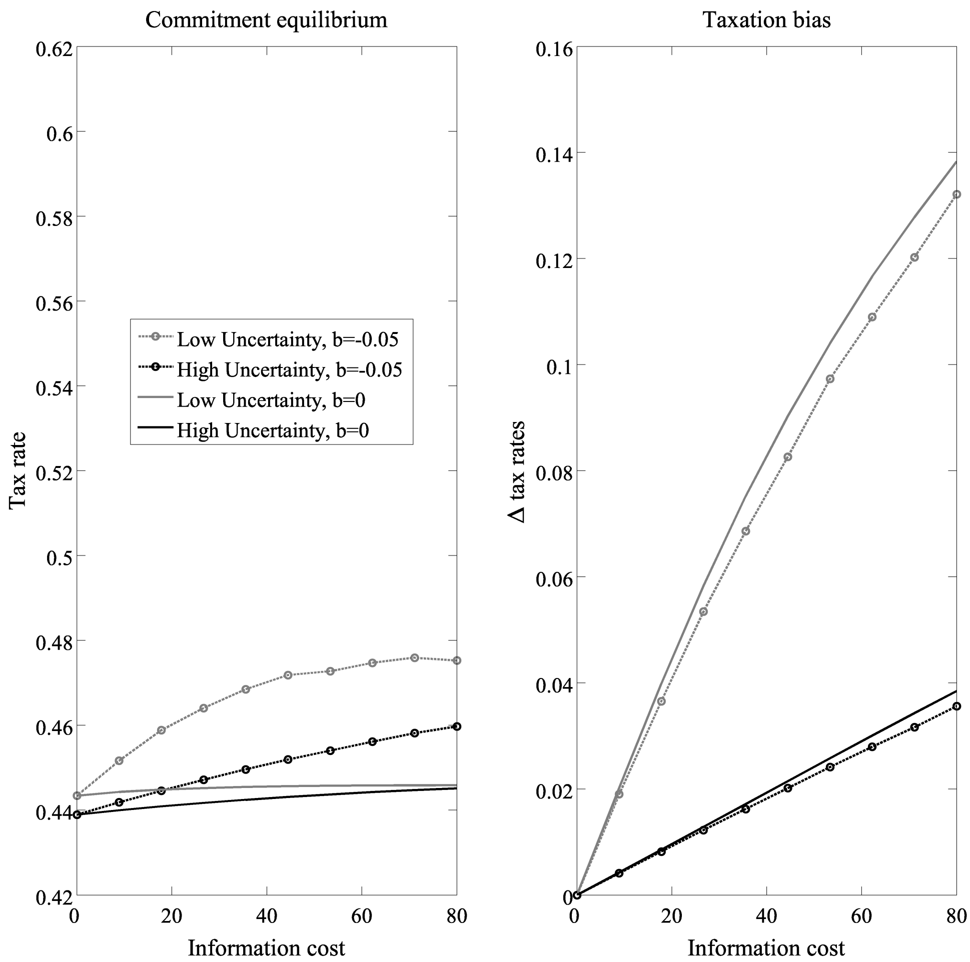
*together with* $E_\vartheta\Big[\int g_{|\tilde{\tau}=\mu}dF(w)\Big]=1$*, the government resource constraint (3.18) and where we have introduced social marginal welfare weights* $g\equiv\frac{G'(\mathcal{V})}{p}$*.*

*Proof.* See Appendix 3.11. □

The direct and equilibrium responses to a change in taxes is now captured through the term $\frac{dy^\star}{d\tilde{\tau}}\Big(1-\frac{d\xi}{d\tau_g}(b-\vartheta)\Big)$. In the latter, the factor 1 stands for the fact that the government accounts for the equilibrium adjustment of priors when setting its policy. The supplemental term $\frac{d\xi}{d\tau_g}(b-\vartheta)$ captures the effect of changes in attention $\xi$ following a marginal increase in the tax rate on equilibrium tax perceptions. This term vanishes (in expectation) when agents correctly anticipate the tax rate ($b=0$) and prompts the government to decrease the tax rate when agents underestimate tax rates ($b<0$). Indeed, increasing the tax rate increases attention and thereby reduces tax underestimation in equilibrium which is detrimental to efficiency.

In the left panel of Figure 3.4 we report the commitment equilibrium tax rates for different values of the information cost parameter $\kappa$. In the rational equilibrium ($b=0$), the tax rate is only marginally higher in the presence of information frictions. Indeed, the policymaker finds it optimal to marginally increase taxes to prompt taxpayers to be more attentive to implementation shocks. In downward biased equilibria ($b<0$), the government further increases the tax rate to exploit the efficiency gains from agents tax underestimation. Indeed, because taxpayers remain inattentive in equilibrium and the prior is downward biased, perceived tax rates are lower than the actual rate. Ultimately, this underestimation of tax rates reduces the efficiency costs of taxation allowing for tax increases. The leverage to increase tax rates is however limited here because the policymaker realizes that it also prompts increases in agents' prior and attention. This results in an increase of perceived tax rates which ultimately increases the efficiency costs of taxation. Consequently, commitment equilibrium tax rates are much smaller than discretionary equilibrium tax rates depicted in Figure 3.3.

Figure 3.4: Taxation bias




NOTE: Difference between positive and normative tax rates for different values of the information cost $\kappa$ expressed in annual \$ / bit. Low (resp. high) uncertainty corresponds to Gaussian implementation shocks with a standard deviation equal to 0.05 (resp. 0.1). $b$ is the equilibrium perception bias in agents' prior. The government has a log social welfare function and its policy follows from Proposition $8'$ for the commitment equilibrium. Taxpayers' behavior relies on Assumptions 3.1-3.5 and an iso-elastic disutility to work $v(y,w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data.

As a consequence, and as depicted in the right panel of Figure 3.4, the taxation bias increases as the information cost parameter $\kappa$ grows. Even small information frictions generate a significant taxation bias. Our simulations indicate that in a rational equilibrium, there is a taxation bias of 4 (resp. 3.5) percentage points in the presence of low (resp. high) uncertainty when the income-weighted average attention parameter is 0.8. Moreover, the taxation bias is above 10 percentage points when the income-weighted average attention falls below 0.55.

The taxation bias can thus lead to significant upward distortions in actual tax rates when the income-weighted average attention turns out to be low. We now show theoretically that this is indeed a key sufficient statistic to empirically assess the magnitude of the taxation bias.

### 3.5.3 Sufficient statistics formulas and taxation bias in the US

We derive sufficient statistics formulas for the equilibrium tax policy under discretion and commitment that echo textbook optimal tax formulas and that we combine to obtain a sufficient statistics formula for the taxation bias. To obtain simple sufficient statistics formulas we further assume that preferences are iso-elastic such that the structural labor supply elasticity $e$ – i.e. computed with respect to the perceived marginal net-of-tax rate – is constant and that implementation shocks and perception biases are small.

**Corollary 1.** *A sufficient statistics formula for the Gaussian discretionary equilibrium tax rate characterized in Proposition 7′ is*

$$\tau_g^{\mathrm{eq}} \simeq \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - b\ \frac{\overline{g(1-\xi)y^\star\xi}\ e}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} \tag{3.26}$$

*where all endogenous right hand side quantities are evaluated at $\tau_g^{\mathrm{eq}}$ and we have introduced the mean operator $\bar{x} \equiv \int x(w)f(w)dw$.*

*Proof.* See Appendix 3.11.1. □

The first term in equation (3.26) corresponds to the textbook optimal linear tax formula up to the presence of the income weighted average attention $\overline{y^\star\xi}$. The second term corresponds to the corrective motive of taxation in the presence of perception biases.

**Corollary 2.** *A sufficient statistics formula for the Gaussian commitment equilibrium tax rate characterized in Proposition 8′ is*

$$\tau_g^\star \simeq \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} - b\ \frac{\overline{g(1-\xi)y^\star}\ e}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} \tag{3.27}$$

*where all endogenous right hand side quantities are evaluated at $\tau_g^\star$ and we have introduced the mean operator $\bar{x} \equiv \int x(w)f(w)dw$.*

*Proof.* See Appendix 3.11.1. □

The first term now exactly coincides with the textbook optimal linear tax formula while the second term again corresponds to the corrective motive of taxation in the presence of perception biases. It should however be noted that corrective terms in equations (3.26) and (3.27) are not identical. Their sign are however identical and both are proportional to the perception bias $b$.

Focusing on near-rational equilibria – that is equilibria with small perception biases $b \simeq 0$ such that corrective terms are second-order – we obtain a simple sufficient statistics formula for the taxation bias.

**Proposition 10.** *A sufficient statistics formula for the taxation bias in Gaussian near-rational equilibria is*

$$\tau_g^{\mathrm{eq}} - \tau_g^\star \simeq \frac{\overline{(1-\xi)y^\star}}{\overline{(1-g)y^\star}}\; e\; t^2 \tag{3.28}$$

*where all endogenous right hand side quantities are evaluated at the actual tax rate $t$.*

*Proof.* See Appendix 3.11.1. □

This simple formula for the taxation bias is reminiscent of the one provided Section 3.2 and is a generalization to a situation in which the government has tastes for redistribution and agents' attention is endogenous and hence type-specific.[83] The income-weighted average attention $\overline{\xi y^\star}$ – or equivalently inattention $\overline{(1-\xi)y^\star}$ – thus becomes a key sufficient statistic for the taxation bias. *Ceteris paribus*, the taxation bias increases with the structural elasticity of labor supply $e$, with the square of the actual tax rate $t$ and decreases with the government redistributive tastes.[84]

In an attempt to gauge the empirical magnitude of the taxation bias in the actual US economy, we bring this sufficient statistics formula to the data. The meta-analysis

[83] Equation (3.28) can also be expressed in terms of covariances as

$$\tau_g^{\mathrm{eq}} - \tau_g^\star \simeq \frac{\mathrm{cov}(\xi, y^\star) - \overline{(1-\xi)}\,\overline{y^\star}}{\mathrm{cov}(g, y^\star)}\; e\; t^2$$

[84] Intuitively, the taxation bias *increases* with the government's redistributive tastes as it relatively increases the incentives to implement unanticipated tax increases. However, this first-order effect here transits through an increase in the actual tax rate $t$ and we get the inverse relationship controlling for $t$.

of Gabaix (2019) combines existing measures of attention to sales taxes to trace out the evolution of average attention with the stakes. We find that income taxes in the US are well approximated by a linear tax schedule with a tax rate of $t = 29.46\%$ which would correspond to an average attention parameter of about 0.70. Focusing on the US personal income tax, Rees-Jones and Taubinsky (2020) estimate that agents' attention parameter to their marginal tax rate is equal to 0.81. Accordingly, we consider an average attention of 0.75 to taxes as our baseline. We are then able to compute the associated income-weighted average attention using our model of endogenous attention and the actual distribution of income.

Turning to other sufficient statistics, we take the structural elasticity parameter $e = 0.33$ estimated by Chetty (2012) and use an inverse optimum approach to deduce the US government's redistributive tastes from the actual tax policy.[85]

Table 3.1: Estimated taxation bias in the actual US economy

| Taxation bias (percentage points) | | Average attention parameter 0.65 | 0.75 | 0.85 |
|---|---|---|---|---|
| Benthamite | 0.27 | 5.21 | **3.66** | 2.17 |
| redistribution | 0.50 | 3.11 | 2.18 | 1.29 |
| parameter | 1.00 | 1.85 | 1.30 | 0.77 |

NOTE: Our estimation of the taxation bias (in percentage points) follows from the characterization in Proposition 10. A larger Benthamite parameter corresponds to a more redistributive objective. The value in bold corresponds to our baseline estimate for the 2016 US economy.

In the actual US economy, we estimate that the taxation bias is roughly equal to 3.66 percentage points in our baseline calibration. This means that the US income tax rate is 12% higher than what would be optimal holding the government's redistributive objective constant. Table 3.1 provides a sensitivity analysis varying average attention and the government's redistributive objective. For the latter we use a Benthamite social welfare function for which we vary the value of the parameter that shapes the desire for redistribution. The value of 0.27 closely approximates the welfare weights we estimate using an inverse optimum approach and a value of 1 corresponds to a logarithmic social welfare function – which captures rather extreme redistributive tastes. For realistic redistributive tastes and attention parameters, the magnitude of the taxation bias in the actual US

[85] That is, we deduce $\overline{(1-g)y^{\star}}$ from equation (3.26) assuming $b \simeq 0$.

economy ranges from 1.29 to 5.21 percentage points and our baseline estimate of 3.66 lies in the middle of this range.

## 3.6 Welfare implications

This section analyzes the welfare implications of information frictions. It first decomposes the variation in aggregate social welfare between potential welfare gains that may be attained with information rigidities (commitment) and the welfare losses associated to actual policy distortions (discretion). It then quantifies the relative importance of the different channels through which information rigidities ultimately affect welfare and redistribution at the individual level.

### 3.6.1 Information rigidities and aggregate welfare

Let $SW^{\text{eq}}(b,\kappa)$ be the social welfare from equation (3.17) evaluated at the discretionary equilibrium. The total welfare impact of information rigidities writes $\Delta SW^{\text{eq}}(b,\kappa) \equiv SW^{\text{eq}}(b,\kappa) - SW^{\text{eq}}(0,0)$.[86] It may be decomposed between the potential welfare gains from misperceptions and the welfare costs induced by the taxation bias as follows

$$\Delta SW^{\text{eq}}(b,\kappa) = \underbrace{SW^{\text{eq}}(b,\kappa) - SW^{\star}(b,\kappa)}_{\text{Taxation bias } (\leq 0)} + \underbrace{SW^{\star}(b,\kappa) - SW^{\star}(0,0)}_{\text{Potential gains}} \tag{3.29}$$

where $SW^{\star}(b,\kappa)$ is the social welfare attained under commitment.

The welfare impact of the taxation bias is negative since the commitment tax policy is by definition the welfare-maximizing feasible policy. As a result, information rigidities are welfare improving if and only if this negative welfare impact is dominated by the welfare gains induced by misperceptions, that is $|SW^{\text{eq}}(b,\kappa) - SW^{\star}(b,\kappa)| \leq SW^{\star}(b,\kappa) - SW^{\text{eq}}(0,0)$.

This condition requires a downward bias in priors $b < 0$ such that agents underestimate tax rates. This can be easily seen when looking at the sufficient statistics formula for the commitment tax rate in equation (3.27). Indeed, when $b = 0$ the commitment tax rate

[86] Note that $SW^{\text{eq}}(0,b) = SW^{\text{eq}}(0,0)$ since as soon as the information cost $\kappa$ is nil, agents have perfect information and whether priors are biased is irrelevant.

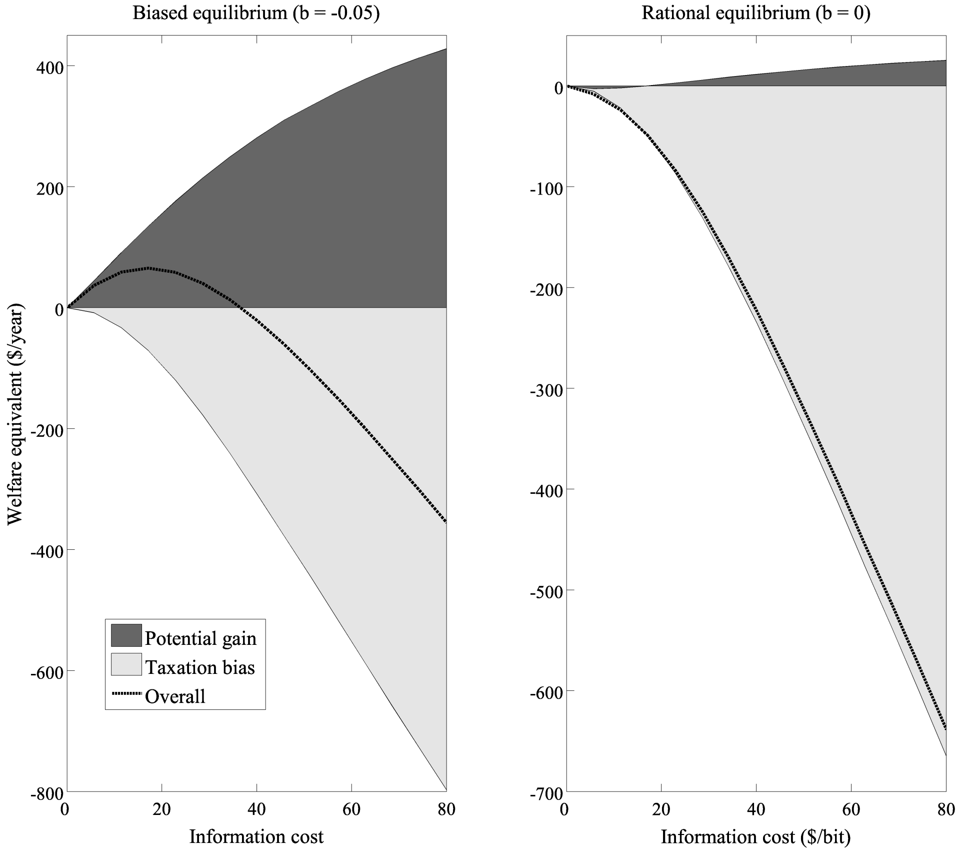
is equal to the optimal tax rate without information up to a first order approximation. Therefore, there cannot be first order gains from information rigidities. However, equation (3.28) indicates that the taxation bias is nonetheless positive. Consequently, the total welfare impact is negative. Panel B of Figure 3.5 illustrates this result for Gaussian implementation shocks.[87]

While a negative perception bias is necessary for information rigidities to be welfare improving, it is not a sufficient condition. Information rigidities should also not be too large to ensure taxpayers are sufficiently attentive to tax policy. Indeed, as the information cost parameter $\kappa$ grows, the welfare losses induced by the taxation bias increase more rapidly than the welfare gains from the negative perception bias. Indeed, the former is convex while the latter is concave.

Panel A of Figure 3.5 illustrates this mechanism with Gaussian implementation shocks. When the downward equilibrium perception bias is equal to 5 percentage points, the welfare gains induced by information rigidities dominate the welfare cost induced by the taxation bias as long as the information cost parameter $\kappa$ is lower than \$25/bit/year. Above this threshold, inattention and the associated deviation from the commitment policy become too important such that information rigidities are welfare decreasing. Therefore, downwards biases in perceived marginal tax rates will be typically associated with a *decrease* – rather than an *increase* – in aggregate social welfare when agents are not sufficiently attentive to tax policy.

[87]Strangely enough, the potential gain is first decreasing and then increasing when $b$ is small or nil. While the magnitude of the potential gain is small and thus negligible in comparison to the impact of the taxation bias, it deserves to be briefly explained. Consider the two extreme cases where $\kappa = 0$ and $\kappa \mapsto \infty$. Hence, $\xi$ is respectively equal to one or zero for each taxpayer. Everything else being equal, aggregate earnings are larger when $\kappa \mapsto \infty$ as agents behave as if there were no implementation shocks when deciding how much to earn (individual earnings are a concave function of the perceived rate), while they fully adjust to these shocks when $\kappa = 0$. Consequently, the potential gain converges to a positive value as $\kappa$ tends to infinity. However, when $\kappa$ is small but strictly positive, some taxpayers noisily observe the implementation shocks so that the variance of their earnings choices increases. Ultimately, it lowers aggregate earnings and generates a negative and decreasing potential gain for small values of $\kappa$. Simulations indicate that the above described variations in aggregate earnings dominate other second order effects (e.g. misoptimization costs).

Figure 3.5: Welfare decomposition




NOTE: Welfare decomposition from equation (3.29) for different values of the information cost $\kappa$. The standard deviation for the Gaussian implementation shocks is equal to 0.05. $b$ is the equilibrium perception bias in agents' prior. The government has a log social welfare function and its policy follows from Proposition **??**. Taxpayers have an iso-elastic disutility to work $v(y, w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data.

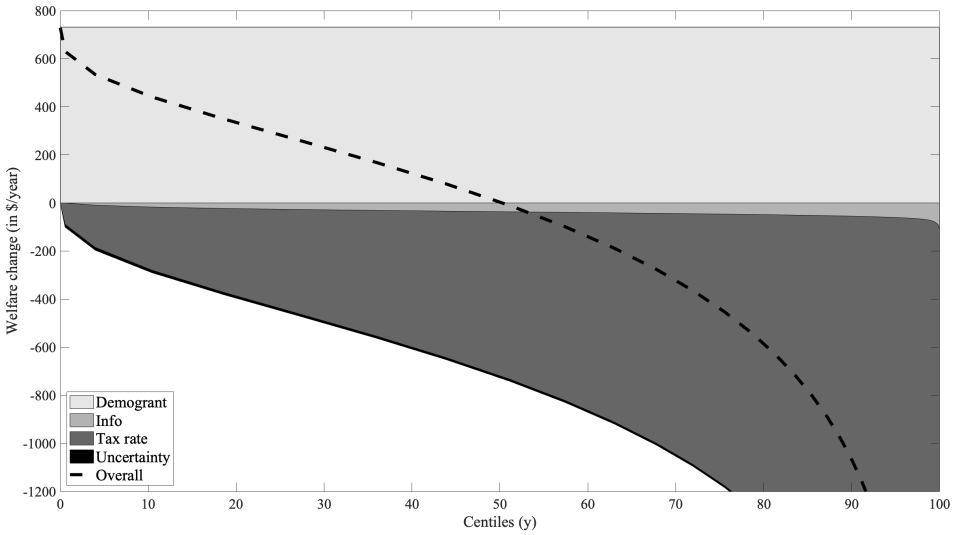
### 3.6.2 Redistributive impacts

We now turn to an analysis of the welfare implications of information rigidities at the individual level. Let $\Delta\mathcal{V} \equiv \mathcal{V}_\kappa(\tau_g^{\text{eq}}(\kappa), w) - \mathcal{V}_0(\tau_g^\star(0), w)$ be the variation in the expected utility of a taxpayer with skill $w$ between the discretionary equilibrium when the information cost is $\kappa$ and a counterfactual with perfect information. Using a quasi-linear and separable utility function allows us to decompose the variation in expected utility induced by information rigidities in the following way (see Appendix 3.13 for precise definitions)

$$\Delta\mathcal{V} \;=\; \Delta^{\mathcal{V}}R + \Delta^{\mathcal{V}}\tau_g + \Delta^{\mathcal{V}}\text{info cost} + \Delta^{\mathcal{V}}b + \Delta^{\mathcal{V}}\text{uncertainty} \tag{3.30}$$

that is, the welfare impact of information rigidities at the individual level arises from a variation in the demogrant $R$, the tax rate $\tau_g$, the cost of information acquisition $\kappa\mathcal{I}(\sigma^\star)$, the misoptimization costs induced by potential perception bias $b$, and the change in overall uncertainty. The increase in the demogrant has a positive effect on expected utility while all other terms are negative.

Figure 3.6: Variation in expected utility




NOTE: Expected utility decomposition from equation (3.30) by deciles of the productivity level $w$. The standard deviation for the Gaussian implementation shocks is equal to 0.1. There is no perception bias in agents' prior $b = 0$ and the information cost is $\kappa = 50$. The government has a log social welfare function and its policy follows from Proposition 7′. Taxpayers' behavior relies on Assumptions 3.1-3.5 and an iso-elastic disutility to work $v(y, w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_w(w)$ is calibrated using 2016 CPS data.

Figure 3.6 plots the above expected utility decomposition with Gaussian implementation shocks ($\sigma_\vartheta = 0.1$), no perception bias ($b = 0$) and an information cost $\kappa$ of \$50/bit per year.

Information rigidities induce policy distortions that create losers and winners. Indeed, the redistributive impact of information rigidities is driven by the increase in the tax rate and thus in redistribution through the demogrant. This naturally benefits low skill workers at the extent of high skill workers.

Somewhat surprisingly, information costs represent a relatively small deadweight loss for society in comparison to the large indirect impact of information frictions on tax policy and welfare. Moreover, it turns out that these information costs are higher for high skill workers because they have higher incentives to collect information and are thus more attentive. Extending our analysis to non-linear tax schedules, we show in the next section that this regressivity of attention has an impact on actual tax progressivity.

## 3.7 Tax progressivity and the taxation bias

In this section we extend the analysis to non-linear tax schedules. We find that the taxation bias becomes heterogeneous across income levels and ultimately reduces tax progressivity.

### 3.7.1 Introducing nonlinear tax schedules

We allow the government to use a nonlinear tax schedule $T(y)$ but the setup introduced in Section 3.3 is otherwise unchanged. In particular, we maintain Assumption 2 that individuals use a linear representation of the tax schedule $\tilde{T}(y) = \tilde{\tau}y - \tilde{R}$ which now raises a new question: in the continuum of marginal tax rates $\{T'(y)\}_y$, what is the marginal tax rate $T'(y_w)$ agent $w$ gathers information about?

Absent income effects, the perceived marginal tax rate $\tilde{\tau}_w$ remains a sufficient statistics for labor supply and uniquely pins down earnings $y^\star(\tilde{\tau}_w; w)$.[88] Using this mapping, we define an agent $w$ ex ante – before information acquisition – optimal earnings level $\hat{y}_w = y^\star(\hat{\tau}_w; w)$ and make the following assumption.

**Assumption 5** (prior reliance)**.** *Taxpayer $w$ gathers information aboute the actual marginal tax rate $\tau_0(\hat{\tau}_w, w) \equiv T_0'(\hat{y}_w)$ at her ex-ante optimal earnings level $\hat{y}_w$.*

Essentially, Assumption 5 guarantees the internal consistency of the perception formation process by ensuring agents have no additional information ex ante than that contained in their prior. Moreover, it gives a novel allocative role for the prior as taxpayers now

[88] While the perceived marginal tax rate $\tilde{\tau}$ was already type-specific in the previous sections – through type-specific attention choices –, we here introduce the subscript $w$ to emphasize that it will in addition be type-specific through agents' type-specific priors (see Assumption 5).

linearize the tax schedule around the income level that they deem optimal ex ante.[89]

The presence of a nonlinear tax schedule does not fundamentally affect equilibrium concepts. Two refinements are nevertheless necessary. First, for the sake of simplicity we assume that implementation shocks $\vartheta$ uniformly affect marginal tax rates at all earnings levels $y$ such that $T_0'(y) = T_g'(y) + \vartheta$. Second, the equilibrium condition $(c)$ from Definition 1 – which characterizes the equilibrium adjustment of priors – now becomes

*(c′) The type-specific prior distribution $\hat{q}_w(\tau)$ is the pdf of $T_g'(\hat{y}_w) + b + \vartheta$.*

That is, each taxpayer's prior is consistent with her marginal tax rate of interest up to an arbitrary perception bias $b$. Incidentally, the prior average $\hat{\tau}_w \equiv E_{\hat{q}_w(.)}[\tau]$ is thus necessarily type-specific in equilibrium when the government implements a nonlinear tax schedule. While natural in our context, this poses a potential challenge for the resolution of this nonlinear tax model.

We rely on a perturbation approach in order to derive the optimal tax schedule. Following Jacquet and Lehmann (2017), one needs three assumptions to solve for the optimal non-linear tax schedule using a tax perturbation approach. (i) The tax function $T_g(.)$ must be twice differentiable. (ii) The optimization program of each taxpayer must admit a unique global maximum. (iii) Agents' second-order conditions must hold strictly. While (i) and (ii) are generic requirements to ensure the global smoothness of the problem so that tax perturbations will not induce individuals to jump between different maxima, condition (iii) has less intuitive consequences.

In standard models, condition (iii) – combined with a single crossing assumption on individuals preferences – ensures the existence of an increasing mapping between earnings $y$ and skills $w$. This is known as a monotonicity condition on allocations.[90] It is a

[89]To illustrate this new allocative role, consider the limit where the information cost $\kappa$ goes to zero. Perceptions are then perfect $\tilde{\tau}_w = \tau_0(\hat{\tau}_w, w)$ and each agent chooses earnings $y_{w|\tilde{\tau}_w} = y^\star(\tau_0(\hat{\tau}_w, w); w)$. This is in contrast to the full information case in which earnings are the solution to a fixed-point problem characterized by $y_w = y^\star(T_0'(y^\star(.)); w)$. In a rational equilibrium ($b = 0$), both income concepts coincide. They will however differ in biased equilibria $b \neq 0$.

[90]It follows from agents incentive constraints in a mechanism design approach.

requirement for the tax perturbation approach which disciplines the curvature of the tax function $T''_g(.)$. Here, allowing for type-specific priors and a perception bias $b$ poses a potential threat to the existence of an increasing mapping between earnings $y$ and skills $w$. In the Online Appendix, we show that under our assumptions the monotonicity condition is also expected to hold when $T''_g(.)$ is smooth enough. As a result, we solve for the optimal tax schedule assuming the monotonicity condition is verified and check ex post that it holds at the optimum.

### 3.7.2 ABCD tax formula

We can now solve for the optimal nonlinear tax schedule. The government chooses a target nonlinear tax schedule $T_g(.)$ that consists in a continuum of marginal tax rates $\{T'_g(y)\}_y$ and a tax level indexed by the demogrant $T_g(0)$. It is implemented up to an implementation shock $\vartheta$ on marginal tax rates and the tax level adjusts such as to satisfy the government budget constraint ex post. The government problem writes

$$\max_{T'_g(.),T_g(0)} \quad E_\vartheta\left[\iint G\Big(\mathcal{V}(\tilde{\tau}_w, T_0(.);\kappa,w)\Big)\ f_{\tilde{\tau}}(\tau|\tau_0(\hat{\tau}_w,w);w)\ f_w(w)\ d\tau dw\right] \tag{3.31}$$

$$\text{s.t.} \quad \iint T_0(y^\star(\tilde{\tau}_w;w))\ f_{\tilde{\tau}}(\tau|\tau_0(\hat{\tau}_w,w);w)\ f_w(w)\ d\tau dw \geq E \tag{3.32}$$

where $E$ is an exogenous expenditure requirement, $f_{\tilde{\tau}}(\tau|\tau_0(\hat{\tau},w);w)$ is the posterior distribution of agent $w$ perceived tax rate and with the indirect utility function

$$\mathcal{V}(\tilde{\tau}_w, T_0(.);\kappa,w) = y^\star(\tilde{\tau}_w;w) - T_0(y^\star(\tilde{\tau}_w;w)) - v\Big(y^\star(\tilde{\tau}_w;w);w\Big) - \kappa\mathcal{I}(\sigma^\star) \tag{3.33}$$

We solve this problem using a perturbation approach. Namely, we consider the effect of a reform that consists in a small increase $\Delta\tau^r$ in marginal tax rates in a small bandwidth of earnings $[y^r - \Delta y, y^r]$ and characterize its impact on the objective function of the government. Following the tax perturbation literature, this reform may be apprehended through three mechanisms: a mechanical effect, a welfare effect and a behavioral effect. However, analyzing the impact of a reform in this setting with information frictions in tax perceptions calls for a careful identification of the agents affected by the reform.

The standard mechanical and welfare effects capture the change in taxes and welfare for individuals $w$ whose earnings are higher than $y^r$ given their perceived tax rates $\tilde{\tau}_w$. Following from the aforementioned monotonicity condition, it corresponds to all agents with a productivity $w \geq w^r$ where $y^\star(\tilde{\tau}_{w^r}; w^r) \equiv y^r$. In contrast, the behavioral effect comes from taxpayers who are learning the marginal tax rate affected by the reform. That is, all agents whose ex ante optimal earnings level $\hat{y}$ belong to $[y^r - \Delta y, y^r]$. Again using the monotonicity condition, we can equivalently identify these agents as those with a productivity $w \in [\hat{w}^r - \Delta\hat{w}, \hat{w}^r]$ where $y^\star(\hat{\tau}_{\hat{w}^r}; \hat{w}^r) \equiv y^r$. Note that the two cut-offs $w^r$ and $\hat{w}^r$ differ almost surely.[91]

We then characterize the discretionary and commitment equilibrium tax schedules assuming Gaussian implementation shocks. As before, these conditions are easier to interpret after applying a small implementation shocks approximation which is what we report in Proposition 11, relegating general conditions to the Online Appendix.

**Proposition 11** (ABCD formula)**.** *Assuming small Gaussian implementation shocks, the equilibrium non-linear tax schedule is to a first-order approximation characterized by*

$$\begin{aligned} &\frac{T_g'(y^\star(\mu_{\hat{w}^r}; \hat{w}^r)) + g(\hat{w}^r)|_{\tilde{\tau}=\mu_{\hat{w}^r}}\Big(\mu_{\hat{w}^r} - T_g'(y^\star(\mu_{\hat{w}^r}; \hat{w}^r))\Big)}{1-\mu_{\hat{w}^r}} \\ &\qquad = \frac{1}{e^{\frac{d\mu_{\hat{w}^r}}{d\tau_g}}|_{\tilde{\tau}=\mu_{\hat{w}^r}}} \frac{1}{y^\star(\mu_{\hat{w}^r}; \hat{w}^r)} \frac{\frac{dy^\star(\hat{\tau}_{\hat{w}^r}; \hat{w}^r)}{dw}}{f_w(\hat{w}^r)} \int_{w^r}^{\infty} \Big(1 - g(w)|_{\tilde{\tau}=\mu_w}\Big) f_w(w)\, dw \end{aligned} \tag{3.34}$$

*together with the transversality condition $\int g(w)_{|\tilde{\tau}_w=\mu_w}\ dF(w) = 1$ and the government budget constraint (3.32), where all endogenous quantities are evaluated at their equilibrium values.*

*Moreover the ex-post average perceived marginal tax rate is $\mu_w \equiv \xi\tau_0(\hat{\tau}_w, w) + [1-\xi]\hat{\tau}_w$ such that $\frac{d\mu_w}{d\tau_g} = \xi$ under discretion and $\frac{d\mu_w}{d\tau_g} = 1 + \frac{d\xi}{d\tau_g}[\tau_0(\hat{\tau}_w, w) - \hat{\tau}_w]$ under commitment.*

*Proof.* See Online Appendix. □

Under commitment and absent perception biases ($b = 0$), the ABCD formula boils

[91] The two cut-offs coincide only when $\hat{\tau}_w = \tilde{\tau}_w$. That is, when $b = 0$ and $\vartheta = 0$. Since we focus on Gaussian implementation shocks here, it is never the case (a.s.).

down to the ABC formula derived in Diamond (1998) and the standard interpretation prevails.[92] The presence of perception biases ($b \neq 0$) has several effects. First, it creates a disconnect between $w^r$ and $\hat{w}^r$ new to this non-linear setting. Second, it adds a welfare effect ($g(\hat{w}^r)(1-\xi)b$ in the numerator of the LHS) related to the failure of the envelope theorem when agents misoptimize. Third, it adds an efficiency term (fraction on the RHS with $\frac{d\mu_w}{d\tau_g} = 1 + \frac{d\xi}{d\tau_g}[\tau_0(\hat{\tau}_w, w) - \hat{\tau}_w]$ in the denominator) accounting for the variation in agents equilibrium misperception when their attention $\xi$ changes in response to tax reforms.

As before, the emergence of a taxation bias comes from the discrepancy between the estimated impact of a reform under discretion and under commitment. Under discretion, the government fails to internalize the equilibrium impact of the reform on perceptions and accordingly considers that an increase $\Delta\tau^r$ in marginal tax rates only increases perceived marginal tax rates by $\xi\Delta\tau^r \leq \Delta\tau^r$. Increasing marginal tax rates is thus perceived as less costly in terms of efficiency than it really is. As a result, marginal tax rates are in a discretion equilibrium higher than in a commitment equilibrium. In other words, marginal tax rates are higher than they should be from a normative perspective: this is the taxation bias.

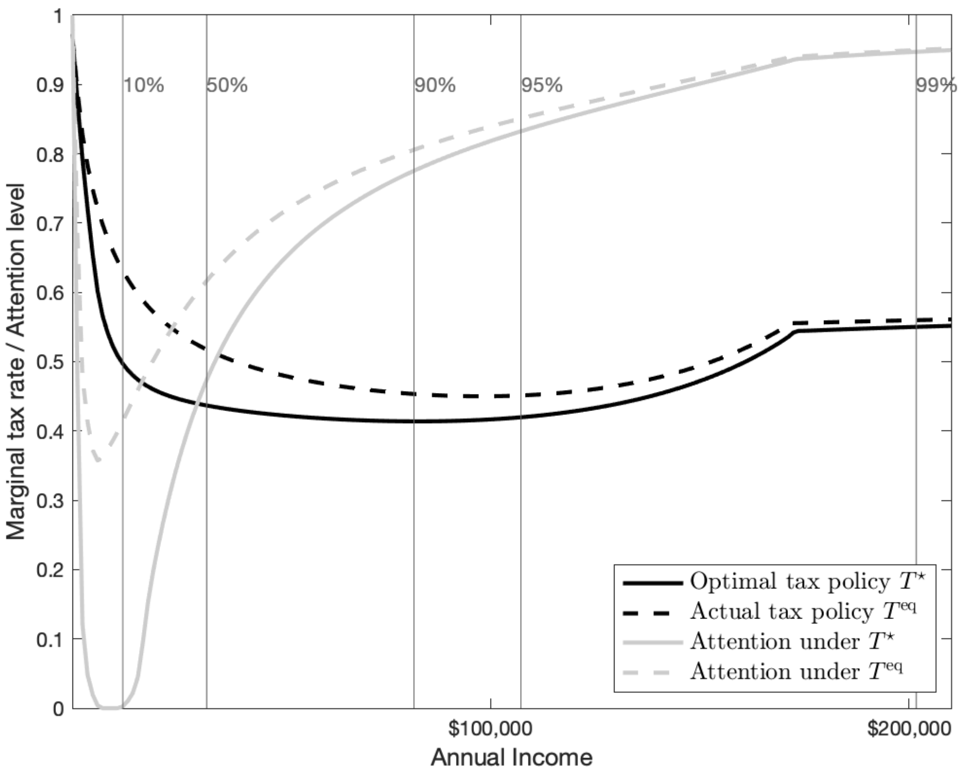
What is new to this non-linear setting is that the taxation bias affecting the marginal tax rate $T_g'(y^r)$ at a given level of earnings $y^r$ is driven by the attention level of agents of type $\hat{w}^r$. Surprisingly, agents $\hat{w}^r$ may not even be located at earnings $y^r$ in the presence of perception biases. More importantly, if attention levels vary across the earnings distribution, the taxation bias will have an impact of the progressivity of the tax schedule.

### 3.7.3 Numerical illustration

To illustrate this property we represent in Figure 3.7 the target nonlinear tax schedules implemented under a discretion equilibrium (actual tax policy – dashed black line) and under commitment (optimal tax policy – full black line). Simulations are carried out absent systematic perception biases ($b = 0$) such that the optimal nonlinear tax schedule corresponds to the textbook optimal nonlinear tax schedule of Saez (2001). We thus

[92]The additional term on the RHS disappears since $\frac{d\mu_w}{d\tau_g} = 1$ while we have that $\hat{\tau}_w^{\mathrm{eq}} \approx \tilde{\tau}_w^{\mathrm{eq}}$ ensuring both $w^r \approx \hat{w}^r$ and $\tilde{\tau}_w^{\mathrm{eq}} \approx T_g'(y^\star(\mu_{\hat{w}^r}^{\mathrm{eq}}; \hat{w}^r))$.

Figure 3.7: Target non-linear tax schedules




NOTE: Target non-linear tax schedules (black curves) and attention levels (grey curves) under a discretion equilibrium (actual tax policy) and under commitment (optimal tax policy) for a value of the information cost $\kappa = 30\$/bit/yr$. The government has a log social welfare function and follows the optimal policy from Proposition 11. Taxpayers' have an iso-elastic disutility to work $v(y, w) = (y/w)^{1+\epsilon}/(1+\epsilon)$ with $\epsilon = 1/0.33$ (Chetty, 2012). The distribution of skills $f_W(w)$ is calibrated using 2016 CPS data and extended with a Pareto-tail of parameter $a = 2$ (Saez, 2001).

naturally retrieve the known U-shape pattern of marginal tax rates.

Because of the taxation bias, marginal tax rates are higher under discretion than under commitment. Strikingly, this difference in marginal tax rates is not constant across earnings levels. For instance, agents located at the first decile (resp. the median) of the earnings distribution face a marginal tax rate of 50% (resp. 44%) under commitment and a marginal tax rate of 63% (resp. 52%) under discretion. In contrast, the marginal tax rate faced by individuals in the top decile (resp. top percentile) increases by at most 4 (resp. 1) percentage points. This reflects the impact of the taxation bias on the progressivity of the tax schedule coming from the variation in attention $\xi$ across earnings.

Attention levels represented in Figure 3.7 (grey lines) are indeed generally increasing in earnings. In our model, more productive agents have intuitively more latitude to choose the earnings level they see fit and attach thus a higher value to being informed about the tax schedule. As a result, attention globally increases with productivity and thus – through the monotonicity condition – with earnings.[93] Note that this pattern is obtained assuming all individuals have the same cost of information $\kappa$. Therefore, assuming that more able workers are also more efficient at collecting information would only reinforce the striking result that, because it decreases with income, inattention to taxes induces regressive tax increases.

## 3.8 Conclusions

We develop a positive theory of tax policy in the presence of information frictions and show that agents' inattention to taxes leads to a taxation bias. We find that this taxation bias is undesirable, large and regressive. Because it reflects a commitment problem, our results suggest that the welfare gains from using precommited tax rules could be large. Alternatively, the model identifies a key parameter to limit the government's deviations from the optimal tax policy: the information cost $\kappa$. Therefore, it would be interesting for future research to investigate the determinants behind information costs driving agents' inattention. Indeed, if the latter are related to the complexity of tax systems, monitoring and restricting tax complexity may be a simple and effective way to prevent the implementation of such inefficient and regressive tax increases.

While some of our results may be model-specific, our analysis sheds a new light on the welfare consequences of information frictions in agents' tax perceptions. It underlines that downward biases in tax perceptions are not necessarily welfare improving. They do lower the efficiency costs of taxation in existing tax systems, but existing tax systems without misperceptions are arguably not the right counterfactual to use for welfare analysis. Indeed, there may be other (equilibrium) effects at play – which here take the form of a taxation bias. We show that the welfare consequences of such effects may be dominant

[93]The pattern reverts at the very beginning of the earnings distribution because as the marginal tax rate approaches one, we approach the origin of the labor supply function where earnings become infinitely responsive to changes in the marginal tax rates. As a result, very low productive agents end up choosing very high attention levels.

thereby delivering counterintuitive welfare implications. One should thus be very careful with the welfare implications drawn from the measurement of misperceptions. We believe that this general lesson applies outside of the realm of taxation.

## 3.9 Solution to the stylized model and extensions

### 3.9.1 Solution to the stylized model

The government seeks to maximize tax revenue taking the prior as given. Its problem writes $\max_\tau \tau Y(1-\tilde{\tau})$ such that $\tilde{\tau} = \xi\tau + (1-\xi)\hat{\tau}$, $\{\tau, \hat{\tau}\} \in [0,1]^2$ and $\xi \in (0,1)$. The associated Lagrangian is $\mathscr{L}(\tau, \lambda) = \tau Y(1 - \xi\tau - (1-\xi)\hat{\tau}) + \lambda(\tau - 1)$. Following from the first order Kuhn and Tucker conditions, $\tau = 1$ if and only if $\hat{\tau} \leq 1 - \frac{\xi}{1-\xi}e$ and $\tau = \frac{1-(1-\xi)\hat{\tau}}{\xi(1+e)}$ otherwise. These conditions are also sufficient since the problem is convex under the assumption that $\tau Y(1-\tau)$ is concave.

At the rational equilibrium, the prior is correct $\hat{\tau} = \tau^\star$. Guess that the rational equilibrium is interior. Hence, $\tau^\star = \frac{1}{1+\xi e}$. Because $e > 0$, it implies that $\hat{\tau} > 1 - \frac{\xi}{1-\xi}e$ in equilibrium, thus confirming the guess. It is then straightforward to prove that $\tau^\star Y(1-\tau^\star) < \tau^r Y(1-\tau^r)$ where $\tau^r \equiv \frac{1}{1+e}$ as $\tau^r = \arg\max_{\tau\in[0,1]} \tau Y(1-\tau)$. Moreover, the taxation bias $\tau^\star - \tau^r = \frac{(1-\xi)e}{(1+\xi e)(1+e)}$ is strictly positive for all $\xi \in (0,1)$.

### 3.9.2 Beyond rational equilibria

Most discussions in the paper revolve around the assumption of rational equilibrium – we rely on this assumption because rational equilibria are often seen as restrictive enough to naturally rule out the type of inefficiencies that we highlight in this paper. Nevertheless, it should be clear that rational equilibrium is not a prerequisite for a taxation bias to exist. In particular, it can be shown that the taxation bias arises in setups where agents' prior do not fully adjust in equilibrium and taxpayers can have large misperceptions about what the tax rate is in equilibrium. In the following, we discuss other equilibrium concepts that are found in the literature within the framework of the stylized model to illustrate the main mechanisms behind the emergence of the taxation bias.

Recall that the perceived rate for the representative taxpayer is $\tilde{\tau}$. We depart from the assumption that the perceived rate is a linear combinaison of the true tax rate and a prior (equation (3.2)), and define two new quantities: (i) the partial marginal change in the perceived rate when agents' prior belief is fixed (denoted $\frac{\partial\tilde{\tau}}{\partial\tau}|_{\hat{\tau}}$), and (ii) the marginal change in the perceived rate ($\frac{\partial \tilde{\tau}}{\partial \tau}$). Following the discussion in Section 3.2, a discretionary government does not commit to a predefine policy rule and, therefore, cannot change the representative taxpayer's prior belief. Consequently, its optimal policy solves (interior solution)

$$Y(1-\tilde{\tau}) + \tau \left.\frac{\partial \tilde{\tau}}{\partial \tau}\right|_{\hat{\tau}} \frac{\partial Y}{\partial \tilde{\tau}} = 0 \tag{3.35}$$

On the other hand, under commitment the government may credibly disclose its policy target and affect the representative taxpayer's prior belief. Hence, the optimal policy under commitment solves (interior solution)

$$Y(1-\tilde{\tau}) + \tau \frac{\partial \tilde{\tau}}{\partial \tau} \frac{\partial Y}{\partial \tilde{\tau}} = 0 \tag{3.36}$$

Comparing equations (3.35) and (3.36), it is clear that the force driving the taxation bias arises from a potential difference between the (i) the partial marginal change in the perceived rate when agents' prior belief is fixed ($\frac{\partial \tilde{\tau}}{\partial \tau}|_{\hat{\tau}}$) and (ii) the marginal change in the perceived rate ($\frac{\partial \tilde{\tau}}{\partial \tau}$). These terms will differ as long as prior beliefs adjust (in some way) to the actual tax policy, thus resulting in a taxation bias.

In the paper, we highlight two mechanisms – documented in the literature (see section 3.3.2 for a discussion) – implying that these parameters are likely to differ. First, an anchoring effect: inattentive taxpayers anchor their perceptions on their prior beliefs. This anchoring effect may generate a taxation bias if there exists an indirect (e.g. equilibrium) adjustment in prior beliefs following a tax reform, albeit imperfect. Second, a debiasing effect: taxpayers may adjust their attention, albeit without an adjustment in the anchor. In the following, we illustrate how these mechanisms may generate a taxation bias in setups that have been studied in the literature to explain persistent misperceptions in taxes. These examples demonstrate that the assumption of rational equilibrium is not a necessary condition for a taxation bias to hold.

**Anchoring effect with a partial equilibrium adjustment**

To illustrate that the anchoring effect is not the consequence of the rational equilibrium assumption, we consider an example with an imperfect adjustment of the prior in equilibrium. Such situations are well-suited to generate persistent over- or under-estimation of the actual tax rate in equilibrium.

Instead of assuming that agents' prior fully adjusts to the actual tax rate in equilibrium, we can extend the analyse to any equilibrium condition such that

$$\hat{\tau} = \max(\min(\alpha\tau, 1), 0) \tag{3.37}$$

This alternative equilibrium condition allows for an under-adjustement of the prior (when $0 < \alpha < 1$), full-adjustment ($\alpha = 1$), over-adjustement ( $\alpha > 1$) and opposite-adjustment ($\alpha < 0$).

**Lemma 8.** *(Partial adjustment of the prior) Let the perceived tax rate be given by equation (3.2) and the equilibrium condition by (3.37). Then, when the tax policies under discretion and commitment are interior, the taxation bias is given by*

$$\tau^{eq} - \tau^{\star} \quad = \quad \frac{1}{\xi(e+1) + (1-\xi)\alpha} - \frac{1}{[\xi + \alpha(1-\xi)](1+e)} \tag{3.38}$$

Lemma 8 indicates that the condition for a taxation bias not to exist is quite restrictive as it requires that agents' prior should be unrelated to the actual tax policy (i.e. $\alpha = 0$). Otherwise, we should observe a taxation bias. In particular, this taxation bias is positive whenever the equilibrium condition requires the prior to evolve in the same direction as the actual tax policy ($\alpha > 0$).

While simplistic and presented for illustration purposes, the equilibrium condition (3.37) nevertheless encompasses familiar instances where one should not observe a full adjustment in taxpayers' prior. In particular, it coincides with the use of linear heuristics such as ironing – i.e. when taxpayers anchor their perception on their average tax rate.[94]

[94] Rees-Jones and Taubinsky (2020) 'find that a simple model including only ironers and correct fore-

**Debiasing effect: endogenous attention and salience**

There may be instances where taxpayers anchor their perception on a constant value that is independent of the fiscal policy. Examples of such behavioral biases include models of salience (Chetty et al., 2009; Finkelstein, 2009) and sparsity (Gabaix, 2014).

These models may also result in a taxation bias when they account for an ex-ante endogenous allocation of attention. Intuitively, the endogenous adjustment in attention generates a discrepancy between the key parameters $\frac{\partial \tilde{\tau}}{\partial \tau}|_{\hat{\tau}}$ and $\frac{\partial \tilde{\tau}}{\partial \tau}$. Obviously, this endogenous allocation of attention implicitly requires an adjustment of the prior (albeit partial) – otherwise we shouldn't observe a correlation between the tax size and attention in models with an ex-ante choice of attention. This requirement is in line with Chetty et al. (2009) finding that consumers have (relatively) accurate prior beliefs about the tax rate, but failed to fully account for the tax when making purchasing decisions.

### 3.9.3 Reformulation of the Tractable Gaussian Learning

The indirect utility of a taxpayer is given by equation (3.12). Performing a second order Taylor approximation of the latter around $\tau_0$ gives

$$V^2_{\tau_0}(\tilde{\tau}, \tau_0, R_0; w) = V(\tau_0, \tau_0, R_0; w) + (\tilde{\tau} - \tau_0)\frac{\partial V}{\partial \tilde{\tau}}\Big|_{\tilde{\tau}=\tau_0} + \frac{(\tilde{\tau} - \tau_0)^2}{2}\frac{\partial^2 V}{\partial \tilde{\tau}^2}\Big|_{\tilde{\tau}=\tau_0} \tag{3.39}$$

where $\frac{\partial V}{\partial \tilde{\tau}}|_{\tilde{\tau}=\tau_0} = 0$ and $\frac{\partial^2 V}{\partial \tilde{\tau}^2}\Big|_{\tilde{\tau}=\tau_0} = \frac{\partial^2 y^\star}{\partial \tilde{\tau}^2}$ from (3.7). Hence,

$$\iint V^2_\tau(\tilde{\tau}, \tau, R; w)\phi(s; \tau, \sigma)\phi(\tau; \hat{\tau}, \hat{\sigma})ds d\tau = \int \left[V(\tau, \tau, R; w) + \frac{\tilde{\sigma}^2}{2}\frac{\partial^2 y^\star}{\partial \tilde{\tau}^2}\Big|_{\tilde{\tau}=\tau}\right]\phi(\tau; \hat{\tau}, \hat{\sigma})d\tau$$

where $\tilde{\sigma}^2$ is the posterior variance and we are using the fact that with a Gaussian prior and a Gaussian signal, the posterior is also Gaussian. Accordingly, the expected information

casters accurately predicts average underestimation of marginal tax rates.' When the government redistributes a constant share of its revenue from taxation in the form of a demogrant, their model is equivalent to imposing conditions (3.2) – where $\xi$ is interpreted as the share of correct forecasters – and (3.37) – where $\alpha = 1/(1+a)$ with $a$ the share of tax revenue that are redistributed in lump-sum transfers. Note that when $a = 0$, the marginal tax rate equals the average tax rate and we are back to the rational equilibrium solution presented in Appendix 3.9.1.

reduction writes

$$\mathcal{I}(\sigma) = \frac{1}{2}\Big(\log(2\pi e\hat{\sigma}^2) - \log(2\pi e\tilde{\sigma}^2)\Big) = \frac{1}{2}\log\frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \tag{3.40}$$

where $\frac{1}{2}\log(2\pi e\sigma^2)$ is the differential entropy (in bits) of a Gaussian distribution with variance $\sigma^2$. Therefore, in a Gaussian model, problem (3.11) becomes

$$\max_{\tilde{\sigma}\geq\hat{\sigma}} \quad \tilde{\sigma}^2 \int \frac{\partial^2 y^\star}{\partial\tilde{\tau}^2}\Big|_{\tilde{\tau}=\tau} \phi(\tau;\hat{\tau},\hat{\sigma})d\tau - \kappa\log\frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \tag{3.41}$$

This problem has been extensively studied in the literature. For instance, a step-by-step derivation of the solution is provided in Mackowiak et al. (2018). It shows that the perceived tax rate is $\tilde{\tau} = \xi s + (1-\xi)\hat{\tau}$ where $\xi \in [0,1]$ is a measure of the attention level set optimally to

$$\xi = \max\left(0, 1 + \frac{\kappa}{\hat{\sigma}^2 \int \frac{\partial^2 y^\star}{\partial\tilde{\tau}^2}\Big|_{\tilde{\tau}=\tau}\phi(\tau;\hat{\tau},\hat{\sigma})d\tau}\right) \tag{3.42}$$

## 3.10 Proofs of Proposition 7 and 8

We here prove both propositions at the same time since the only difference between the two problems is in the nature of responses to tax changes that are taken into account. We thus solve the general problem where all agents' responses are taken into account (including equilibrium adjustments) to obtain Proposition 8 and from which Proposition 7 naturally follows.

The Lagrangian associated to problem (3.17) writes

$$\begin{aligned}\mathscr{L}(\tau_g, R, p) \quad = \quad & E_\vartheta\Bigg[\iint \Big[G\Big(\mathcal{V}(\tilde{\tau}, \tau_g+\vartheta, R, \kappa; w)\Big) \\ & +p\Big((\tau_g+\vartheta)y^\star(\tilde{\tau}; w) - R_0 - E\Big)\Big] f_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta; w)f_w(w)d\tilde{\tau}dw\Bigg]\end{aligned} \tag{3.43}$$

The first-order condition associated with the choice of the marginal tax rate $\tau_g$ is

$$\begin{aligned}\frac{1}{p}\frac{d\mathscr{L}}{d\tau_g} \quad = \quad & E_\vartheta\bigg[\int\bigg\{\int\Big[\frac{G'(\mathcal{V})}{p}\frac{d\mathcal{V}}{d\tau_g}+y^\star\Big]f_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)d\tilde{\tau} \\ & +\int\Big[\frac{G(\mathcal{V})}{p}+(\tau_g+\vartheta)y^\star-R_0-E\Big]\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}d\tilde{\tau}\bigg\}f_w(w)dw\bigg]\end{aligned} \tag{3.44}$$

where $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}$ is the change in the posterior distribution of perceived tax rate for type $w$ and captures agents' responses to tax changes.

By definition $\int f_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)d\tilde{\tau}=1$ thus $\int\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}d\tilde{\tau}=0$. Moreover, the quasi-linearity of utility implies that $\frac{d\mathcal{V}}{d\tau_g}=-y^\star(\tilde{\tau};w)$. Therefore, the optimality condition $\frac{1}{p}\frac{d\mathscr{L}}{d\tau_g}=0$ writes

$$\begin{aligned}E_\vartheta\bigg[\int\bigg\{\int\Big[-\frac{G'(\mathcal{V})}{p}y^\star+y^\star\Big]f_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)d\tilde{\tau} \\ +\int\Big[\frac{G(\mathcal{V})}{p}+(\tau_g+\vartheta)y^\star\Big]\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}d\tilde{\tau}\bigg\}f_w(w)dw\bigg]=0\end{aligned} \tag{3.45}$$

This is equation (3.21) from Proposition 7 and characterizes the commitment tax rate. Equation (3.19) from Proposition 8 which characterizes the tax rate chosen by a discretionary government is obtained by when agents' responses to a change in the tax rate is computed holding agents' prior $\hat{q}$ constant. That is $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}\Big|_{\hat{q}(.)}$ replaces $\frac{df_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)}{d\tau_g}$ in equation (3.45).

The first-order condition associated with the choice of the demogrant $R$ is

$$\frac{1}{p}\frac{d\mathscr{L}}{dR}=E_\vartheta\bigg[\iint\Big[\frac{G'(\mathcal{V})}{p}\frac{d\mathcal{V}}{dR}-1\Big]f_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)f_w(w)d\tilde{\tau}dw\bigg] \tag{3.46}$$

By quasi-linearity we have $\frac{d\mathcal{V}}{dR}=1$. The optimality condition $\frac{1}{p}\frac{d\mathscr{L}}{dR}=0$ thus writes

$$E_\vartheta\bigg[\iint\Big[\frac{G'(\mathcal{V})}{p}-1\Big]f_{\tilde{\tau}}(\tilde{\tau}|\tau_g+\vartheta;w)f_w(w)d\tilde{\tau}dw\bigg]=0 \tag{3.47}$$

This is equation (3.20) from Proposition 7 and equation (3.22) from Proposition 8.

## 3.11 Optimal policies in tractable Gaussian case

Conditions (3.45) and (3.47) apply to any learning leading to a differentiable posterior distribution of perceptions $f_{\tilde{\tau}}(\tau|\tau_0;w)$ with positive support on $[0,1]$, where $\tau_0 = \tau_g + \vartheta$. Further insights may be gained by using a tractable Gaussian learning (Assumption 4). Indeed, in this case $f_{\tilde{\tau}}(\tau|\tau_0;w)$ is a Gaussian pdf $\phi(\tau;\mu,\sigma^2)$ with mean $\mu = \xi\tau_0 + (1-\xi)\hat{\tau}$ and variance $\sigma^2 = \sigma^{\star 2}$ . We can thus express agents' responses to tax reforms in terms of changes in the true tax rate $\tau_0$, changes in the prior mean $\hat{\tau}$ and induced changes in attention $\xi$ that correspond to changes in the precision of the signal $\sigma^\star$. To do so, we use a first-order approximation of the objective at the mean $\mu$ and exploit the following Lemma.

**Lemma 9.** *Let $\psi(x)$ be a differentiable real-valued function, $\psi_a(x) = \psi(a) + (x-a)\psi'(a)$ its first-order Taylor approximation evaluated at $a$ and $\phi(x;\mu,\sigma^2)$ the pdf of the Gaussian distribution with mean $\mu$ and variance $\sigma^2$. Then,*

$$\int_{\mathbb{R}} \psi_\mu(x)\phi(x;\mu,\sigma^2)dx = \psi(\mu) \tag{3.48}$$

$$\int_{\mathbb{R}} \psi_\mu(x)\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\mu}dx = \psi'(\mu) \tag{3.49}$$

$$\int_{\mathbb{R}} \psi_\mu(x)\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\sigma}dx = 0 \tag{3.50}$$

*Proof.* Equation (3.48) directly follows from $\int_{\mathbb{R}}(x-\mu)\phi(x;\mu,\sigma^2) = 0$ by definition of the mean. To prove equation (3.49), realize that $\int_{\mathbb{R}} \frac{\partial\phi(x;\mu,\sigma^2)}{\partial\mu}dx = 0$ and $\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\mu} = \frac{x-\mu}{\sigma^2}\phi(x;\mu,\sigma^2)$ so that $\int_{\mathbb{R}}(\psi(\mu) + (x-\mu)\psi'(\mu))\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\mu}dx = \frac{\psi'(\mu)}{\sigma^2}\int_{\mathbb{R}}(x-\mu)^2\phi(x;\mu,\sigma^2) = \psi'(\mu)$. Equation (3.50) follows from the fact that $\int_{\mathbb{R}} \frac{\partial\phi(x;\mu,\sigma^2)}{\partial\sigma}dx = 0$ such that the integral of a constant is nil and that $\frac{\partial\phi(x;\mu,\sigma^2)}{\partial\sigma}$ is symmetric such that the integral of $x$ also nil by a symmetry argument. □

Rewriting equation (3.45) as

$$E_\vartheta\Bigg[\int\bigg\{\int\Big[-\frac{G'(\mathcal{V})}{p}y^\star + y^\star\Big]\phi(\tau;\mu,\sigma^2)d\tau + \int\left[\frac{G(\mathcal{V})}{p} + \tau_0 y^\star\right]\left(\frac{d\phi(\tau;\mu,\sigma^2)}{d\mu}\frac{d\mu}{d\tau_g} + \frac{d\phi(\tau;\mu,\sigma^2)}{d\sigma}\frac{d\sigma}{d\tau_g}\right)d\tau\bigg\}f_w(w)dw\Bigg] = 0 \tag{3.51}$$

allows us to apply Lemma 9 to obtain with $\mu = \xi\tau_0 + (1-\xi)\hat{\tau}$

$$E_\vartheta\Bigg[\int\Bigg\{\Big[-\frac{G'(\mathcal{V})}{p}y^\star + y^\star\Big]\Big|_{\tilde{\tau}=\mu} + \Big[\Big(\frac{G'(\mathcal{V})}{p}(\tilde{\tau}-\tau_0)+\tau_0\Big)\frac{dy^\star}{d\tilde{\tau}}\frac{d\mu}{d\tau_g}\Big]\Big|_{\tilde{\tau}=\mu}\Bigg\}f_w(w)dw\Bigg] = 0 \tag{3.52}$$

since taking $\psi(\tau) = \left[\frac{G(\mathcal{V})}{p} + \tau_0 y^\star\right](\tau)$ implies $\psi'(\mu) = \left[\left(\frac{G'(\mathcal{V})}{p}(\tilde{\tau}-\tau_0)+\tau_0\right)\frac{dy^\star}{d\tilde{\tau}}\right](\mu)$ by the modified envelope condition. Recall that $\mu = \xi\tau_0 + (1-\xi)\hat{\tau}$. Now, in equilibrium we have by definition that $\hat{\tau} = \tau_g + b$ meaning $\mu = \tau_g + \xi\vartheta + (1-\xi)b$ and $\mu - \tau_0 = (1-\xi)(b-\vartheta)$. Hence, in equilibrium,

$$E_\vartheta\Bigg[\int\Bigg\{\Big[-\frac{G'(\mathcal{V})}{p}y^\star + y^\star\Big]\Big|_{\tilde{\tau}=\tau_g+\xi\vartheta+(1-\xi)b} + \Big[\Big(\frac{G'(\mathcal{V})}{p}(1-\xi)(b-\vartheta)+\tau_g+\vartheta\Big)\frac{dy^\star}{d\tilde{\tau}}\frac{d\mu}{d\tau_g}\Big]\Big|_{\tilde{\tau}=\tau_g+\xi\vartheta+(1-\xi)b}\Bigg\}f_w(w)dw\Bigg] = 0 \tag{3.53}$$

Last, we characterize taxpayers' average response to tax reforms $\frac{d\mu}{d\tau_g}$ as computed under discretion and commitment. Under discretion, the policymaker takes agent's priors and thus attention strategies as given, hence $\frac{d\mu}{d\tau_g} = \xi\frac{d\tau_0}{d\tau_g} = \xi$ which yields equation (3.24). Under commitment, the policymaker internalizes the equilibrium condition that priors and thus attention strategies adjust to the tax policy such that $\frac{d\mu}{d\tau_g} = \xi\frac{d\tau_0}{d\tau_g} + (1-\xi)\frac{d\hat{\tau}}{d\tau_g} + \frac{d\xi}{d\tau_g}(\tau_0-\hat{\tau}) = 1 + \frac{d\xi}{d\tau_g}(\vartheta - b)$ in equilibrium. This yields equation (3.25).

Transversality conditions follow from a direct application of Lemma 9 to equation 3.47 with again $\mu = \tau_g + \xi\vartheta + (1-\xi)b$:

$$E_\vartheta\Bigg[\int\frac{G'(\mathcal{V})}{p}\Big|_{\tilde{\tau}=\mu}f(w)dw\Bigg] = 1 \tag{3.54}$$

### 3.11.1 Sufficient statistics formulas in tractable Gaussian case

Taking a small noise approximation, characterizations of equilibrium tax rates under discretion $\tau_g^{\text{eq}}$ and commitment $\tau_g^\star$ in this tractable Gaussian model write

$$\int \Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star + \Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b+\tau_g^{\text{eq}}\Big)\frac{dy^\star}{d\tilde{\tau}}\xi\Big]\Big|_{\tilde{\tau}=\tau_g^{\text{eq}}+(1-\xi)b} f_w(w)dw = 0$$
$$\int \Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star + \Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b+\tau_g^{\star}\Big)\frac{dy^\star}{d\tilde{\tau}}\Big(1-\frac{d\xi}{d\tau_g}b\Big)\Big]\Big|_{\tilde{\tau}=\tau_g^{\star}+(1-\xi)b} f_w(w)dw = 0$$

Assuming preferences are iso-elastic, $U(c,y;w)=c-\frac{(y/w)^{1+\varepsilon}}{1+\varepsilon}$, the elasticity of earnings with respect to the perceived marginal net-of-tax rate $e$ is constant

$$\forall \tilde{\tau}, w, \quad e \equiv \frac{1-\tilde{\tau}}{y^\star}\frac{dy^\star}{d(1-\tilde{\tau})} = \frac{1}{\varepsilon} \iff \frac{dy^\star}{d\tilde{\tau}} = -e\frac{y^\star}{1-\tilde{\tau}} \tag{3.55}$$

Plugging in $e$ we get

$$\int \Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star - \Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b+\tau_g^{\text{eq}}\Big)e\frac{y^\star}{1-\tilde{\tau}}\xi\Big]\Big|_{\tilde{\tau}=\tau_g^{\text{eq}}+(1-\xi)b} f_w(w)dw = 0$$
$$\int \Big[\Big(1-\frac{G'(\mathcal{V})}{p}\Big)y^\star - \Big(\frac{G'(\mathcal{V})}{p}(1-\xi)b+\tau_g^{\star}\Big)e\frac{y^\star}{1-\tilde{\tau}}\Big(1-\frac{d\xi}{d\tau_g}b\Big)\Big]\Big|_{\tilde{\tau}=\tau_g^{\star}+(1-\xi)b} f_w(w)dw = 0$$

To further simplify these formulas we now make a small perception bias approximation $b << 1$. This allows us to use the approximation $\frac{1}{1-\tau_g-(1-\xi)b} \approx \frac{1}{1-\tau_g}$ and to assume $\frac{d\xi}{d\tau_g}b << 1$ to simplify some terms[95]. Defining social marginal welfare weights $g(w) \equiv \frac{G'(\mathcal{V})}{p}$ and the mean operator $\bar{x} = \int x(w)f(w)dw$ we get

$$\Big\{\overline{(1-g)y^\star} - \frac{\tau_g^{\text{eq}}}{1-\tau_g^{\text{eq}}}\overline{y^\star\xi}e - \frac{b}{1-\tau_g^{\text{eq}}}\overline{g(1-\xi)y^\star\xi e}\Big\}\Big|_{\tilde{\tau}=\tau_g^{\text{eq}}+(1-\xi)b} = 0$$
$$\Big\{\overline{(1-g)y^\star} - \frac{\tau_g^{\star}}{1-\tau_g^{\star}}\overline{y^\star}e - \frac{b}{1-\tau_g^{\star}}\overline{g(1-\xi)y^\star e}\Big\}\Big|_{\tilde{\tau}=\tau_g^{\star}+(1-\xi)b} = 0$$

which simplify to the compact sufficient statistics formulas

[95] In our simulations we do check that $\frac{d\xi}{d\tau_g}$ does not take large values (it takes values between 0.2 and 1 in equilibrium) as a way to confirm the validity of this approximation.

$$\tau_g^{\text{eq}} = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - b\ \frac{\overline{g(1-\xi)y^\star\xi}\ e}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} \tag{3.56}$$

$$\tau_g^\star = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} - b\ \frac{\overline{g(1-\xi)y^\star}\ e}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} \tag{3.57}$$

where all endogenous quantities on the right hand-side of the equations are evaluated at respectively $\tilde{\tau} = \tau_g^{\text{eq}} + (1-\xi)b$ and $\tilde{\tau} = \tau_g^\star + (1-\xi)b$. In other words formulas are expressed in terms of sufficient statistics evaluated at the optimum.

## 3.12 Taxation bias in tractable Gaussian case

A difficulty in comparing $\tau_g^{\text{eq}}$ and $\tau_g^\star$ is that some right-hand side quantities are endogenous to the tax rate and thus evaluated at different tax rates. To overcome this difficulty, we use a small taxation bias approximation $\tau_g^{\text{eq}} \approx \tau_g^\star = t$ such that quantities can be evaluated to a first-order approximation at the same tax rate. Furthermore, we assume that corrective motives associated to the presence of a perception bias $b$ are evaluated at tax rate $t$ such that we can finally directly compare

$$\tau_g^{\text{eq}} = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - b\ \frac{\overline{g(1-\xi)y^\star\xi}\ e}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} \tag{3.58}$$

$$\tau_g^\star = \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} - b\ \frac{\overline{g(1-\xi)y^\star}\ e}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} \tag{3.59}$$

The first terms on the right-hand side corresponds to the standard optimal tax formula (e.g. Piketty and Saez (2013)) whereas the second are corrective terms associated to the existence of a perception bias $b$. For small perception biases, these corrective terms are second-order and go in the same direction for both positive and normative tax rates. They are thus not driving the difference between the two and we disregard them to derive the following simple sufficient statistics formula for the taxation bias

$$\begin{aligned}
\tau_g^{\text{eq}} - \tau_g^\star &= \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e} - \frac{\overline{(1-g)y^\star}}{\overline{(1-g)y^\star} + \overline{y^\star}\ e} && (3.60)\\
&= \frac{\overline{(1-g)y^\star}\left(\overline{(1-g)y^\star} + \overline{y^\star}\ e\right) - \overline{(1-g)y^\star}\left(\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e\right)}{\left(\overline{(1-g)y^\star} + \overline{y^\star\xi}\ e\right)\left(\overline{(1-g)y^\star} + \overline{y^\star}\ e\right)} && (3.61)\\
&= \frac{e\ \tau_g^{\text{eq}}\ \tau_g^\star}{\overline{(1-g)y^\star}}\left(\overline{y^\star} - \overline{y^\star\xi}\right) && (3.62)\\
&\simeq \frac{\overline{(1-\xi)y^\star}}{\overline{(1-g)y^\star}}\ e\ t^2 && (3.63)
\end{aligned}$$

## 3.13 Utility decomposition

Let $\mathcal{V}_\kappa(\tau_g^\star(\kappa), w)$ be the expected utility of taxpayer $w$ at the positive equilibrium when the information cost is $\kappa$ and the optimal target tax rate of the government is $\tau_g^\star(\kappa)$. Then, with a separable utility,

$$\begin{aligned}
\mathcal{V}_0(\tau_g^\star(0), w) &= E_{\tau_0|\tau_g^\star(0)}\Big[R_0 + (1-\tau_0)y^\star(\tau_0; w) - v(y^\star(\tau_0; w); w)\Big]\\
\mathcal{V}_\kappa(\tau_g^\star(\kappa), w) &= E_{\tau_0|\tau_g^\star(\kappa)}\Big[\int \Big(R_0 + (1-\tau_0)y^\star(\tau; w) - v(y^\star(\tau; w); w)\Big) f_{\tilde{\tau}}(\tau|\tau_0, w) d\tau\Big] - \kappa\mathcal{I}(\sigma^\star(\hat{q}(\tau), \kappa, w))
\end{aligned}$$

Using straightforward algebra,

$$\mathcal{V}_\kappa(\tau_g^\star(\kappa), w) - \mathcal{V}_0(\tau_g^\star(0), w) = \Delta^{\mathcal{V}} R + \Delta^{\mathcal{V}} \tau + \Delta^{\mathcal{V}} b + \Delta^{\mathcal{V}}\text{uncertainty} + \Delta^{\mathcal{V}}\text{info cost} \quad (3.64)$$

where

$$\Delta^{\mathcal{V}} R \equiv E_{\tau_0|\tau_g^\star(\kappa)}[R_0] - E_{\tau_0|\tau_g^\star(0)}[R_0]$$

is the change in the average demogrant,

$$\Delta^{\mathcal{V}} \tau \equiv E_{\tau_0|\tau_g^\star(\kappa)}[(1-\tau_0)y^\star(\tau_0; w) - v(y^\star(\tau_0; w); w)] - E_{\tau_0|\tau_g^\star(0)}[(1-\tau_0)y^\star(\tau_0; w) - v(y^\star(\tau_0; w); w)]$$

is the change in the expected utility due to the change in the tax target $\tau_g^\star$,

$$\begin{aligned}\Delta^{\mathcal{V}} b \;\; \equiv \;\; & E_{\tau_0|\tau_g^\star(\kappa)}\Big[(1-\tau_0)\Big(y^\star(\tau_0+(1-\xi)b;w)-y^\star(\tau_0;w)\Big) \\ & -\Big(v(y^\star(\tau_0+(1-\xi)b;w);w)-v(y^\star(\tau_0;w);w)\Big)\Big]\end{aligned}$$

is the change in the expected utility due to the bias $b$,

$$\begin{aligned}\Delta^{\mathcal{V}}\text{uncertainty} \;\; \equiv \;\; & E_{\tau_0|\tau_g^\star(\kappa)}\Big[\int(1-\tau_0)\Big(y^\star(\tau;w)-v(y^\star(\tau;w);w\Big) \\ & \phi\Big(\tau;\xi\tau_0+(1-\xi)(\tau_g^\star(\kappa)+b),(\xi\sigma^\star)^2\Big)d\tau \\ & -\Big((1-\tau_0)\Big(y^\star(\tau_0+(1-\xi)b;w)-v(y^\star(\tau_0+(1-\xi)b;w);w)\Big)\Big]\end{aligned}$$

is the change in the expected utility due to noisy information and $\Delta^{\mathcal{V}}\text{info cost} = -\kappa\mathcal{I}(\sigma^\star(\hat{q}(\tau),\kappa,w))$.

This online appendix first provides detailed information on the numerical simulations. Second, it gives all proofs and derivations for the extension to nonlinear taxation. Third, we show how to incorporate income effects in the analysis.

## 3.14 Numerical simulations

Simulations are implemented using Matlab and the algorithm may be summarized as follows. We first estimate a log-normal distribution of skills that we extend with a Pareto tail. This distribution of skills is then binned into a discrete approximation and taken as given for the rest of the exercise. Second, we find the optimal policy of the government using an iterative routine. Starting with a guess for the optimal policy, we compute the optimal attention strategies and allocations in equilibrium (i.e. when the priors are adjusted). We then compute a new optimal policy given taxpayers' choices and iterate until convergence to a fixed point solution.

This appendix provides details on these different steps. We first present the calibration strategy for the skill distribution. Second, we explain how to solve for the optimal attention strategies and allocations for a given tax schedule. Finally, we discuss how the government's problem is solved in the linear tax setting before turning to the nonlinear case.

### 3.14.1 Skill distribution

Simulations require an exogenous distribution of skills $f_w(.)$. We fit the adjusted gross incomes from the 2016 March CPS data to a log-normal distribution. The parameters of the log-normal are chosen to exactly match the mean and median of the observed distribution. Following Saez (2001), we extend the log-normal distribution with a Pareto tail ($k = 2$) for annual incomes above \$200,000. We then discretize the income distribution using evenly distributed bins over the $[200; 200,000]$ interval and evenly distributed bins (in ln scale) over the $[200,000; 4,000,000]$ interval. This allows us to approximate integrals with Riemann sums.

To translate this income distribution into a skill distribution, we invert agents' first-order conditions for labor supply. We first use OECD data on 2016 labor taxes in the US and fit a linear tax schedule $\{\tau_{obs}, R_{obs}\}$. Then, we impose a quasi-linear utility specification $u(c, y; w) = c - (y/w)^{1+\epsilon}/(1+\epsilon)$ with $e = 1/\epsilon = 0.33$ (Chetty, 2012). Assuming we are in a no bias equilibrium (i.e. rational expectation) such that agents' perceived tax rate coincide with the observed one $\tau_{obs}$, this allows us to compute skills through $w = \left(y^{\epsilon}/(1-\tau_{obs})\right)^{\frac{1}{1+\epsilon}}$. We also use the estimated linear tax system $\{\tau_{obs}, R_{obs}\}$ together with the actual distribution of earnings to deduce an exogenous expenditure requirement $E$ for the government budget constraint.

### 3.14.2 Taxpayers' behavior

Taxpayers' choices are presented in Section 3.3. For the simulations, we consider Gaussian implementation shocks. Under this assumption, the equilibrium prior distribution is Gaussian as well. Consequently, one may easily compute the attention parameter ($\xi$), income ($y$) and consumption ($c$) for each taxpayer. Given an attention cost $\kappa$, a marginal tax rate $\tau$ – that potentially varies for each individual – and an uncertainty parameter $\sigma_{\vartheta}$, the attention strategy in equilibrium follows from equation (3.14). Gaussian integrals are approximated using Gauss-Hermite quadratures. Using an agent's first-order condition (3.7) and budget constraint, we compute her income, consumption and utility for different signal realizations. These computations are made for each type of agent $w$. The demogrant $R$ is computed from the government budget constraint.

### 3.14.3 Optimal linear tax

Unless stated otherwise, we assume throughout our numerical exercise that the social planner has a log objective $G(.) = \log(.)$.

In order to compute the optimal linear tax under discretion, we start with a guess $\tau_{g,0}$. Using this guess, we can deduce each taxpayer's attention strategy when the prior is adjusted to the guess $\hat{\tau}_0 = \tau_{g,0} + b$. We then consider this distribution of attention strategies as constant and use a Matlab optimization routine to find a new $\tau_{g,1}$ which maximizes social welfare for these attention strategies. We then update the prior $\hat{\tau}_1 = \tau_{g,1} + b$, recompute the attention strategies and re-optimize until convergence $|\hat{\tau}_i - \tau_{g,i+1}| \leq 1e^{-5}$. This method is intuitive and captures the essence of the discretionary policy: the government maximizes its objective taking attention strategies as fixed.

We also implement an alternative algorithm where instead of maximizing social welfare numerically we directly pick a new tax rate using the government FOCs in Proposition 7 under a small signals approximation. We find comparable equilibrium rates. Similarly, we compute the optimal policy under commitment using the FOCs in Proposition 8.

### 3.14.4 Optimal nonlinear tax

In order to compute the optimal nonlinear tax, we again use an iterative routine. We start with a guess – namely, a constant marginal rate – and iterate until convergence of the nonlinear tax schedule. We only present results for the unbiased equilibrium $b = 0$. We proceed in the same spirit as for the linear tax schedule:

1. Start with a guess for the nonlinear tax schedule

2. Compute the attention strategies ($\forall w$) for a given adjusted prior $\hat{\tau}_w$

3. Compute allocations given attention strategies and tax schedule

4. Solve for the government FOCs at each $w$ to deduce a new tax schedule

5. Repeat steps 1-4 until convergence.

To maintain the numerical stability of the algorithm we impose a slow adjustment of attention strategies $\xi$ at each iteration. Indeed, marginal tax rates being sensitive to attention, one shall avoid large jumps in the attention parameter. The convergence criteria we use is the infinite norm for both marginal tax rates and attention strategies.

## 3.15 Proofs for the extension to non-linear taxation

We here provide the proofs on the monotonicity condition and Proposition 11 (ABCD formula) of the main text.

### 3.15.1 Monotonicity

In this section, we demonstrate that the monotonicity condition is expected to hold for the quasi-linear and iso-elastic separable utility function that we consider in our simulations. For alternative specifications, we recommend to proceed using a guess-and-verify method. The latter is already implemented in our code and a warning is automatically displayed when the monotonicity does not hold ex post.

With a quasi-linear and iso-elastic separable utility function the first-order condition defining $y^\star(\tilde{\tau}_w; w)$ is

$$(FOC)_y : 1 - \tilde{\tau}_w - \frac{1}{w}\left(\frac{y^\star}{w}\right)^\epsilon = 0 \tag{3.65}$$

Differentiating this equation with respect to $w$ yields

$$\frac{\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon-1}\frac{dy^\star(\tilde{\tau}_w; w)}{dw} = \frac{1+\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^\epsilon - \frac{d\tilde{\tau}_w}{dw} \tag{3.66}$$

Now – in expectation of the realization of the implementation shock $\vartheta$ – we also have $\tilde{\tau}_w = T_g'(y^\star(\hat{\tau}_w; w)) + (1-\xi)b$ which allows us to get

$$\frac{d\tilde{\tau}_w}{dw} = T_g''(y^\star(\hat{\tau}_w; w))\frac{dy^\star(\hat{\tau}_w; w)}{dw} + \frac{d}{dw}\Big[(1-\xi)b\Big] \tag{3.67}$$

and we can show that

1. If agents correctly perceive marginal tax rates ($b = 0$), the equilibrium condition $\hat{\tau}_w = T_g'(y^\star(\hat{\tau}_w; w)) + b$ becomes $\hat{\tau}_w = T_g'(y^\star(\hat{\tau}_w; w)) = \tilde{\tau}$. We then have $\frac{dy^\star(\hat{\tau}_w;w)}{dw} = \frac{dy^\star(\tilde{\tau}_w;w)}{dw}$ such that plugging (3.67) with $b = 0$ into (3.66) the monotonicity condition boils down to

$$\frac{dy^\star}{dw} = \frac{\frac{1+\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon}}{\frac{\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon-1} + T_g''(y^\star)} \geq 0 \iff -T_g''(y^\star) \leq \frac{\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon-1} \tag{3.68}$$

2. If agents exhibit a small perception bias ($b \approx 0$) such that we have $\frac{dy^\star(\hat{\tau}_w;w)}{dw} \approx \frac{dy^\star(\tilde{\tau}_w;w)}{dw}$ plugging (3.67) into (3.66) the monotonicity condition rewrites

$$\frac{dy^\star}{dw} = \frac{\frac{1+\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon} - \frac{d}{dw}\Big[(1-\xi)b\Big]}{\frac{\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon-1} + T_g''(y^\star)} \geq 0 \iff \begin{cases} -T_g''(y^\star) \leq \frac{\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon-1} \\ \frac{d}{dw}\Big[(1-\xi)b\Big] \leq \frac{1+\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon} \end{cases} \tag{3.69}$$

   where the equivalence comes from the fact that the other case in which we would have $-T_g''(y^\star) \geq \frac{\epsilon}{w^2}\left(\frac{y^\star}{w}\right)^{\epsilon-1}$ is infeasible.

Hence, the monotonicity condition will hold if the tax function $T_g(y)$ is sufficiently smooth such that its second derivative is bounded (in absolute value).

### 3.15.2 Proposition 11 (ABCD formula)

We proceed with a tax perturbation approach in order to characterize the nonlinear tax schedule chosen under discretion and under commitment. Consider a tax schedule $T_g(.)$ and a reform that consists in a small increase $\Delta\tau^r$ in marginal tax rates in a small bandwidth of earnings $[y^r - \Delta y^r, y^r]$ and let us compute its impact on the government's objective (written in Lagrangian form)

$$\mathscr{L} = E_\vartheta\Bigg[\iint \Big\{G\Big(\mathcal{V}(\tilde{\tau}_w, T_0(.); \kappa, w)\Big) + p\Big(T_0(y^\star(\tilde{\tau}_w; w)) - E\Big)\Big\}\ f_{\tilde{\tau}_w}(\tau|\tau_0; w)\ f_w(w) d\tau dw\Bigg] \tag{3.70}$$

where $p$ is the multiplier associated to the government's budget constraint and is equal to the social marginal value of public funds at the optimum.

**Impact of the reform** For a given target tax schedule $T_g(.)$, the reform has

- a mechanical effect $dM$ and a welfare effect $dW$ that translate the lump-sum increase of $\Delta\tau^r \Delta y^r$ in the tax liabilities of agents $w \in [w^r, \infty[$ defined by $y^\star(\tilde{\tau}_{w^r}; w^r) \equiv y^r$ where $E_s[\tilde{\tau}_w|\hat{\tau}_w] = \xi T_0'(y^\star(\hat{\tau}_w; w)) + (1-\xi)\hat{\tau}_w$ with $T_0' = T_g' + \vartheta$
- a labor supply or behavioral effect $dB$ that translates an increase $\Delta\tau^r$ in marginal tax rates that impacts the perceived marginal tax rates $\tilde{\tau}_w$ of agents $w \in [\hat{w}^r - \Delta\hat{w}^r, \hat{w}^r]$ defined by $y^\star(\hat{\tau}_{\hat{w}^r}; \hat{w}^r) \equiv y^r$ and $y^\star(\hat{\tau}_{\hat{w}^r}; \hat{w}^r - \Delta\hat{w}^r) \equiv y^r - \Delta y^r$

such that the total impact on the government's objective is

$$\frac{d\mathcal{L}}{p} = \frac{dM}{p} + \frac{dW}{p} + \frac{dB}{p} \tag{3.71}$$

with

$$\begin{aligned}
\frac{dM}{p} + \frac{dW}{p} &= \int_{w^r}^{\infty} E_\vartheta\Bigg[\int \left\{\Delta\tau^r\Delta y^r - \frac{G'(\mathcal{V}(w))}{p}\frac{\partial U}{\partial c}\Delta\tau^r\Delta y^r\right\} f_{\tilde{\tau}}(\tau|\tau_0; w)\ d\tau\Bigg] f_w(w)\ dw \\
&= \int_{w^r}^{\infty} E_\vartheta\Bigg[\int \Big(1 - g(w)\Big)\Delta\tau^r\Delta y^r\ f_{\tilde{\tau}}(\tau|\tau_0; w)\ d\tau\Bigg] f_w(w)\ dw \qquad (3.72)
\end{aligned}$$

since we here have, holding $\tilde{\tau}_w$ constant,

$$d\mathcal{V} = \frac{d}{dc}\Big\{U\Big(\underbrace{y^\star(\tilde{\tau}_w; w) - T_0(y^\star(\tilde{\tau}_w; w))}_{c}, y^\star(\tilde{\tau}_w; w); w\Big) - \kappa\mathcal{I}(\sigma^\star)\Big\}dc = -\frac{\partial U}{\partial c}dT_0 \tag{3.73}$$

and

$$\begin{aligned}
\frac{dB}{p} &= \int_{\hat{w}^r - \Delta\hat{w}^r}^{\hat{w}^r} E_\vartheta\Bigg[\int \left\{\frac{G(\mathcal{V}(w))}{p} + T_0(y^\star(\tilde{\tau}; w))\right\} \frac{df_{\tilde{\tau}}(\tau|\tau_0; w)}{d\tau_g}\Delta\tau^r\ d\tau\Bigg] f_w(w)\ dw \\
&\approx E_\vartheta\Bigg[\int \left\{\frac{G(\mathcal{V}(\hat{w}^r))}{p} + T_0(y^\star(\tilde{\tau}; \hat{w}^r))\right\} \frac{df_{\tilde{\tau}}(\tau|\tau_0; \hat{w}^r)}{d\tau_g}\Delta\tau^r\ d\tau\Bigg] f_w(\hat{w}^r)\ \Delta\hat{w}^r \qquad (3.74)
\end{aligned}$$

since we here have, holding $\tilde{\tau}_w$ constant,

$$d\mathcal{V} = \frac{d}{d\tau_g}\Big\{U\Big(y^\star(\tilde{\tau}_w; w) - T_0(y^\star(\tilde{\tau}_w; w)), y^\star(\tilde{\tau}_w; w); w\Big) - \kappa\mathcal{I}(\sigma^\star)\Big\}d\tau_g = 0 \tag{3.75}$$

**Characterization of tax policy** The optimality condition for the choice of tax policy $\frac{d\mathcal{L}}{p} = 0$ thus writes

$$E_{\vartheta}\left[ \int \left\{ \frac{G(\mathcal{V}(\hat{w}^r))}{p} + T_0(y^{\star}(\tilde{\tau}_{\hat{w}^r}; \hat{w}^r)) \right\} \frac{df_{\tilde{\tau}_{\hat{w}^r}}(\tau|\tau_0; \hat{w}^r)}{d\tau_g} \, d\tau \right] \frac{f_w(\hat{w}^r)}{\frac{dy^{\star}(\hat{\tau}_{\hat{w}^r}; \hat{w}^r)}{dw}}$$
$$+ \int_{w^r}^{\infty} E_{\vartheta}\left[ \int \Big(1 - g(w)\Big) \, f_{\tilde{\tau}_w}(\tau|\tau_0; w) \, d\tau \right] f_w(w) \, dw = 0 \tag{3.76}$$

where we have simplified through by $\Delta\tau^r \Delta y^r$ noting that

$$y^{\star}(\hat{\tau}; \hat{w}^r - \Delta\hat{w}^r) \equiv y^r - \Delta y^r \quad \Longrightarrow \quad \Delta\hat{w}^r \frac{dy^{\star}(\hat{\tau}_{\hat{w}^r}; \hat{w}^r)}{dw} \approx \Delta y^r$$

Assuming we are in the tractable Gaussian case, the ex post (after learning) distribution of the perceived marginal tax rate is Gaussian $f_{\tilde{\tau}_w}(\tau|\tau_0; w) \sim \mathcal{N}(\mu_w, \sigma^2)$ with mean $\mu_w = \xi\tau_0 + (1-\xi)\hat{\tau}_w$ and variance parameter $\sigma = \sigma^{\star}$. Applying Lemma 9 we can thus rewrite the optimality condition as

$$E_{\vartheta}\left[ \left[ \left\{ \frac{G'(\mathcal{V}(\hat{w}^r))}{p} \Big( \tilde{\tau}_{\hat{w}^r} - T_0'(y^{\star}(\tilde{\tau}_{\hat{w}^r}; \hat{w}^r)) \Big) \; + \; T_0'(y^{\star}(\tilde{\tau}_{\hat{w}^r}; \hat{w}^r)) \right\} \frac{dy^{\star}}{d\tilde{\tau}} \frac{d\mu_{\hat{w}^r}}{d\tau_g} \right] \Bigg|_{\tilde{\tau}=\mu_{\hat{w}^r}} \frac{f_w(\hat{w}^r)}{\frac{dy^{\star}(\hat{\tau}_{\hat{w}^r}; \hat{w}^r)}{dw}} \right.$$
$$\left. + \; \int_{w^r}^{\infty} \Big[ 1 - g(w) \Big] \Big|_{\tilde{\tau}=\mu_w} f_w(w) \, dw \right] = 0 \tag{3.77}$$

where $\frac{d\mu_w}{d\tau_g} = \xi \frac{d\tau_0}{d\tau_g} = \xi$ under discretion since the government takes agents' priors as given whereas $\frac{d\mu_w}{d\tau_g} = \xi \frac{d\tau_0}{d\tau_g} + (1-\xi)\frac{d\hat{\tau}_w}{d\tau_g} + \frac{d\xi}{d\tau_g}(\tau_0 - \hat{\tau}_w)$ under commitment since the government internalizes that priors adjust to the choice of tax policy and are thus an endogenous object.

In addition, the Lagrange multiplier is – absent income effects – determined by the same transversality condition as before

$$E_{\vartheta}\left[ \int_0^{\infty} \Big[ 1 - g(w) \Big] \Big|_{\tilde{\tau}=\mu_w} f_w(w) \, dw \right] = 0 \tag{3.78}$$

which can be obtained in a perturbation approach by computing the impact of a uniform lump-sum increase in taxes.

**ABCD formula** To obtain our ABCD formula from equation (3.77), let us introduce

$e = \frac{1-\tilde{\tau}_w}{y^\star(\tilde{\tau}_w;w)} \frac{dy^\star(\tilde{\tau}_w;w)}{d(1-\tilde{\tau}_w)}$ and assume that the shock $\vartheta$ is small to use $E_\vartheta[\psi(\vartheta)] \approx \psi(E_\vartheta[\vartheta])$ regardless of function $\psi$'s curvature such that $E_\vartheta[\psi(\tau_0)] \approx \psi(\tau_g)$. This yields

$$\frac{T_g'(y^\star(\mu_{\hat{w}^r};\hat{w}^r)) + g(\hat{w}^r)|_{\tilde{\tau}=\mu_{\hat{w}^r}}\Big(\mu_{\hat{w}^r} - T_g'(y^\star(\mu_{\hat{w}^r};\hat{w}^r))\Big)}{1-\mu_{\hat{w}^r}} \tag{3.79}$$
$$= \frac{1}{e\frac{d\mu_{\hat{w}^r}}{d\tau_g}|_{\tilde{\tau}=\mu_{\hat{w}^r}}} \frac{1}{y^\star(\mu_{\hat{w}^r};\hat{w}^r)} \frac{\frac{dy^\star(\tilde{\tau}_{\hat{w}^r};\hat{w}^r)}{dw}}{f_w(\hat{w}^r)} \int_{w^r}^\infty \Big(1 - g(w)|_{\tilde{\tau}=\mu_w}\Big) f_w(w)\ dw$$

where $\mu_w = \xi T_g'(y^\star(\hat{\tau}_w;w)) + (1-\xi)\hat{\tau}_w$ and $\frac{d\mu_w}{d\tau_g} = \xi\frac{dT_g'(y^\star(\hat{\tau}_w;w))}{d\tau_g} = \xi$ under discretion since the government takes agents' priors as given whereas $\frac{d\mu_w}{d\tau_g} = \xi + (1-\xi)\frac{d\hat{\tau}_w}{d\tau_g} + \frac{d\xi}{d\tau_g}(T_g'(y^\star(\hat{\tau}_w;w)) - \hat{\tau}_w)$ under commitment since the government internalizes that priors adjust to the policy rule and are thus an endogenous object.

Note that with a quasi-linear and iso-elastic separable utility function we have $y^\star(\tilde{\tau}_w;w) = w^{1+\frac{1}{\epsilon}}(1-\tilde{\tau}_w)^{\frac{1}{\epsilon}}$ and $e = \frac{1}{\epsilon}$ such that

$$\frac{dy^\star(\tilde{\tau}_w;w)}{dw} = \left(1+\frac{1}{\epsilon}\right) w^{\frac{1}{\epsilon}}(1-\tilde{\tau}_w)^{\frac{1}{\epsilon}} = \frac{1+e}{w} y^\star(\tilde{\tau}_w;w) \tag{3.80}$$

Assuming small (or no) perception biases such that $\tilde{\tau}_w \approx \hat{\tau}_w$ and $\frac{dy^\star(\tilde{\tau}_w;w)}{dw} \approx \frac{dy^\star(\hat{\tau}_w;w)}{dw}$ yields

$$\frac{T_g'(y^\star(\mu^{\text{eq}}_{\hat{w}^r};\hat{w}^r)) + g(\hat{w}^r)(1-\xi)b}{1-\tilde{\tau}^{\text{eq}}_{\hat{w}^r}} = \frac{1}{\frac{d\mu_{\hat{w}^r}}{d\tau_g}} \frac{1+e}{e} \frac{1}{\hat{w}^r f_w(\hat{w}^r)} \int_{w^r}^\infty \Big(1-g(w)\Big) f_w(w)\ dw \tag{3.81}$$

## 3.16 Income effects

In this final section of the Online Appendix, we illustrate how the (linear tax) model in the paper could be extended to account for income effects and accordingly characterize tax policy under discretion and commitment. We now have to account for the fact that the average posterior tax rate is no longer a sufficient statistics for taxpayers' earnings choices. This requires a mere reformulation of the initial problem without income effects: integration in the government's problem is now with respect to the signal distribution.

In order to introduce income effects, it will prove useful to slightly reformulate taxpayers' problem introduced in Section 3.3. To this end, consider that there is a continuum of individuals at each skill $w$ of size $f(w)$ and let $Y(\tau_0) \equiv \iint y^\star(\centerdot)\phi(s;\tau_0,\sigma^\star)dsdF(w)$ be the aggregate earnings. Then, because the government budget constraint is binding at the optimum, the demogrant writes $R(\tau_0) = \tau_0 Y(\tau_0) - E$ as the overall population remains of size one. Further, and given that a taxpayer's budget constraint binds ex post, consumption adjusts such that $c_0 = R(\tau_0) + (1-\tau_0)y$. Therefore, an agent's utility is $u(R(\tau_0) + (1-\tau_0)y, y)$ for a realization $\tau_0$ and earnings choice $y$.

Given the above reformulation, the only uncertainty arises from the randomness in the realized tax rate. An individual therefore chooses the signal precision $\sigma$ and income $y$ to maximize her expected utility

$$\sup_{\sigma,y|s} \iint u(R(\tau) + (1-\tau)y, y; w)\phi(s;\tau,\sigma)\hat{q}(\tau)dsd\tau - \kappa\mathcal{I}(\sigma) \tag{3.82}$$

where admissible earnings policies for this individual's choice may depend on the signal $s$. Now, guess that the optimal attention strategy $\sigma^\star$ depends only on $w$, $\hat{q}(.)$, and $\kappa$. As a consequence, the optimal earnings choice $y^\star(s,w;\sigma^\star,\hat{q}(.))$ now solves

$$\begin{aligned}\int [(1-\tau)u_c(R(\tau) + (1-\tau)y^\star, y^\star; w) \\ +u_y(R(\tau) + (1-\tau)y^\star, y^\star; w)] f\Big(\tau|s;\sigma^\star,\hat{q}(.)\Big)d\tau = 0\end{aligned} \tag{3.83}$$

where $f(\tau|s;\sigma^\star,\hat{q}(.)) = \frac{\phi(s;\tau,\sigma^\star)\hat{q}(\tau)}{\int \phi(s;\tau,\sigma^\star)\hat{q}(\tau)d\tau}$ from Bayes rule. Assume that a solution to equation (3.83) exists. In turn, it implies that

$$\sigma^\star(w,\hat{q}(.),\kappa) \quad = \quad \arg\sup_{\sigma} \iint u(R(\tau) + (1-\tau)y^\star, y^\star; w)\phi(s;\tau,\sigma)\hat{q}(\tau)dsd\tau - \kappa\mathcal{I}(\sigma) \tag{3.84}$$

thus confirming the guess on $\sigma^\star$ (when it exists). We can now define agents' indirect utility function

$$\mathcal{V}(s,\tau_0;w,\kappa,\hat{q}(.)) \equiv u(R(\tau_0) + (1-\tau_0)y^\star, y^\star; w) - \kappa\mathcal{I}(\sigma^\star) \tag{3.85}$$

Turning to the government problem, it requires a mere variation from (3.17)

$$\max_{\tau_g} \qquad E_\vartheta\Bigg[\iint G\Big(\mathcal{V}\Big(s,\tau_0;w,\kappa,\hat{q}(.)\Big)\Big)\ \phi(s;\tau_0,\sigma^\star)\ f_w(w)d\tau dw\Bigg] \tag{3.86}$$

Note that the inner integration is now with respect to the signal distribution $\phi(s;\tau_0,\sigma^\star)$ and no longer with respect to the posterior distribution of perceived rates. This is because the perceived tax rate $\tilde{\tau}$ is no longer a sufficient statistics for earnings choices.

The first order condition for the target tax rate under discretion writes

$$E_\vartheta\Big[\int\Big\{\int\Big[G'(\mathcal{V})\frac{d\mathcal{V}}{d\tau_g}\phi(s;\tau_0,\sigma^\star)ds+\int G(\mathcal{V})\frac{d\phi(s;\tau_0,\sigma^\star)}{d\tau_g}ds\Big\}f_w(w)dw\Big]=0 \quad (3.87)$$

and the first order condition for the target tax rate under commitment writes

$$\begin{aligned}E_\vartheta\Big[\int\Big\{\int\Big[G'(\mathcal{V})\Big(\frac{d\mathcal{V}}{d\tau_g}+\frac{d\mathcal{V}}{d\hat{q}(.)}\frac{d\hat{q}(.)}{d\tau_g}\Big)\phi(s;\tau_0,\sigma^\star)ds \quad (3.88)\\ +\int G(\mathcal{V})\frac{d\phi(s;\tau_0,\sigma^\star)}{d\tau_g}ds\Big\}f_w(w)dw\Big]=0\end{aligned}$$

This characterizes tax policy under discretion and commitment in the presence of income effects. The key difference between the two equations is the fact that the commitment tax policy takes into account the adjustment in the prior $\frac{d\hat{q}(.)}{d\tau_g}$ whereas the discretion tax policy does not. This leads to a taxation bias.

# Chapter 4

# Forecast Stickiness along the Business Cycle

**Abstract:** This paper estimates a two margin forecast formation process that allows for forecast rounding on individual and consensus forecast data. Forecasters decide when to revise their forecast (extensive margin). When they do, they slowly incorporate new information (intensive margin) and may report a rounded value for their new forecast (rounding). It finds that these three rigidities simultaneously exist and estimate their respective contribution. The overall forecast stickiness is almost exclusively the consequence of the rigidities at the intensive margin. It then derives quarterly time series for the evolution of information frictions and proposes a simple mapping to account for these variations in economic models.[96]

Keywords: forecasts; information frictions.

JEL Classification: D83, D84, E30, E70

[96] I am extremely grateful to Philippe Andrade, Antoine Ferey, Erwan Gautier and Xavier Ragot for helpful comments and suggestions.

## 4.1 Introduction

The stickiness in economic agents' expectations, which arises as a consequence of information frictions, drives the persistence of aggregate variables and the transmission of monetary and fiscal policies (e.g. Sims (2003); Mankiw and Reis (2006); Maćkowiak and Wiederholt (2015); Fuhrer (2017); Carroll et al. (2020)). It is therefore a central topic in macroeconomics. However, and as is further discussed below in the literature review, economists have yet to concur on (i) the source(s) of these frictions, (ii) whether they are a characteristics of individual data, (iii) their evolution over time at high frequencies, and (iv) their respective contribution to the excess smoothness of expectations.

This paper provides an empirical contribution to these important questions. It estimates a two margin forecast formation process that allows for forecast rounding. Forecasters decide when to revise their forecast (extensive margin). When they do, they slowly incorporate new information (intensive margin) and may report a rounded value for their new forecast (rounding). Simultaneously accounting for these three forms of rigidities allows to encompass a broad class of potential frictions in forecast formation.[97] Using individual data from the ECB Survey of Professional Forecasters, the paper tests new predictions confirming that (i) rigidities exist at both margins and rounding is a salient feature of the data that must be accounted for. It then (ii) estimates the degree of rigidities at each margin using individual forecasts and (iii) derives quarterly time series for inflation, GDP growth and unemployment information frictions. Furthermore, it demonstrates that, for all three types of forecast, (iv) the overall forecast stickiness is almost exclusively the consequence of the rigidities at the intensive margin.[98]

As a starting point, the paper studies a framework inspired from the two margin forecast model developed in Andrade and Le Bihan (2013). They consider, as we do, that forecasters might not systematically revise their forecasts, e.g., because of Sticky

[97]For instance, Sticky information (Gabaix and Laibson, 2001; Mankiw and Reis, 2002; Carroll, 2003), Noisy information (Lucas, 1972; Sims, 2003), consideration sets (Caplin et al., 2019), communication costs (Bec et al., 2017), generalized overreactions and overconfidence (Bordalo et al., 2020), cognitive discounting (Gabaix, 2020).

[98]More specifically, the rigidities at the extensive margin barely affect the overall forecast stickiness (increases it by 4 to 5 %) and play no role in explaining its variation over time. The effect of rounding is negligible for the overall stickiness (decreases it by less than 0.4%).

information (Mankiw and Reis, 2002), and slowly incorporate new information when they do, e.g., because of Noisy information (Lucas, 1972; Sims, 2003). Moreover, we also rely on the same dataset: the European survey of professional forecasters. Our objective is however different. We aim at estimating information frictions directly from the stickiness in individual forecast revisions.[99]

More specifically, Section 4.2 derives novel implications for the two margin forecast model that allow to estimate the rigidities at each margin directly from forecast revisions. The key parameter to estimate at the extensive margin is the share of forecasts that are revised.[100] The key parameter to estimate at the intensive margin is the rate at which new information is incorporated into revised forecasts.[101] We provide two ways to estimate these parameters. A direct estimation from individual forecast revisions and an indirect estimation from consensus (average) revisions. Both show that a key identifying assumption is that the participation at the extensive margin is observable.

Fortunately, the two margin information model predicts that the participation at the extensive margin is directly observable from zero-revisions: a forecaster reports the same forecast at two successive dates if and only if she did not revise her forecast.[102] This prediction has been widely used in the literature to estimate the share of forecasts that are revised directly from the observed zero-revisions (e.g. Dräger and Lamla (2012); Andrade and Le Bihan (2013); Pfajfar and Santoro (2013); Binder (2017); Fuhrer (2018); Baker et al. (2020); Giacomini et al. (2020)).

However, we find that the data clearly reject this prediction. To show that, we propose a simple test that compares the estimates obtained from consensus revisions and individual panel revisions. Both require a prior identification of the participation at the extensive margin. When assuming that the participation is observable from zero-revisions, we find that, on average, about 75% of forecasts are revised each quarter (panel estimates). Similarly, one can also obtain an estimate of the average participation at the extensive margin

[99]Looking at individual forecast revisions allows to asses the contribution of each form of rigidities, avoid problems related to aggregation in consensus forecasts and provide more efficient estimates. See the literature review for a discussion.

[100]This parameter is generally interpreted as a measure of the frequency of the sporadic updating of the information set (e.g. Sticky information).

[101]This parameter is generally interpreted as a measure of the precision of the information observed by forecasters (e.g. the Kalman gain in Noisy information models).

[102]This equivalence holds almost surely.

by comparing the autocorrelation in the consensus revision when successively computed on the subset of revised forecasts and all forecasts. The results that we obtain from this second method are at odds with the aforementioned panel estimates, and hence the prior assumption regarding the observability of participations. Instead, they indicate that we should observe much more revisions (about 90%). The difference are significant for the three types of forecasts (inflation, GDP growth and unemployment) that we analyse and robust to horizon effects.

A reason for the rejection of the two margin information model could be that forecasters report rounded forecasts (Andrade and Le Bihan, 2013). We therefore extend the model to account for forecast rounding. In this new framework, zero-revisions are either a consequence of forecast rounding or the (absence of) participation at the extensive margin.

Section 4.3 develops a method to simultaneously estimate the participation at the extensive margin and information rigidities at the intensive margin in the presence of rounding. In doing so, we consider that the absence of participation at the extensive margin coincides with those zero-revisions that are unlikely to be explained by forecast rounding. Consequently, each zero-revision is ranked according to the probability that it results from rounding. We develop an iterative estimation algorithm – that accounts for the endogenous selection of participants and its impact on estimated information frictions and estimated probabilities to participate – that we implement to estimate a threshold governing the inclusion or exclusion of those zero-revisions among participants in order to minimize the residual sum of squares.

Applying the aforementioned test to this second (estimated) sample of participants, the data do not reject the prior identification of the participation at the extensive margin anymore. The results indicate that about 96% of forecasts are revised each quarter on average. That is, most of the observed zero-revisions are the consequence of forecast rounding and not of rigidities at the extensive margin. Nevertheless, rigidities are statistically significant at each margin and accounting for the rigidities at the extensive margin significantly improves the model goodness-of-fit. This is essentially because including the sample of non-revised forecasts drastically increases the estimated rigidities at the intensive margin. The latter are large. Our results indicate that forecasters who revise their forecasts incorporate between 50 and 60% of the news they receive each quarter on average. Applying Sheppard's correction to asses the impact of rounding for revised forecasts, we find that it has a negligible impact.

We then move on to the second objective of the present paper and derive variable-specific time-series for the key parameter at each margin. The identification of participants at the extensive margin allows to directly estimate the share of non-revisers at each quarter. We observe important variations over time. In particular, there are periods when almost all forecasters revise their forecast and periods when almost 20% of forecasters do not revise.

The panel structure of the survey also allows to use interaction terms to compute time-series for the evolution of information frictions at the intensive margin. This approach does not impose smoothing assumptions regarding their evolution over time and is thus well suited to study high frequency variations in information frictions. The trend in information acquisition has remained relatively constant over the 1999-2019 period and there is no sudden structural change over this period. We nevertheless observe medium term variations that are common to all three series (inflation, GDP growth and unemployment forecasts). Information acquisition has slowly decreased over the 2000-2008 period, a period of relative economic stability in the euro area. It started to increase afterward and remained stable in between the two most recent recessions. Since 2016, the trend in information acquisition seems to be on an increasing path again. Notably, information acquisition has become more volatile in the post 2008 area. Looking at the distribution of the estimated Kalman gains, we see that the distributions are generally positively skewed, thus suggesting that most of the time forecasters pay relatively little attention to new information but that there exists some rare periods when they reach near full-information forecasts.

This paper estimates information frictions for both margins. However, in macroeconomic models with representative agents, information rigidities are generally introduced to capture the excess smoothness in average expectations (Fuhrer, 2017, 2018). Therefore, we demonstrate that both forms of rigidities map into a single parameter driving the overall smoothness in the consensus forecast under a mild assumption. Decomposing the contribution of each margin to the overall consensus forecast stickiness, we find that the rigidities at the extensive margin barely affect it (increases it by 4 to 5 %) and play no role in explaining its variation over time.

Finally, the paper discusses the state-dependence of information frictions by studying its dynamics during recessions. This exercise is meant to illustrate that the time series we constructed are not the result of mere randomness. Our analyse closely follows the methodology in Coibion and Gorodnichenko (2015). Averaging information frictions over all forecast types, they find that U.S. forecasters tend to be more attentive in the aftermath of a recession and that this increase is persistent. Our data confirm a similar pattern for European forecasters. However, we can use the time series for each forecast type in order to derive variable-specific dynamics during recessions. Doing so, we observe very different dynamics for inflation, GDP growth and unemployment. More specifically, we find a sharp increase in information rigidities for inflation and GDP growth that is more than compensated by a decrease in information frictions for unemployment. In future work, these time series could be used to analyse the determinants of attention allocation over time.

## Related Literature

Our paper relates to the recent, yet dense, literature measuring information frictions from forecast revisions and errors.[103] This literature provides convincing evidence that

[103] A related literature estimates information frictions from forecast disagreement (e.g. Mankiw et al. (2003); Branch (2007); Coibion and Gorodnichenko (2012); Dovern et al. (2012); Andrade and Le Bihan (2013); Dovern (2015); Andrade et al. (2016); Giacomini et al. (2020)) and also finds that information frictions are important. Within this literature, Andrade and Le Bihan (2013) also estimate a model of Sticky and Noisy information using the method of estimated moments to match some features of forecast disagreement. The current paper instead focuses on the ability of this model to match the stickiness of forecasts. As was recently pointed out by Reis (2020), 'It is this stickiness of expectations that, in varied ways, the models of the last two decades have tried to make sense of.' (p.7). In order to match the disagreement among forecasters, Andrade and Le Bihan find that Sticky information must be extremely large (10% of revisions each quarter). They however find that these same rigidities are much lower in individual data (about 75%) and conclude that the Sticky and Noisy information model is rejected by the data. We reach opposite conclusions. We find that the data reject their identifying assumption and conclude that Sticky information is in fact even much lower than what they find (about 95%, from both individual and consensus forecasts). Nevertheless, the Sticky and Noisy information model is still able to capture the observed excess smoothness of consensus forecasts and is coherent with salient features of individual data once we account for rounding.

consensus forecasts are excessively smooth in comparison to the predictions from the full information rational expectation (FIRE) hypothesis (Coibion and Gorodnichenko, 2015; Dovern et al., 2015; Broer and Kohlhas, 2018; Bordalo et al., 2020). It has a large echo because its findings have significant implications in macroeconomics (Sims, 2003; Mankiw and Reis, 2006; Maćkowiak and Wiederholt, 2015; Fuhrer, 2017; Carroll et al., 2020). This literature is also the primary source used to calibrate the key parameters of widely-used models of information frictions, e.g., the Sticky information model in which agents update their expectations sporadicly (Gabaix and Laibson, 2001; Mankiw and Reis, 2002; Carroll, 2003) or the Noisy information model in which agents observe noisy signals (Lucas, 1972; Sims, 2003).

Nevertheless, existing approaches based on consensus forecasts face some qualitative and quantitative limitations. First, they hardly help to distinguish between competing theories of information frictions.[104] Second, the estimates obtained from consensus forecasts are inefficient.[105] Third, we cannot estimate time series for the evolution of information frictions without making strong smoothing assumptions about their dynamics.[106] Understanding the drivers of sudden variations in information rigidities is, however, of the utmost importance for the conduct of monetary and fiscal policies.

Our paper takes forward on these limitations. First, it develops a novel estimation strategy based on consensus forecasts that allows to separately estimate and test the contribution of Sticky and Noisy information. It finds that both forms of rigidities coexist, but that Noisy information is, by far, the main cause of the excess smoothness

[104]For instance, when information is both Sticky and Noisy (Andrade and Le Bihan, 2013; Giacomini et al., 2020), the smoothness in consensus forecasts is a byproduct of these two forms of rigidities and one cannot asses the relative size of each from consensus forecasts. More generally, there are many alternative explanations for the observed excess smoothness of consensus forecasts. Consensus forecasts may be excessively smooth even in the absence of systematic information frictions at the individual level (Pesaran and Weale, 2006; Crowe, 2010), when individuals over-react to news (Bordalo et al., 2020) or under alternative behavioral biases (Broer and Kohlhas, 2018; Angeletos et al., 2020).

[105]They are estimated on small sample time series and, hence, come with large standard errors. Using the full panel structure of forecast surveys naturally results in more efficient estimates. For instance, Ryngaert (2017) uses Monte Carlo simulations to show that one can reasonably expect a 20% decrease in the standard errors when relying on the panel structure of the data instead of the consensus forecasts.

[106]This is also a direct consequence of the small sample size. For instance, Larsen et al. (2020) have to fit a Latent Threshold Model to estimate the time variation in information frictions. There is evidence that information frictions decrease during recessions (Coibion and Gorodnichenko, 2015; Andrade and Le Bihan, 2013). In comparison, our paper estimates quarterly time series for information frictions. These time series confirms that information frictions decrease during recession. They however provide a more complete picture of the variations in information frictions.

in consensus forecasts. Second, it develops a new framework to directly estimate these rigidities from individual forecast revisions. Using the full panel structure of the data highly increases the precision of our estimates (by 70%) and allow us to estimate time series for the evolution of information frictions. The evidence drawn from individual forecast revisions support the Sticky and Noisy information theories. They predict a large degree of information rigidities and are in line with our evidence from consensus forecasts.

Some papers have already estimated either Sticky or Noisy information rigidities – but not both simultaneously as we do in this paper – from individual forecast data. In the following, we respectively discuss how we relate and contribute to each.

A key prediction of Sticky information is that a forecaster reports the same fixed-event forecast at two successive dates if and only if she did not revise her information set. Building on this prediction, Andrade and Le Bihan (2013) estimate the rigidities at the extensive margin[107] directly from the share of zero-revisions observed in the data. This method has been widely applied (e.g. Dräger and Lamla (2012); Pfajfar and Santoro (2013); Binder (2017); Fuhrer (2018); Baker et al. (2020); Giacomini et al. (2020)).

There are, however, reasonable concerns related to the usefulness of this measure. For instance, we observe about 25% of zero-revisions in the European SPF each quarter, while, when asked, only 10% of the same European forecasters report that they do not revise their forecast at least once each quarter (Meyler and Rubene, 2009).[108] Second, the frequency of surveys seems to matter, with more frequent surveys resulting in relatively higher shares of revisions.[109]

[107]Since we do not directly observe individuals' information sets, we use individual forecast revisions as a proxy for information set revisions. However, other forces, such as communication costs and strategic behaviors, may imply that forecast revisions are an imperfect proxy for a revised information set. In the paper, we encompass all the rigidities that may prevent forecast revisions in one category: the rigidities at the extensive margin. Sticky information is just one form of the rigidities at the extensive margin.

[108]Similarly, only 4% of the US-SPF forecasters declare that they neither update their forecasts at a monthly or quarterly frequency (Stark, 2013).

[109]For instance, Binder (2017) uses a monthly forecast survey and finds that 95% of respondents revised their forecast at least once in a 6 months window. However, when comparing forecasts made at period t and t+6, she finds that only 70% of forecasts have been revised. This result is puzzling. Indeed, if a forecaster updates her forecast for a continuous variable after a month, then we should also observe a change in the less frequent survey. Indeed, since she is forecasting a continuous variable, the probability that she reports again her forecast from period $t$ after having updated at least once is zero (with FIRE or Noisy information). Without further frictions, both frequencies should be the same. An explanation could be that forecasters have a discrete consideration set, e.g., they round their forecasts.

Motivated by these intriguing observations, we develop a test for the internal coherence of this measure. Our test relies on a prior identification of unrevised forecasts and estimates the implied share of revision from consensus revisions. We find that it strongly rejects the hypothesis that the prior identification coincides with zero revisions. Instead, it indicates that we should observe much more revisions.

An explanation could be that forecasters round their forecast. This is a well-documented feature of forecast survey data. The impact of rounding on the estimated rigidities at the extensive margin is however unclear. Some argue that it may be small, e.g., Andrade and Le Bihan (2013),[110] while others argue that it may be important, e.g., Dovern et al. (2015); Binder (2017).[111] In this paper, we explicitly estimate the impact of forecast rounding on individual forecasts. The resulting (estimated) prior identification of unrevised forecasts is not rejected anymore. We find that rounding explains about 80% of the observed zero revisions. Consequently, the implied rigidities at the extensive margin are even smaller than usually reported in previous studies using quarterly surveys. Only 5% of forecasts are not revised each quarter. This finding is coherent with the 10% reported by the European forecasters themselves and the conclusions from monthly surveys.[112] Nevertheless, we also conclude that these rigidities are statistically significant and should not be ignored when estimating other information rigidities from individual data.

In comparison, there is relatively less work on the slow incorporation of information in individual forecasts.[113] In this paper, we show that the coefficient from regressing the

[110]Andrade and Le Bihan (2013) asses the impact of rounding by fitting a VAR on macroeconomic time series and look at the forecasts from this VAR. This approach implicitly presupposes that forecasters know the model economy (here a VAR) and have access to full information when updating. Both hypothesis are discutable and, in particular, the latter hypothesis is not supported by the fact that they, and we, find that information is imperfect for revised forecasts, i.e., forecasters slowly incorporate information when revising. Their measure is therefore a lower bound for the impact of rounding.

[111] Dovern et al. (2015) base their argument on Andrade and Le Bihan's VAR exercise. Binder (2017) makes her argument using fixed-horizon forecasts which cannot be related to the rigidities at the extensive margin without making assumptions on the dynamics of the variable being forecasted, e.g., random walk.

[112]Even though these monthly surveys generally look at fixed-horizon (one-year ahead) forecasts.

[113]In an unpublished paper, Ryngaert (2017) extends the methodology in Coibion and Gorodnichenko (2015) to individual forecast data. The main limitation of this approach is that it requires a guess for the dynamics of the forecasted macro variable. Furthermore, it requires to measure forecast errors and, therefore, relies on external data. Thus, this approach also raises questions regarding whether the econometrician should use real-time data or corrected time series to measure forecast errors, and is costly to extend to multiple countries and macroeconomic variables. Also in a working paper, Fuhrer (2018) estimates the excess smoothness in individual forecasts from forecast revisions. His work is related to

deviation of an individual's forecast revision w.r.t. the average revision on the deviation of her prior forecast w.r.t. the previous average forecast equals the rate at which she incorporates new information, i.e., the Kalman gain in Noisy information models. Moreover, we illustrate that it is important to control for the presence of rigidities at the extensive margin when estimating these rigidities at the intensive margin.

A key prediction of the *(rational)* noisy information model is that individual forecasts are rational, i.e., individual forecast errors are not predictable given one's information. As was recently pointed out, this strong prediction is useful in order to distinguish the noisy *rational* expectation framework from other competing theories of information frictions and behavioral biases (Broer and Kohlhas, 2018; Fuhrer, 2018; Bordalo et al., 2020; Angeletos et al., 2020).[114] Instead, in this paper we do not asses the optimality of noisy information filtering, but are interested in the excess smoothness in individual forecasts. We measure the latter as the weight individuals put on their respective prior beliefs. Arguably, this weight captures the essence of Noisy information: individuals smooth their expectations to prevent from their imperfect idiosyncratic information. Whether this smoothing is optimal in the MSE-sense is a different, yet also fundamental, question.

The paper is organized as follows. Section 4.2 develops and estimates a two margin forecast model. It shows that this model is rejected by the data. Section 4.3 extends this model to account for forecast rounding. This model is not rejected by the data anymore. Section 4.4 discusses the evolution of information frictions over time for each margin. Finally, Section 4.5 illustrates how the rigidities at each margin map into the consensus

ours (at the intensive margin). The main difference is that we estimate a within transformation of his regression that is robust to unobserved common information. Furthermore, we analyse the variation of the excess smoothness over time.

[114]For sake of comparison, we reproduce the analyse from these papers using our data from the European survey of professional forecasters. The results are reported in Table 4.5 in the Appendix. It shows that this prediction is generally not rejected by our data. We only reject this prediction at the 5% (or 10% with Driscoll and Kraay robust standard errors which are robust to general forms of cross-sectional and temporal dependence) for unemployment forecasts. As can be seen from the right panel in Table 4.5, this rejection is driven by non 'rational' behaviors at specific horizons, e.g., for the one year ahead forecasts. When introducing horizon interactions, we do not reject it at the 5% level for 5 out of the 8 horizons.
Broer and Kohlhas (2018) also looked at HIPC inflation and GDP growth forecasts from the European survey of professional forecasters. We reach a different conclusion for GDP growth forecasts. The only difference with respect to their study is that we use fixed-event calendar forecasts, while they use fixed-event rolling forecasts. Analyzing wether individual calendar forecasts are generally more 'rational' than rolling forecasts is behind the scope of this paper.

forecast stickiness and analyzes the dynamics of this stickiness during recessions.

## 4.2 A two margin forecast model

In this section, I introduce a two margin forecast formation process with Noisy and Sticky information and propose some statistics to estimate the key parameters measuring information frictions at each margin from panel data. These statistics assume a correct identification of the participation at the extensive margin, i.e., whether a forecast is revised. I estimate the model using the ECB survey of professional forecasters and show that a forecasting model with these two rigidities only is incoherent with the data.

### 4.2.1 The Forecast Formation Process

I consider the problem of a forecaster who faces two forms of rigidities. First, information may be imperfect and the forecaster slowly incorporates noisy information about the variable she is forecasting. Second, forecasters may face an adjustment cost to revise their forecasts. This cost could represent a communication cost associated to the publication of a new forecast, the cost of collecting new data releases and to estimate a forecasting model, a credibility loss that would result from too frequent forecast revisions, or a fixed cognitive cost for processing new information. This section discusses the behavior of a forecaster facing both rigidities.

#### 4.2.1.1 The Intensive Margin: Noisy Information

Let $x_{t+h|t}$ denote the full-information rational forecast of variable $x_{t+h}$ at time $t$. Following the Noisy information literature, I assume that, conditionally on revising her forecast, forecaster $i$ is unable to observe $x_{t+h|t}$, but instead observes an unbiased Gaussian signal $s_{i,t,h} = x_{t+h|t} + \omega_{i,t,h}$. I allow the precision of the signal to vary over time and forecast horizons. Furthermore, and again conditionally on revising, a forecaster $i$'s forecast is a linear combinaison of the signal and her prior forecast[115]

$$F_{i,t,h} = (1-k_{t,h})F_{i,t-1,h+1} + k_{t,h}s_{i,t,h} \tag{4.1}$$

$$\iff \quad r_{i,t,h} = k_{t,h}(\bar{x}_{t+h|t} - F_{i,t-1,h+1}) + k_{t,h}\omega_{i,t,h} \tag{4.2}$$

where $F_{i,t,h}$ is forecaster's $i$ forecast of the $h$-periods ahead aggregate variable $x_{t+h}$ at period $t$, $r_{i,t,h} \equiv F_{i,t,h} - F_{i,t-1,h+1}$ is her forecast revision and $k_{t,h}$ is the Kalman gain, i.e., the weight she puts on the signal $s_{i,t,h}$. Equation (4.2) may not be directly estimated because the full-information rational expectation of $x_{t+h}$ at time $t$ is unknown. Nevertheless, we can define $\eta_{t,h} = \bar{F}^{u(t)}_{t,h} - \bar{x}_{t+h|t}$ the average forecast error (with respect to the full-information rational forecast of variable $x_{t+h}$ at time $t$) among updating forecasters. The superscript $u(t)$ is used to highlight that the average is computed among revising forecasters at time $t$. Introducing this expression in equation (4.2) generates a fixed effect that is correlated with the independent variable. I will thus estimate a within transformation of equation (4.2)

$$r_{i,t,h} - \bar{r}^{u(t)}_{t,h} = k_{t,h}(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}) + \varepsilon_{i,t,h} \tag{4.3}$$

where the error term $\varepsilon_{i,t,h} = k_{t,h}(\omega_{i,t,h} - \bar{\omega}^{u(t)}_{t,h})$ depends on the signal noises at time $t$ and is therefore uncorrelated with forecasts at time $t-1$. Appendix 4.8 demonstrates how to derive this within transformation from (4.2). Equation (4.3) offers a simple way to estimate information rigidities on micro-data. In comparaison to the other methods proposed in the literature, it has the advantage of providing a micro estimate that directly maps into the Kalman filter $k$ and does not require to make an assumption for the dynamics of the aggregate variable $x_t$, nor to take a stand on an appropriate external data source for $x_t$ (e.g. real-time data). Moreover, this within transformation would be unaffected by the presence of an additive noise $\alpha_{t,h}$ common to all forecasts.

[115] These results naturally obtain for optimal forecasts in the MSE-sense (Coibion and Gorodnichenko, 2015) and other models of information filtering (Broer and Kohlhas, 2018; Fuhrer, 2018; Bordalo et al., 2020; Angeletos et al., 2020).

#### 4.2.1.2 The Extensive Margin: Sporadic Revisions

Forecasters may also face an adjustment cost and, therefore, revise their forecast only sporadically. Let $\lambda_{t,h}$ be the share of $h$-periods ahead forecast revisions at period $t$. Following the literature on adjustment costs, the choice to revise a forecast may be random, i.e., revising forecasters are randomly selected with a constant probability $\lambda_{t,h}$; time-dependent, i.e., forecasters revise on a predetermined and periodic basis; state-dependent, i.e., a forecaster's choice to revise depends on the actual state of the economy; information-dependent, i.e. a forecaster's choice to revise depends on her perceived state of the economy; or any combinaison of these rules. For non-revising forecasters, we have $r_{i,t,h} = 0$, whereas revising forecasters adjust their forecast according to equation (4.1). Consequently, we have

$$r_{i,t,h} = \begin{cases} k_{t,h}(\bar{x}_{t+h|t} - F_{i,t-1,h+1}) + k_{t,h}\omega_{i,t,h} & \text{if} \quad u_{i,t,h} = 1 \\ 0 & \text{if} \quad u_{i,t,h} = 0 \end{cases} \tag{4.4}$$

where $u_{i,t,h}$ is a dummy variable taking the value one when forecaster $i$ revises her $h$-periods ahead forecast at period $t$.

### 4.2.2 Estimation

The key parameters to estimate here are the information rigidities at the intensive margin ($k_{t,h}$) and the participation rate at the extensive margin ($\lambda_{h,t}$). In order to provide a transparent discussion, I disregard the time variation of these parameters and delay the analyse of this source of heterogeneity to a latter section.

The only difficulty in estimating $k_h$ and $\lambda_h$ relates to the identification of the participation at the extensive margin ($u_{i,t,h}$). Equation (4.4) however implies that $P(r_{i,t,h} = 0|u_{i,t,h} = 1) = 0$. That is, the model predicts that revised forecasts are, in principle, almost surely identified from observed revisions $r_{i,t,h}$.[116]

[116] As is highlighted in Andrade and Le Bihan (2013), there are however strong evidence that ECB forecasters round their forecasts. When this is the case, it is not possible to directly identify revised forecasts from the value of revisions. I propose a method accounting for the possibility of forecast

Consequently, one can directly estimate the proportions of revised forecasts

$$\overline{\lambda_h} \quad = \quad \frac{\sum_{i,t} \mathbf{1}_{(r_{i,t,h}=0)}}{N_h} \tag{4.5}$$

This is the standard approach in the literature to estimate the parameter $\lambda_{h,t}$ in Sticky information models (Dräger and Lamla, 2012; Andrade and Le Bihan, 2013; Pfajfar and Santoro, 2013; Binder, 2017; Fuhrer, 2018; Baker et al., 2020; Giacomini et al., 2020)

Following from equation (4.3), the OLS regression with interaction terms

$$r_{i,t,h} - \bar{r}^{u(t)}_{t,h} \quad = \quad \sum_j k_j \ \mathbf{1}_{h=j}(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}) + \varepsilon_{i,t,h} \tag{4.6}$$

estimated on the subsample of observations such that $r_{i,t,h} \neq 0$ provides estimates for $k_h$. These estimates are denoted $\hat{k}_h$. Appendix 4.8 demonstrates that the equivalent estimates obtained from the full sample of observations is a downward biased estimate of the rigidities at the intensive margin in the presence of rigidities at the extensive margin. More specifically, pooled regressions – let it be on consensus or individual data – estimate a byproduct of the rigidities at both margin and may not be used to assess the relative contribution of each margin.

These estimates are model-consistent. They are however highly sensitive to the identification of the participation at the extensive margin. In order to asses the validity of this identification, I therefore also provide an indirect estimate of the participation at the extensive margin based on consensus forecasts.

Averaging equation (4.1) on revising forecasters and computing the average revision as in Dovern et al. (2015), we get

$$\bar{r}^{u(t)}_{t,h} \quad = \quad (1-k_h)\bar{r}^{u(t)}_{t-1,h+1} + \vartheta_{t,h} \tag{4.7}$$

where the error term is a martingale difference, i.e., the difference between the rational expectations of $x_{t+h}$ at two successive periods.

rounding in Section 4.3 and demonstrate that accounting for forecast rounding is crucial for the model estimation.

Accounting for both margins and averaging over all forecasts, it also holds

$$\begin{aligned}\bar{r}_{t,h} &= \lambda_h \bar{r}^{u(t)}_{t,h} \\ &= \lambda_h (1-k_h) \bar{r}^{u(t)}_{t-1,h+1} + \lambda_h \vartheta_{t,h} \end{aligned} \tag{4.8}$$

The first equation states that the average revision among all forecasts is a weighted average of the average revision of revised forecasts and the average revision of non-revised forecast which is nil. Using the expression for the average revision of revised forecasts from equation (4.7), we obtain equation (4.8). The latter equation indicates that the stickiness in the consensus forecast is driven by both margins. Importantly, the ratio of the slope estimates from fitting equations (4.8) and (4.7) provides an estimate for $\lambda_h$. Because of the limited number of observations for consensus forecasts, I estimate a random slopes and intercepts model. The (common-effect) estimate obtained from this ratio is denoted $\hat{\lambda}$.

Interestingly, under the assumption that the identification of the participation at the extensive margin is correct, the difference between the indirect and direct estimates ($\hat{\lambda}-\overline{\lambda}$) should be small. Any significant difference between these estimates would raise doubts on the identification of the participation at the extensive margin.

### 4.2.3 Results

**The data** come from the ECB survey of professional forecasters. I use this survey because it combines some characteristics which are important for this study and that makes it quite unique. First, the surveys directly ask respondents to provide a rate for inflation, GDP growth and unemployment. We therefore do not have to construct these rates from actual data or previous forecasts.[117] Second, forecasters are asked to report the same calendar forecast each quarter and for a couple of years. Thanks to these two features, we observe forecasters' revisions in their own forecasts at each quarter for a long period as they report

[117]Table 1 (panel B) in Bordalo et al. (2020) illustrates the importance of these ad-hoc corrections on forecast data. Using the US SPF and Blue Chip Financial Forecasts, they find an average share of zero revisions of about 15% (min 9%, max 22%) when forecasters are asked to report a rate, while these shares are on average equal to 4% (min 0%, max 10%) when they compute rates from forecasts that are initially reported in levels. In conclusion, it is very likely, and not surprising, that potential ad-hoc data corrections may overturn sensible features of the data such as zero revisions.

it and not as it was computed from the econometrician. Arguably, a drawback of this survey is that it focuses on a specific population. Nevertheless, professional forecasters are likely more sophisticated than other agents in the economy. Therefore, the information rigidities estimated on this population may represent a lower bound for other populations.

Many papers have already used the European survey of professional forecasters and Andrade and Le Bihan (2013) provide a general presentation of the survey characteristics. The survey covers a period from the first quarter of 1999 to the last quarter of 2019. I focus on short-term (up to 2 years) inflation, GDP growth and unemployment rate calendar forecasts. On average, we observe 46.7 forecasts per quarter for a given calendar year. In each quarter, forecasters are surveyed about their macroeconomic forecasts for fixed events, e.g., inflation rates in the current and next year. Therefore, we are able to observe successive revisions of a same forecast across time for a given forecaster. For each calendar forecast between 2000 and 2018, we observe up to 8 individual forecasts. However, there are at most 4 forecasts per forecaster for inflation in 1999 (first year in the survey) and at most 4 forecasts for 2020.

**Descriptive statistics on revisions** clearly indicate that forecasts are not systematically revised. Indeed, and thanks to structure of the ECB survey of professional forecasters, one can directly observe forecast revisions from the calendar forecasts. Each quarter, 76%, 79% and 75% of forecasters respectively revise their inflation, growth and unemployment rate forecasts on average. As is reported in Figure 4.1, the probability to revise a forecast varies with forecasters' perception gap and revision history. The perception gap is defined as the difference between a forecaster's prior forecast $F_{i,t-1,h+1}$ and the average forecast of revising forecasters $\bar{F}_{t,h}^{u(t)}$.[118] We observe that the larger this error in absolute value, the more likely a forecaster is to revise her forecast. Similarly, the probability to update also depends on the number of quarters without revisions. Surprisingly enough, the unconditional hazard rates are neither constant nor increasing. Instead, forecasters who did not update for at least one quarter are less likely to update than forecasters who last updated one quarter ago. The statistics in Figure 4.1 provide indirect evidence on the endogeneity of the forecasters' choice to revise their forecast. Similarly, they also indi-

[118]This perception gap is an estimate of a forecaster's prior forecast error.

cate that forecasters do not continuously revise their forecasts, thus rolling out a forecast formation model with a noisy intensive margin only.

We now turn to our **estimation results** for the key parameters $k_h$ and $\lambda_h$ following the procedure discussed in Section 4.2.2. Table (4.1) reports the estimates $k_h$ from estimating the within transformation in equation (4.6). The estimates reported in the left-left panel are estimated on the subsample of the observations such that $r_{i,t,h} \neq 0$. It indicates that at any horizon $h$, the Kalman gain $k_h$ is significantly different from zero at the one percent level. Furthermore, we can also reject the null hypothesis of full-information, i.e., $k_h = 1$, at the one percent level at any horizon $h$. The left-right panel reports the estimated participation rate at the intensive margin. We can systematically reject that they are equal to one. Overall, the information reported in the left panel confirms the existence of information rigidities at each margin.

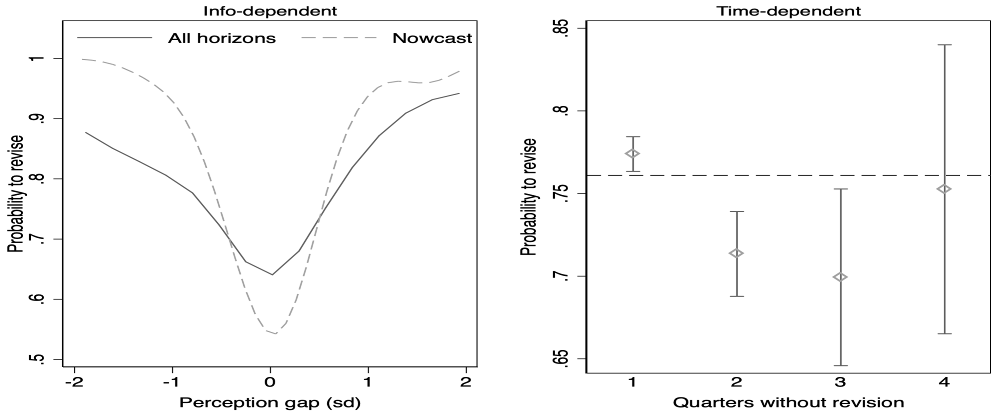
The right panel reports the estimates from the consensus forecasts. They systematically reject the null of full information. Moreover, we observe that the estimated degree of information rigidities that we get from consensus forecasts are not significantly different to those obtained on panel data. For inflation, the estimated Kalman gain is 0.518 while it is 0.619 with the panel data (respectively 0.328 and 0.555 for GDP growth, and 0.544 and 0.579 for unemployment). These differences are generally not statistically significant at the 5% level (excepted for GDP growth for which the difference is barely significant at the 5% level, but not at the 1%).

However, we get contradictory results regarding the participation rate at the extensive margin between the consensus forecast and panel estimates. According to the consensus forecasts, we should expect much more forecasts to be revised on average. Indeed, the estimated revisions rates are equal to 0.92, 0.96 and 0.88 based on the consensus forecasts for respectively inflation, GDP growth and unemployment. The values that we obtain from panel data generally lie outside the 95% confidence interval obtained from consensus forecasts (the only exception being unemployment where the value is equal to the lower bound of the confidence interval).

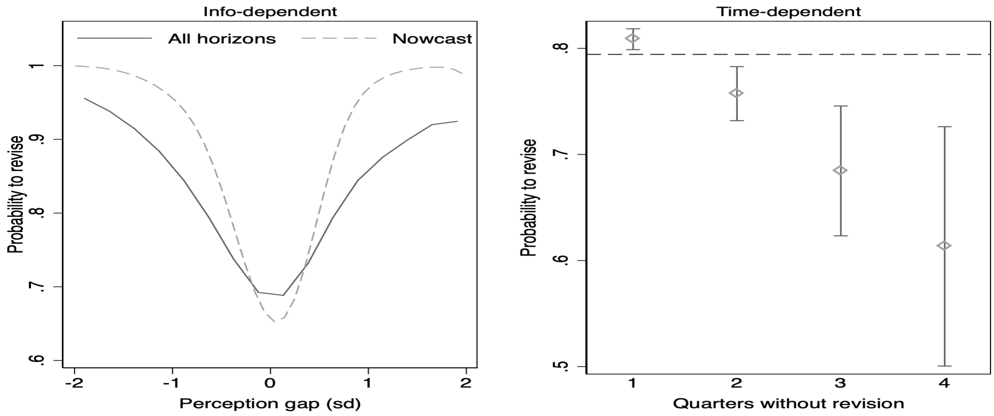
Overall, looking at either panel data or consensus forecasts, the data provide clear evidence on the existence of information rigidities at each margin. They however reject

Figure 4.1: Evidence on Forecast Revisions

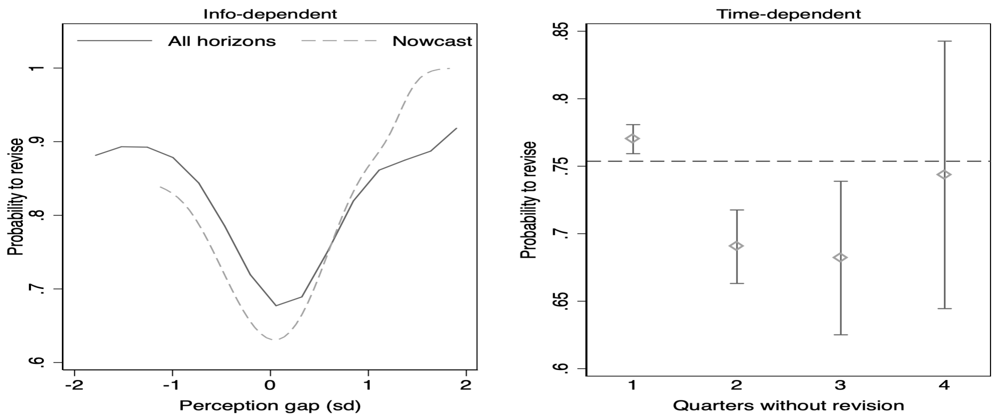
*Inflation*




*GDP Growth*




*Unemployment Rate*




Note: Panel A reports a Kernel-weighted local polynomial smoothing of the probability to revise a forecast as a function of the perception gap. The perception gap is the difference between a forecaster's prior forecast and the average forecast of revising forecasters. The later is an estimate of the full-information forecast. Panel B reports the hazard rate and its 95% confidence interval as a function of the number of quarters without forecast revision. The horizontal line is the unconditional probability to revise a forecast. Durations without revision for more than a year are not reported because of too few observations.

Table 4.1: Information Rigidities for Each Margin: Zero Revisions

| | **Panel Estimates** | | | | | **Consensus Forecasts** | | |
|---|---|---|---|---|---|---|---|---|
| | Intensive $k_h$ | | Extensive $\overline{\lambda_h}$ | | | Revisions $\beta_1 = (1 - k_h)$ | Pool $\beta_2 = \lambda_h(1 - k_h)$ | Ratio $\hat{\lambda}_h = \beta_2/\beta_1$ |
| | coef. | s.e. | coef. | s.e. | | | | |
| | | | | | ***Inflation*** | | | |
| Horizon | | | | | | | | |
| 0 | 0.890*** | 0.041 | 0.717 | 0.014 | coef. | 0.482 | 0.442 | **0.917** |
| 1 | 0.796*** | 0.024 | 0.817 | 0.013 | s.e. | 0.139 | 0.133 | 0.038 |
| 2 | 0.710*** | 0.051 | 0.842 | 0.011 | pval | 0.001 | 0.001 | 0.000 |
| 3 | 0.561*** | 0.023 | 0.817 | 0.012 | 95ci (low) | 0.210 | 0.182 | 0.842 |
| 4 | 0.551*** | 0.064 | 0.746 | 0.014 | 95ci (up) | 0.754 | 0.702 | 0.992 |
| 5 | 0.521*** | 0.058 | 0.704 | 0.015 | N | 162 | 162 | 162 |
| 6 | 0.436*** | 0.048 | 0.710 | 0.014 | | | | |
| 7 | 0.424*** | 0.030 | 0.723 | 0.015 | | | | |
| Sample Av. | 0.619*** | 0.021 | **0.761** | 0.005 | | | | |
| $(N, R^2)$ | 5,863 | 0.348 | 7,705 | . | | | | |
| | | | | | ***GDP Growth*** | | | |
| Horizon | | | | | | | | |
| 0 | 0.786*** | 0.040 | 0.833 | 0.012 | coef. | 0.672 | 0.642 | **0.955** |
| 1 | 0.669*** | 0.073 | 0.804 | 0.013 | s.e. | 0.110 | 0.108 | 0.021 |
| 2 | 0.562*** | 0.034 | 0.834 | 0.012 | pval | 0.000 | 0.000 | 0.000 |
| 3 | 0.475*** | 0.039 | 0.855 | 0.011 | 95ci (low) | 0.456 | 0.430 | 0.915 |
| 4 | 0.478*** | 0.068 | 0.820 | 0.013 | 95ci (up) | 0.888 | 0.853 | 0.996 |
| 5 | 0.451*** | 0.060 | 0.744 | 0.014 | N | 162 | 162 | 162 |
| 6 | 0.467*** | 0.079 | 0.704 | 0.015 | | | | |
| 7 | 0.530*** | 0.050 | 0.751 | 0.015 | | | | |
| Sample Av. | 0.555*** | 0.022 | **0.794** | 0.005 | | | | |
| $(N, R^2)$ | 6,038 | 0.308 | 7,602 | . | | | | |
| | | | | | ***Unemployment*** | | | |
| Horizon | | | | | | | | |
| 0 | 0.662*** | 0.109 | 0.682 | 0.015 | coef. | 0.535 | 0.470 | **0.879** |
| 1 | 0.567*** | 0.127 | 0.720 | 0.015 | s.e. | 0.179 | 0.176 | 0.064 |
| 2 | 0.630*** | 0.072 | 0.776 | 0.013 | pval | 0.003 | 0.007 | 0.000 |
| 3 | 0.666*** | 0.053 | 0.788 | 0.013 | 95ci (low) | 0.185 | 0.126 | 0.753 |
| 4 | 0.562*** | 0.045 | 0.783 | 0.014 | 95ci (up) | 0.886 | 0.814 | 1.005 |
| 5 | 0.507*** | 0.047 | 0.745 | 0.015 | N | 162 | 162 | 162 |
| 6 | 0.484*** | 0.046 | 0.756 | 0.014 | | | | |
| 7 | 0.542*** | 0.055 | 0.779 | 0.015 | | | | |
| Sample Av. | 0.579*** | 0.051 | **0.754** | 0.005 | | | | |
| $(N, R^2)$ | 5,560 | 0.373 | 7,377 | . | | | | |

NOTE: **(Intensive)** Results from estimating the Kalman gain $k_h$ from equation (4.3) on revised forecasts only ($r_{i,t,h} \neq 0$) and all observations. Two-way clustered standard errors at the forecaster and horizon levels. The sample average is the average marginal effect and its standard error is computed using the delta-method. *, ** and *** respectively denotes significance at the 10, 5 and 1% level. **(Extensive)** $\overline{\lambda_h}$ is the proportion of zero revisions in the data (equation (4.5). We can reject at the 1% level that any of the estimates $\overline{\lambda_h}$ is equal to either zero or one. **(Consensus Forecasts)** The first two columns respectively report the results from estimating equations (4.7) and (**??**) with a random slopes and intercepts model. The last column reports the ratio of the two coefficients obtained from these regressions. This ratio provides an estimate of the participation rate at the extensive margin. The standard errors are obtained from stratified bootstrap at the horizon level with 500 replications and inference assumes that the sampling distribution is normal.

one key prediction of the two margin forecast model. Namely, that the participation at the extensive margin can be directly identified from looking at whether a revision is nil. To the extend that the estimation of information frictions at each margin requires a correct identification of the participation at the extensive margin, one should remain cautious when interpreting the results from Figure 4.1. In the next section, I show that one can significantly improve the identification at the extensive margin by account for forecast rounding and, therefore, get accurate estimates for information frictions.

## 4.3 Extensive margin: accounting for rounding

As was first pointed out by Andrade and Le Bihan (2013), ECB forecasters generally report rounded macroeconomics forecasts. This behavior is clearly illustrated in Figure 4.2. An histogram of revisions shows systematic peaks at multiples of 0.1, with very few observations between two peaks.

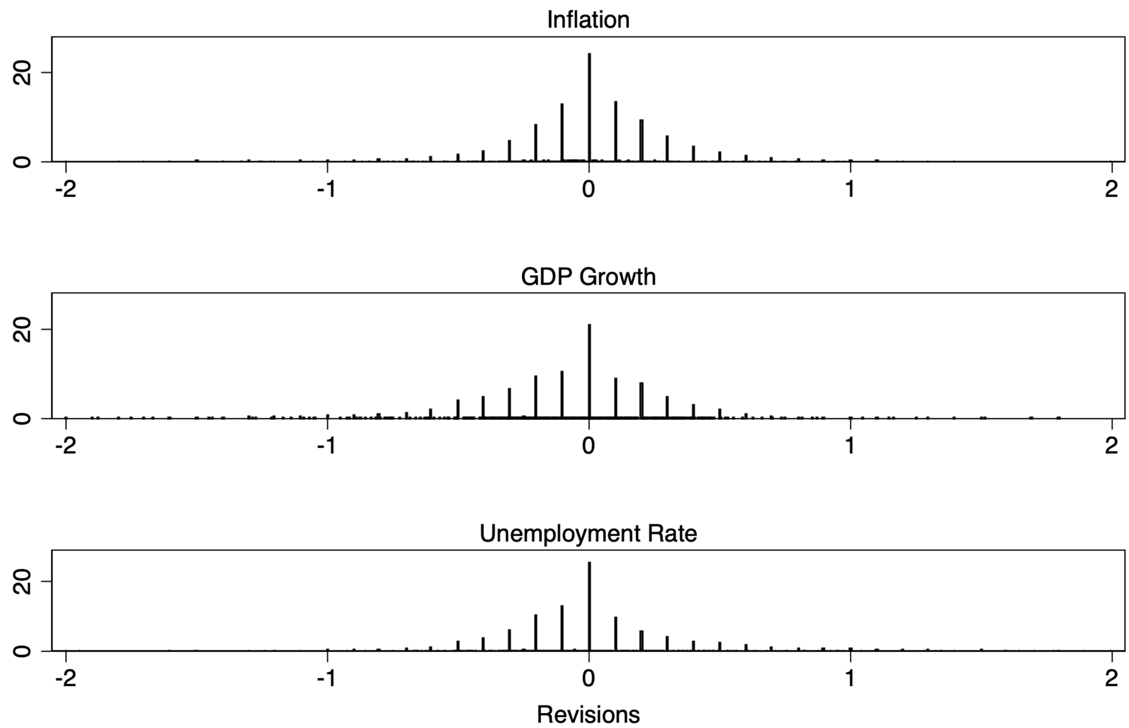
Accounting for the rounding of forecasts is essential as it may explain the observed excess number of zero revisions, the endogeneity of $u_{i,t,h}$ and, therefore, some of the empirical patterns discussed in the previous section. Consequently, this section proposes a conservative methodology that will allow us to test for the presence of forecast adjustment rigidities that may not be the consequence of forecasts rounding. It therefore provides an alternative identification of the participation at the extensive margin ($u_{i,t,h}$).

### 4.3.1 Forecast Formation with Rounding

In the presence of rounding, we are not able to observe the latent revisions $r^{\star}_{i,t,h}$ but instead observe $r_{i,t,h}$ a rounded value of these latent revisions. In the following, stared variables refer to the latent and unobserved variables. As a result, in the presence of rounding $P(r_{i,t,h} = 0|u^{\star}_{i,t,h} = 1) \geq 0$ and we cannot identify the participation at the extensive margin $u^{\star}_{i,t,h}$ directly from zero revisions anymore. In other words, there are potentially two explanations for the observed zero revisions: (i) the forecaster did not revise her forecast, (ii) or she did but reported the same forecast again because of rounding.

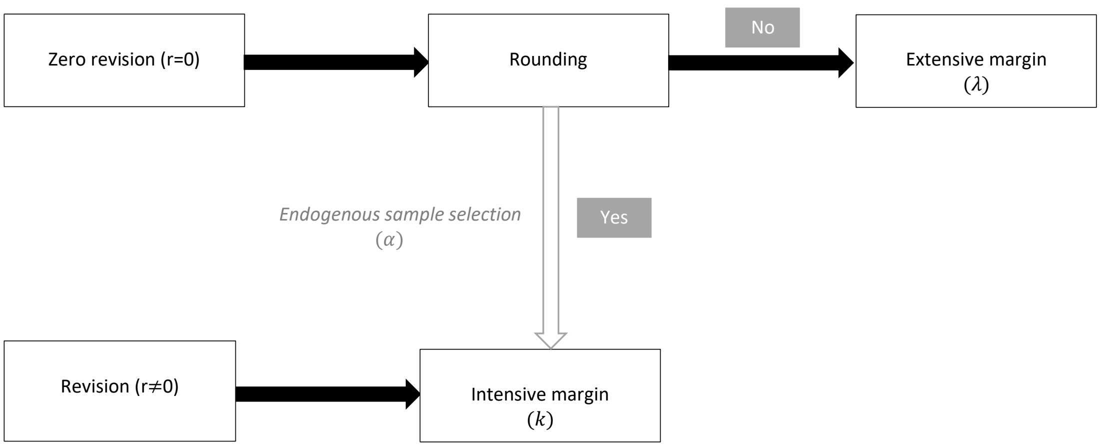
As is explicit from Figure 4.3, accounting for the existence of rounding is essential

Figure 4.2: Rounded Revisions




NOTE: Histogram (in percent) of revisions with a width of 0.005. The data are truncated to revisions within a $[-2, 2]$ interval.

Figure 4.3: A Two Margin Forecast Model with Rounding



as it generates an endogenous selection of the participation at the extensive margin. In the following, I extend the model presented in Section 4.2 to allow for the presence of rounding.

The latent data generating process follows from equation (4.4). Ignoring the time variation in the Kalman gains[119] and using the formulation in equation (4.3), we obtain

$$r^{\star}_{i,t,h} - \bar{r}^{u^{\star}(t)}_{t,h} = u^{\star}_{i,t,h}\Big(k_h(\bar{F}^{u^{\star}(t)}_{t-1,h+1} - F_{i,t-1,h+1}) + \varepsilon_{i,t,h}\Big) - (1 - u^{\star}_{i,t,h})\bar{r}^{u^{\star}(t)}_{t,h} \qquad (4.9)$$

and the probability of observing a rounded zero revision is

$$\begin{aligned} P(r_{i,t,h} = 0 | u^{\star}_{i,t,h} = 1) &= \Phi\left(\frac{a - k_h(\bar{F}^{u^{\star}(t)}_{t-1,h+1} - F_{i,t-1,h+1}) - \bar{r}^{u^{\star}(t)}_{t,h}}{\sigma_h}\right) \\ &\quad - \Phi\left(\frac{-a - k_h(\bar{F}^{u^{\star}(t)}_{t-1,h+1} - F_{i,t-1,h+1}) - \bar{r}^{u^{\star}(t)}_{t,h})}{\sigma_h}\right) \end{aligned} \qquad (4.10)$$

where $a$ is half the rounding precision, $\Phi(.)$ the standard normal c.d.f. and $\sigma_h$ the standard deviation of $\varepsilon_{i,t,h}$. This probability may be used to identify those zero revisions which are hardly likely to result from rounding. Building on the formalism of statistical hypothesis testing, we can define a risk $\alpha$ such that we can reject that the observed zero revision arrises as a consequence of rounding. Doing so, we are able to define a conservative decision rule for the identification of the participation at the extensive margin in the presence of rounding:

$$\hat{u}_{i,t,h} = 1 - \mathbf{1}_{\{r_{i,t,h}=0\}} \cdot \mathbf{1}_{\{P(r_{i,t,h}=0|u^{\star}_{i,t,h}=1)<\alpha\}} \qquad (4.11)$$

*Remark on the impact of rounding for the intensive margin:* If forecasts are rounded, it may, in principle, affect the estimation of information rigidities at the intensive margin. This is however unlikely to be problematic here. As is demonstrated in e.g. Schneeweiß and Komlos (2009), the OLS estimate of $k_h$ is biased (because the independent variable is rounded) and we can apply Sheppard's correction to approximately correct for this bias. Given the half rounding precision $a$ and the standard error of the independent

[119] The results are barely affected when one allows the Kalman gains to vary over time – through an interaction term – but keeps on assuming that the error terms are homoskedastic within forecast horizons.

variable, the corrected estimates are (on average) equal to 1.004, 1.003 and 1.002 time the estimates that we report for the intensive margin for respectively inflation, GDP growth and unemployment. That is, the impact of rounding is negligible at the intensive margin.[120]

### 4.3.2 Estimation Strategy

The parameters to be estimated in this model are the Kalman gains at the intensive margin $k_h$ and the threshold $\alpha$ for the participation at the extensive margin. They are estimated in order to minimize the residual sum of squares of the following simultaneous system of three equations

$$\begin{aligned} r_{i,t,h} - \bar{r}_{t,h}^{\hat{u}(t)} &= \hat{u}_{i,t,h}\Big(\hat{k}_h(\bar{F}_{t-1,h+1}^{\hat{u}(t)} - F_{i,t-1,h+1}) + \varepsilon_{i,t,h}\Big) - (1-\hat{u}_{i,t,h})\bar{r}_{t,h}^{\hat{u}(t)} && (4.12) \\ \hat{u}_{i,t,h} &= 1 - \mathbf{1}_{\{r_{i,t,h}=0\}} \cdot \mathbf{1}_{\{\hat{p}_{i,t,h}<\hat{\alpha}\}} && (4.13) \\ \hat{p}_{i,t,h} &= \Phi\left(\frac{a-\hat{r}_{i,t,h}}{\hat{\sigma}_h^{\hat{y}}}\right) - \Phi\left(\frac{-a-\hat{r}_{i,t,h}}{\hat{\sigma}_h^{\hat{y}}}\right) && (4.14) \end{aligned}$$

where $\hat{r}_{i,t,h}$ is the prediction for the revision, $\hat{\sigma}_h^{\hat{y}}$ its standard error and $a$ a constant set to 0.05. Defining $\hat{\varepsilon}_{i,t,h}$ the estimated error at the intensive margin

$$\hat{\varepsilon}_{i,t,h} \equiv r_{i,t,h} - \bar{r}_{t,h}^{\hat{u}(t)} - \hat{k}_h(\bar{F}_{t-1,h+1}^{\hat{u}(t)} - F_{i,t-1,h+1}) \quad (4.15)$$

the optimal estimates are given by

$$\Big(\{\hat{k}_h^\star\}_{h=0}^7, \hat{\alpha}^\star\Big) = \arg \min_{(\{\hat{k}_h\}_{h=0}^7, \hat{\alpha})} \sum_{i,t,h} \Big(\hat{u}_{i,t,h}\hat{\varepsilon}_{i,t,h}^2 + (1-\hat{u}_{i,t,h})(\bar{r}_{t,h}^{\hat{u}(t)})^2\Big) \quad (4.16)$$

In order to solve for this optimization problem, realize that for a given vector of $\hat{u}_{i,t,h}$, $\hat{k}_H$ and $\hat{\sigma}_H^{\hat{y}}$ are the standard OLS estimates – with homoskedastic errors – obtained from regressing equation (4.3) on the subset of observations such that $\hat{u}_{i,t,h} = 1$ and $h = H$. Given a value for $\hat{\alpha}$, one may then derive an updated vector of $\hat{u}_{i,t,h}$, estimate equation (4.3) again on the new subset of observations and repeat until convergence. Finally,

[120] This same remark explains why we do not account for the impact of rounding in the subsequent sections.

repeating this procedure over a grid of $\hat{\alpha}$, we can find the $\hat{\alpha}^{\star}$ that minimizes the overall residual sum of squares in equation (4.16). I therefore rely the estimation Algorithm 1 to solve problem (4.16).

**Algorithm 1** Estimation Algorithm

```
for all α̂ in grid do
    Set û_{i,t,h} = u_{i,t,h}
    while ||û_{i,t,h} − û'_{i,t,h}||_1 > tol do
        for H = 0 to 7 do
            Get OLS estimates k̂_H and σ̂_H^ŷ on observations such that û_{i,t,h} = 1 and h = H.
        end for
        Update û'_{i,t,h} from equations (4.13) and (4.14).
    end while
    Compute the RSS in equation (4.16).
end for
```

This estimation procedure closely follows that used to estimate threshold regression models. Indeed, we are implicitly estimating a two regimes threshold regression model with a step regime switching function given by equation (4.13) and an implicit lower regime where $k_{h,low} = 0 \ \forall h$. However, because the regime switching function depends on the estimates from the OLS regression, we must iterate on these two equations. The parameter standard errors are obtained from 1000 clustered bootstrap replications at the forecaster level.

### 4.3.3 Results

Table 4.2 reports the estimates obtained from estimating algorithm 1 for inflation, GDP growth and unemployment rate.

According to the estimated thresholds $\alpha$, most of the observed zero revisions are attributable to rounding. Indeed, when allowing for forecast rounding, we find that the participation rate at the extensive margin rises drastically to respectively 96.1%, 95.4% and 96.6% for inflation, GDP growth and unemployment.[121] Nevertheless, and even though the observed rigidities at the extensive are now much lower, these rigidities are still important to explain the data. Indeed, in comparaison to a constrained model with

[121] I do not report confidence intervals for the latter shares because it is well-known that bootstrap confidence intervals are generally inaccurate at the frontier of parameter sets.

Table 4.2: Panel Estimates for the Model with Rounding

| | Av. Kalman Gain $k$ | | Threshold $\alpha$ | | Prop. Revisers | |
|---|---|---|---|---|---|---|
| | coeff. (s.e.) | [95CI, bias corr.] [95CI, percentile] | coeff. (s.e.) | [95CI, bias corr.] [95CI, percentile] | $\lambda$ (s.e.) | zero rev. (s.e.) |
| | | $R^2$ | $(R^2)^c$ *(intensive only)* | | *F-stat* | |
| Inflation | 0.597 (0.023) | [ 0.552 ; 0.638 ] [ 0.556 ; 0.641 ] | 0.15 (0.053) | [ 0 ; 0.17 ] [ 0 ; 0.17 ] | 0.961 (0.008) | 0.835 (0.036) |
| | | *0.339* | *0.318* | | *243.80* | |
| GDP Growth | 0.543 (0.022) | [ 0.498 ; 0.582 ] [ 0.504 ; 0.588 ] | 0.13 (0.054) | [ 0.11 ; 0.16 ] [ 0 ; 0.15 ] | 0.954 (0.006) | 0.778 (0.034) |
| | | *0.300* | *0.281* | | *210.97* | |
| Unemployment | 0.566 (0.045) | [ 0.487 ; 0.650 ] [ 0.491 ; 0.653 ] | 0.10 (0.038) | [ 0 ; 0.12 ] [ 0 ; 0.12 ] | 0.966 (0.005) | 0.863 (0.024) |
| | | *0.364* | *0.323* | | *470.17* | |

NOTE: Estimates obtained with Algorithm 1. The standard errors and confidence intervals at the 95% percent level are estimated from 1,000 bootstrap replications clustered at the forecaster level (104 clusters). For robustness, I report the estimated confidence intervals using the bias corrected and percentiles methods. The last two columns respectively report the share of revisers among all forecasts and among zero-revisions only. In order to asses the gain from accounting for the extensive margin, I report the R-squared for the constrained model ($(R^2)^c$) when $\alpha = 0$, that is for the model with an intensive margin only. I also compute the associated $F$-stat associated to the constrained and unconstrained models. The sampling distribution of this $F$-stat under $H_0$ is however not standard.

an intensive margin only ($\alpha = 0$), the $R$-squared are respectively 6.5%, 6.9% and 12.6% larger for inflation, GDP growth and unemployment. As a rule-of-thumb, I also report the associated $F$-statistics. These statistics are also large and would lead to reject the constrained model in favor of the two-margins model at the one percent level if the sampling distribution under the null was the Fisher distribution.[122] Such high values for the $F$-stats and increases in the $R$-squared are nevertheless strong evidence that the model with two margins is a better representation of the data generating process than a model with noisy information only.

Table 4.3 reports the panel and consensus forecast estimates introduced in Section 4.2.2. It is an analog table to Table 4.3 but now computed with respect to the identification of the participation at the extensive margin that allows for forecast rounding. Given this new identification, the panel and consensus forecasts estimates for the average participation at the extensive margin are now very similar. This is a good signal of the internal consistency of the identification strategy, though it cannot be interpreted as a test for whether these values are correct. Furthermore, the estimated confidence intervals obtained for the consensus forecast estimates lead to reject the hypothesis of no frictions at the extensive margin .[123]

Regarding the estimates at the intensive margin, they are barely affected. We find that information rigidities are slightly higher at the intensive margin with this new identification of revised forecasts. However, the effect on the estimates is small (though generally statistically significant) and the conclusions drawn in Section 4.2.3 for the intensive margin continues to hold.

In conclusion, forecast rounding explains the vast majority of the observed zero revisions. Consequently, zero revisions should not be used as a proxy for Sticky information (or any other adjustment friction in forecasts). However, the data indicates that there are small, yet not negligible, frictions at the extensive margin. Accounting for the latter

[122] This is, however, unlikely to be the correct sampling distribution for this test and it would require more investigation to derive the correct sampling distribution.

[123] At the at the 1% level for GDP growth and 5% level for inflation using normal based confidence intervals, and at the 1% level for GDP growth and unemployment and 5% level for inflation using percentiles based confidence intervals.

Table 4.3: Information Rigidities for Each Margin with Rounding

| | **Panel Estimates** | | | | **Consensus Forecasts** | | | |
|---|---|---|---|---|---|---|---|---|
| | Intensive $k_h$ | | Extensive $\overline{\lambda_h}$ | | | Revisions $\beta_1 = (1-k_h)$ | Pool $\beta_2 = \lambda_h(1-k_h)$ | Ratio $\hat{\lambda}_h = \beta_2/\beta_1$ |
| | coef. | s.e. | coef. | s.e. | | | | |
| | | | | | ***Inflation*** | | | |
| Horizon | | | | | | | | |
| 0 | 0.861*** | 0.037 | 0.988 | 0.004 | coef. | 0.500 | 0.482 | **0.963** |
| 1 | 0.773*** | 0.018 | 0.966 | 0.006 | s.e. | 0.137 | 0.135 | 0.017 |
| 2 | 0.704*** | 0.049 | 0.977 | 0.005 | pval | 0.000 | 0.000 | 0.000 |
| 3 | 0.547*** | 0.021 | 0.958 | 0.006 | 95ci (low) | 0.231 | 0.218 | 0.930 |
| 4 | 0.524*** | 0.055 | 0.954 | 0.007 | 95ci (up) | 0.769 | 0.746 | 0.996 |
| 5 | 0.495*** | 0.054 | 0.945 | 0.007 | N | 162 | 162 | 162 |
| 6 | 0.423*** | 0.040 | 0.942 | 0.007 | | | | |
| 7 | 0.405*** | 0.025 | 0.951 | 0.007 | | | | |
| Sample Av. | 0.597*** | 0.019 | **0.961** | 0.002 | | | | |
| $(N, R^2)$ | 7,401 | 0.352 | 7,705 | . | | | | |
| | | | | | ***GDP Growth*** | | | |
| Horizon | | | | | | | | |
| 0 | 0.768*** | 0.032 | 0.986 | 0.004 | coef. | 0.678 | 0.656 | **0.968** |
| 1 | 0.660*** | 0.066 | 0.962 | 0.006 | s.e. | 0.110 | 0.109 | 0.010 |
| 2 | 0.561*** | 0.030 | 0.958 | 0.006 | pval | 0.000 | 0.000 | 0.000 |
| 3 | 0.476*** | 0.037 | 0.960 | 0.006 | 95ci (low) | 0.463 | 0.443 | 0.948 |
| 4 | 0.466*** | 0.059 | 0.950 | 0.007 | 95ci (up) | 0.894 | 0.870 | 0.987 |
| 5 | 0.439*** | 0.057 | 0.933 | 0.008 | N | 162 | 162 | 162 |
| 6 | 0.447*** | 0.065 | 0.945 | 0.007 | | | | |
| 7 | 0.510*** | 0.052 | 0.937 | 0.008 | | | | |
| Sample Av. | 0.543*** | 0.020 | **0.954** | 0.002 | | | | |
| $(N, R^2)$ | 7,255 | 0.318 | 7,602 | . | | | | |
| | | | | | ***Unemployment*** | | | |
| Horizon | | | | | | | | |
| 0 | 0.658*** | 0.103 | 0.984 | 0.004 | coef. | 0.489 | 0.465 | **0.952** |
| 1 | 0.558*** | 0.099 | 0.978 | 0.005 | s.e. | 0.169 | 0.170 | 0.062 |
| 2 | 0.624*** | 0.062 | 0.973 | 0.005 | pval | 0.004 | 0.006 | 0.000 |
| 3 | 0.640*** | 0.047 | 0.971 | 0.005 | 95ci (low) | 0.157 | 0.132 | 0.831 |
| 4 | 0.545*** | 0.040 | 0.968 | 0.006 | 95ci (up) | 0.820 | 0.799 | 1.074 |
| 5 | 0.488*** | 0.045 | 0.953 | 0.007 | N | 162 | 162 | 162 |
| 6 | 0.471*** | 0.039 | 0.960 | 0.006 | | | | |
| 7 | 0.531*** | 0.051 | 0.939 | 0.008 | | | | |
| Sample Av. | 0.566*** | 0.045 | **0.966** | 0.002 | | | | |
| $(N, R^2)$ | 7,129 | 0.380 | 7,377 | . | | | | |

NOTE: This is an analogous table to Table 4.3. See Table 4.3 for information on the statistics reported here. The only difference is that the identification of the participation at the extensive margin now allows for forecast rounding. A forecast is revised whenever the predicted value for $\hat{u}_{i,t,h}$ in equation (4.13) is equal to one. Inference assumes that the identification of the participation at the extensive margin is certain. See Table 4.2 for the effects of the uncertainty introduced by the stochastic identification of revised forecasts.

does not only improves the model goodness-of-fit but is also key for the estimation of information frictions at the intensive margin. These latter are, by far, the main rigidity affecting forecast revisions in our model.

## 4.4 Time-varying information rigidities

In this section, we analyse the variation of information frictions over time using the panel structure of the data. More specifically, we derive time series for frictions at each margin for inflation, GDP growth and unemployment.

### 4.4.1 Time-Variation at the Intensive Margin

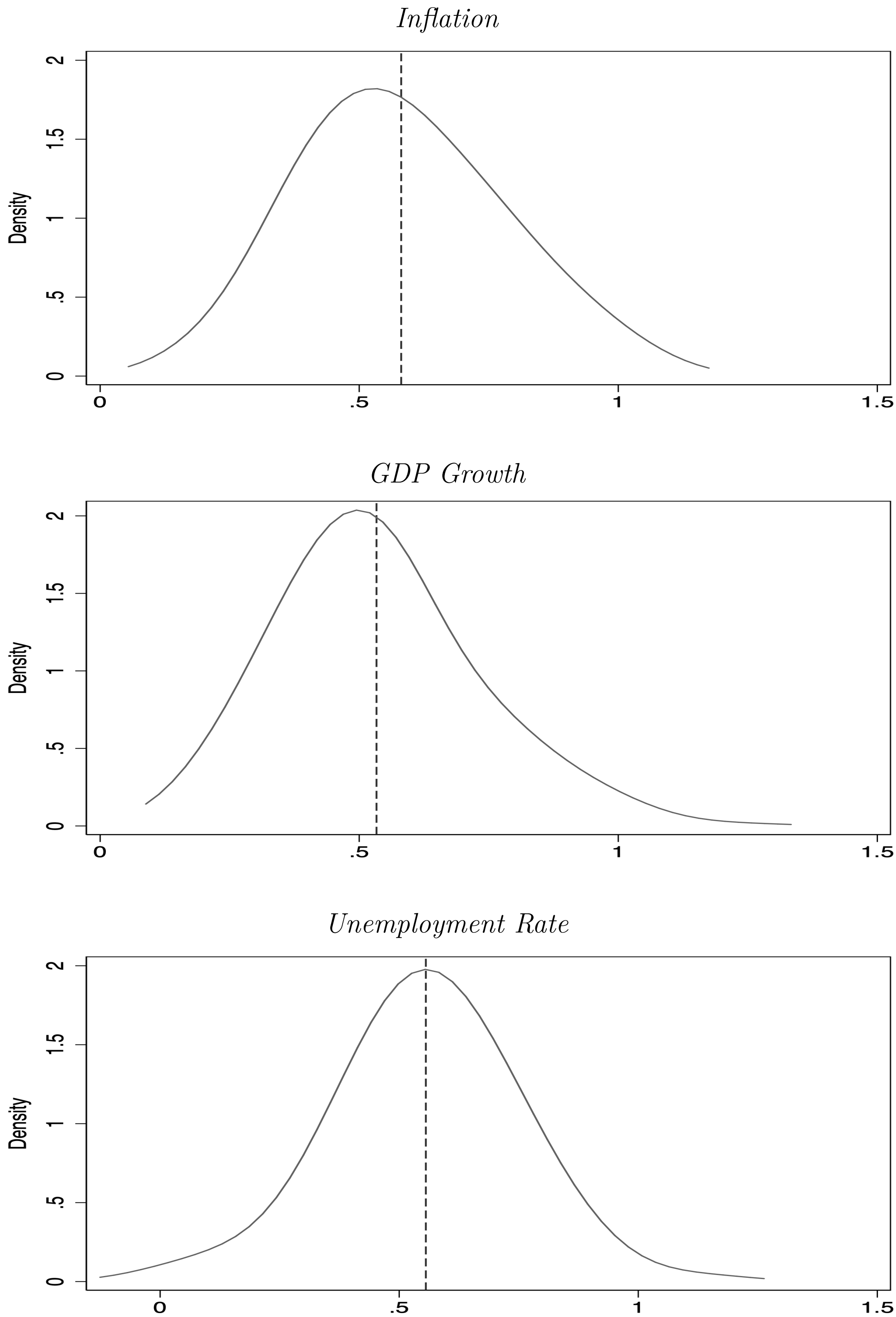
In order to analysis the time-variation of information rigidities at the intensive margin, I estimate equation (4.3) on the sample of revisions (as identified in the previous section) with time and horizon interaction terms. The estimated Kalman gains are thus the sum of a horizon and a time fixed effects. Because the estimation does not impose any constraint on the range of the estimates, a small proportion of the estimated Kalman gains are outside of the unit interval.

Figure 4.4 plots the distribution of the estimated Kalman gains. It shows that information acquisition is highly heterogenous across time, horizons and forecast types. Furthermore, the distribution for inflation and GDP growth are positively skewed, thus suggesting that forecasters are inattentive most of the time but, on rare occasions, are willing to incorporate more information and reach near full-information forecasts.

A sequential decomposition of the variance indicates that the time dimension explains respectively 48.34%, 71.49% and 87.46% of the model variance in the estimated Kalman gains for inflation, GDP growth and unemployment. That is, time variations in information acquisition at the intensive margin are significant and, potentially, more important than the heterogeneity induced by forecast horizons. Therefore, and unsurprisingly, we can reject the hypothesis of constant Kalman gains over time at the 1% level for all three variables.

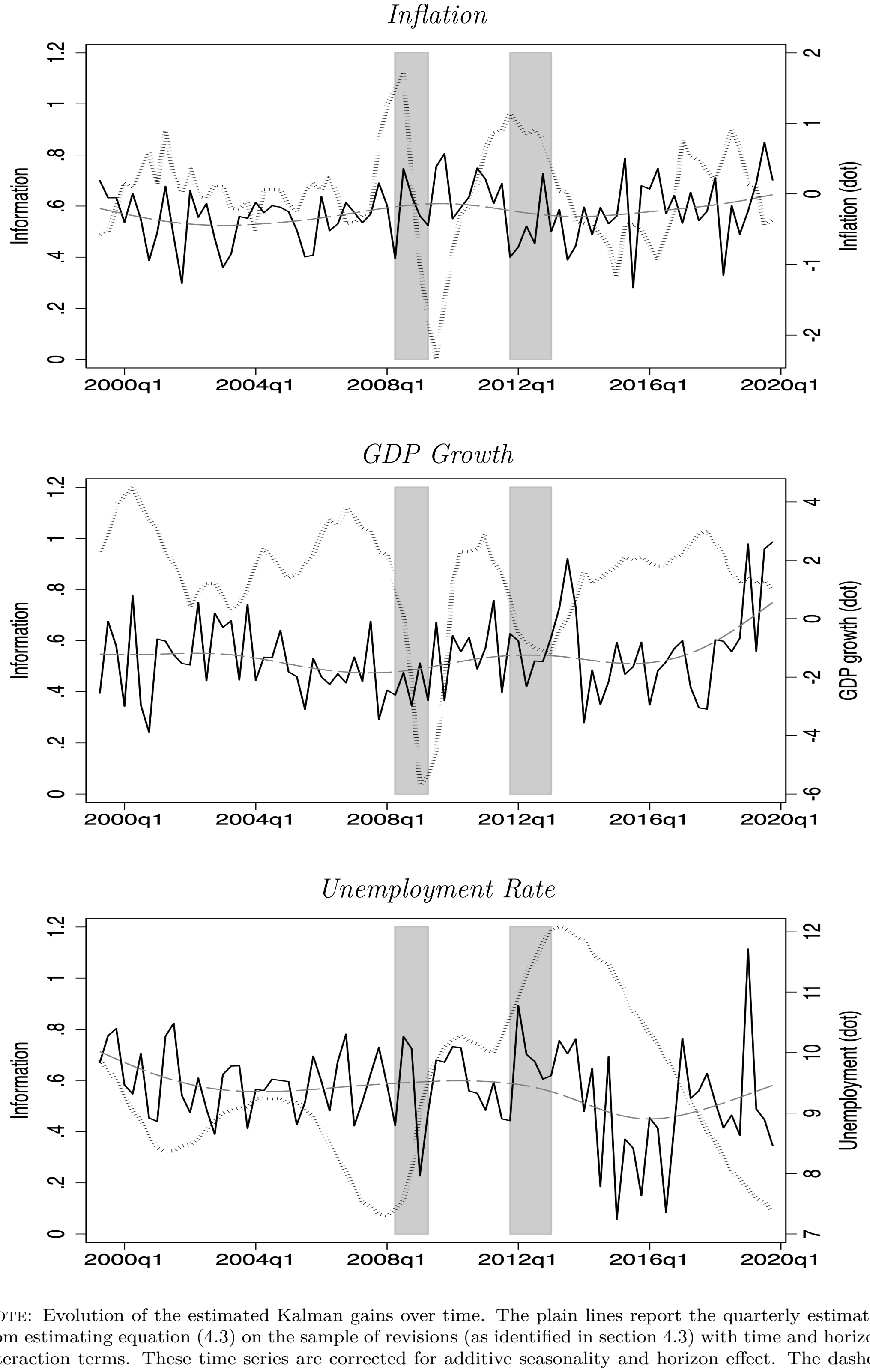
Controlling for the horizon effect, the time series for information acquisition display a

Figure 4.4: Distribution of the Estimated Kalman Gains




NOTE: Sample distribution of the estimated Kalman gains $k_{h,t}$ for inflation, GDP growth and unemployment. Gaussian kernel with a bandwidth equal to 0.1. The dashed vertical line is the sample average.

Figure 4.5: Evolution of the Kalman Gains over Time




Note: Evolution of the estimated Kalman gains over time. The plain lines report the quarterly estimates from estimating equation (4.3) on the sample of revisions (as identified in section 4.3) with time and horizon interaction terms. These time series are corrected for additive seasonality and horizon effect. The dashed lines report the trends estimated using HP filters. Finally, the dotted lines report the macroeconomic variable that is being forecasted. The grey areas are the recessions identified by the CEPR dating committee.

significant seasonality. Interestingly, the observed seasonality differ across the three time series. The seasonal peak in information acquisition is during the first quarter of each year for inflation and during the last quarter for unemployment. Regarding GDP growth, we do not observe a clear quarterly increase in information acquisition, but forecasters tend to be relatively more attentive during the second and third quarters of the year on average.

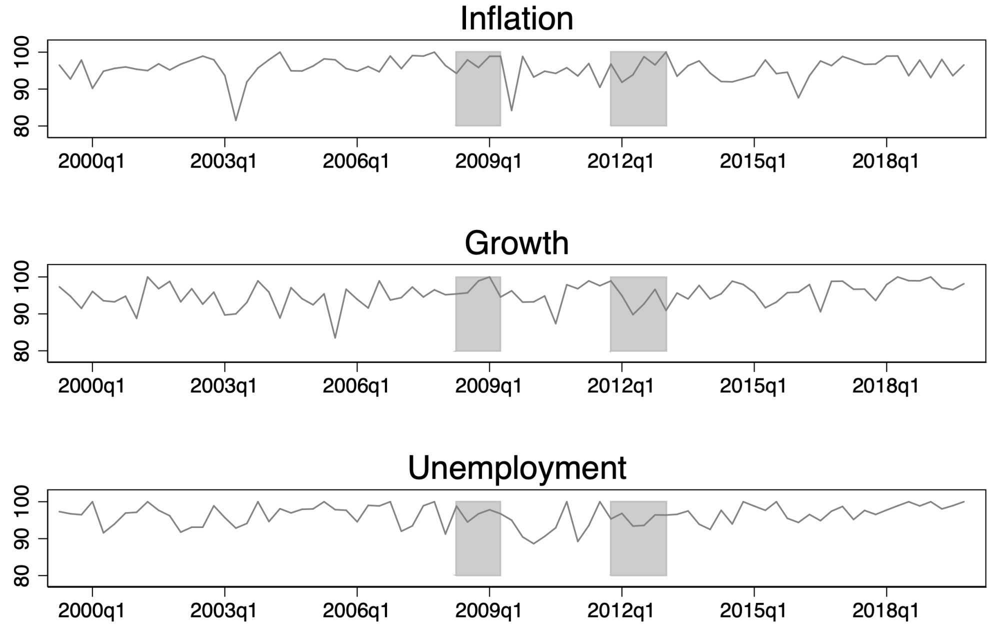
Figure 4.5 displays the evolution of the Kalman gains estimated for the panel of revising forecasters while controlling for the horizon effect and seasonality. The trend in information acquisition has remained relatively constant over the 1999-2019 period and we don't observe sudden structural change over this period. We nevertheless observe medium term variations that are common to all three series. Information acquisition has slowly decreased over the 2000-2008 period, a period of relative economic stability in the euro area. It started to increase afterward and remained stable in between the two most recent recessions. Since 2016, the trend in information acquisition seems to be on an increasing path again.

The panel structure of the survey of professional forecasters allows to estimate higher frequency variations in information acquisition. These changes in information acquisition at the business cycle frequency are significant and we observe large and prolonged deviations from the trend. For instance, information acquisition for inflation remained low during the first year of the 2012 recession, while it remained at a high level for a couple of years for unemployment. Notably, information acquisition has become more volatile in the post 2008 area. Using a one-sided test of variance equality, one favors the alternative of increased variance at, respectively, the 5, 10 and 1 percent levels for inflation, growth and unemployment.

### 4.4.2 Time-Variation at the Extensive Margin

Similarly, the estimation procedure describes in Algorithm 1 also allows to study the evolution of forecasters' participation at the extensive margin. Figure 4.6 reports the evolution of these information rigidities over time. While the average share of revising forecasters is high, we observe important deviations from this average over time with periods when all forecasters update their forecasts and periods when almost 20% of forecasters do not change their forecasts. This was for instance the case for inflation forecasts during the second quarter of 2003.

Figure 4.6: Evolution of the Share of Revised Forecasts




Note: Shares of revised forecast as identified from Algorithm 1. The grey areas are the recessions identified by the CEPR dating committee.

## 4.5 Forecasts stickiness

It is well-known that information frictions are a main driver of the persistence of macroeconomic variables (Sims, 2003; Mankiw and Reis, 2006; Maćkowiak and Wiederholt, 2015; Fuhrer, 2017; Carroll et al., 2020) and, in particular, professional forecasters' information frictions (Carroll, 2003). Therefore, assessing the evolution of professional forecasters consensus forecast stickiness for three major economic aggregates provides a new and potentially important ingredient to the empirical macro literature focusing on the business cycle dynamics. In this section, I illustrate how the estimates from each margin map into a single statistic that captures the consensus forecast stickiness under a plausible approximation. I then derive the evolution of this statistic at a quarterly frequency. In order to illustrate that the time series we constructed are not the result of mere randomness, I analyze their evolution during recessions and reach the well-documented conclusion that information rigidities globally decrease during these periods (Andrade and Le Bihan, 2013; Coibion and Gorodnichenko, 2015). This average evolution however hides important heterogeneity across forecast types.

### 4.5.1 Consensus Forecast Stickiness

Averaging equation (4.1) across revising forecasters and accounting for the extensive margin, the consensus (average) forecast is

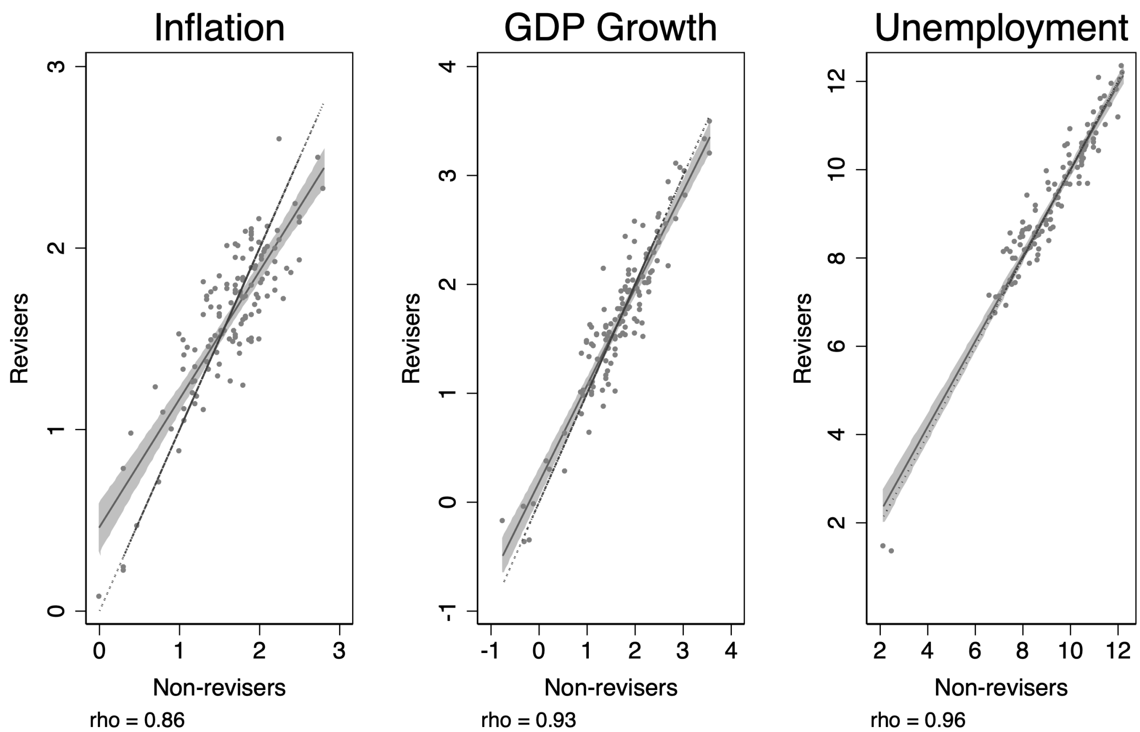
$$\bar{F}_{t,h} = \lambda_{t,h}(1-k_{h,t})\bar{F}^{u(t)}_{t-1,h+1} + (1-\lambda_{t,h})\bar{F}^{nu(t)}_{t-1,h+1} + \lambda_{t,h}k_{h,t}\bar{s}_{t,h} \tag{4.17}$$

where $\bar{s}_{t,h}$ is the average signal of revising forecasters at time $t$ for horizon $h$. There are two sources of stickiness in the consensus forecast. The first one arises as a consequence of the noisy information acquired by revising forecasters who slowly incorporate new information and anchor their forecast on their previous forecast. The second is a direct consequence of the extensive margin. Non-revising forecasters simply report the same forecast. Because the latter forecasters fully anchor their forecasts, even a small proportion of non-revising forecasters can have a noticeable impact on the overall stickiness of the consensus forecast.

The dynamics of the stickiness parameters from equation (4.17) have already been discussed in section 4.4. The consensus forecast stickiness however collapses to a single parameter $(1-\lambda_{t,h}k_{h,t})$ when there is no difference between the previous average forecast of revising and non-revising forecasters.[124] To asses whether this approximation is supported by the data, Figure 4.7 plots these two averages for each forecast type. When the approximation is exact, all data points belong to the dotted 45 degree line. Estimating a regression line, we have evidence that this approximation is likely to be correct in expectation for the unemployment rate. The 45 degree line belongs to the 95% confidence

[124]This would for instance be the case when revising forecasters are randomly and independently selected. It could also arise with an endogenous selection at the extensive margin.

Figure 4.7: Previous Average Forecast of Revising and Non-revising Forecasters




NOTE: Plots of the previous average forecast of revising forecasters ($\bar{F}^{u(t)}_{t-1,h+1}$) against that of non-revising forecasters ($\bar{F}^{nu(t)}_{t-1,h+1}$). The dotted line is the 45 degree line. The plain line is a regression line with its 95% confidence interval. Rho is the correlation coefficient. For readability, two outliers around -4 have been dropped for the middle panel (GDP growth).

interval associated to the regression line. This is also generally true for GDP growth and, in particular, the approximation is accurate where most data points are concentrated. Regarding inflation, the OLS prediction line is significantly different from the 45 degree line thus suggesting that the approximation is unlikely to hold at the population level. Nevertheless, the 45 degree line remains a decent approximation of the regression line.

### 4.5.2 Contribution of Each Margin

We can asses the contribution of each margin to the consensus forecast stickiness. Table 4.4 decomposes the first two moments of the distribution $\lambda_t \times k_t$. The average stickiness is mostly driven by information rigidities at the intensive margin, and the average stickiness increases by 4-5% when accounting for the extensive margin. Regarding the evolution of stickiness over time, the impact of the extensive margin becomes negligible. That is, the extensive margin plays a small role on the level of consensus forecast stickiness and has a negligible effect on its evolution over time.

Table 4.4: Contributions to Consensus Forecast Stickiness

| | $E(\lambda k)$ | $=$ | $cov(\lambda, k)$ | $+$ | $E(k)$ | $\times$ | $E(\lambda)$ |
|---|---|---|---|---|---|---|---|
| Inflation | 0.551 | | 0.000 | | 0.573 | | 0.961 |
| GDP growth | 0.514 | | 0.001 | | 0.537 | | 0.955 |
| Unemployment | 0.542 | | -0.000 | | 0.561 | | 0.966 |
| | $V(\lambda k)$ | $\simeq$ | $V(\lambda)V(k)$ | $+$ | $V(k)E(\lambda)^2$ | $+$ | $V(\lambda)E(k)^2$ |
| Inflation | 0.013 | | 0.000 | | 0.013 | | 0.000 |
| GDP growth | 0.023 | | 0.000 | | 0.022 | | 0.000 |
| Unemployment | 0.029 | | 0.000 | | 0.029 | | 0.000 |

Figure 4.11 in the Appendix displays the evolution of the consensus forecast stickiness over time and compares it with that obtained from the intensive margin only. As can be seen, the two time-series are very similar.

### 4.5.3 Forecast Stickiness during recessions

To asses the dynamics of information rigidities in the euro area in the aftermath of a recession, I follow the methodology in Coibion and Gorodnichenko (2015) and estimate the following specification:

$$\beta_t \quad = \quad \alpha + \sum_{j=0}^{J} \phi_j \mathbf{1}_{t-j}^{REC} + error_t \tag{4.18}$$

where $\mathbf{1}_{t-j}^{REC}$ is a dummy variable equal to one in the $j$th quarter of each recession. The estimates for $\phi_j$ therefore provide an impulse response for $\beta_t$ during recessions. In order to smooth this impulse response function, I fit a polynomial distributed lag model with a polynomial order equal to 4, as in Coibion and Gorodnichenko (2015). The lag order $J$ is restricted to belong in $[12, 13, \ldots, 20]$ and is selected to minimize the Akaike information criterion (AIC). This methodology is applied separately for each forecast type, and for each time series for information rigidities at the intensive margin alone, extensive margin alone, and overall information rigidities.

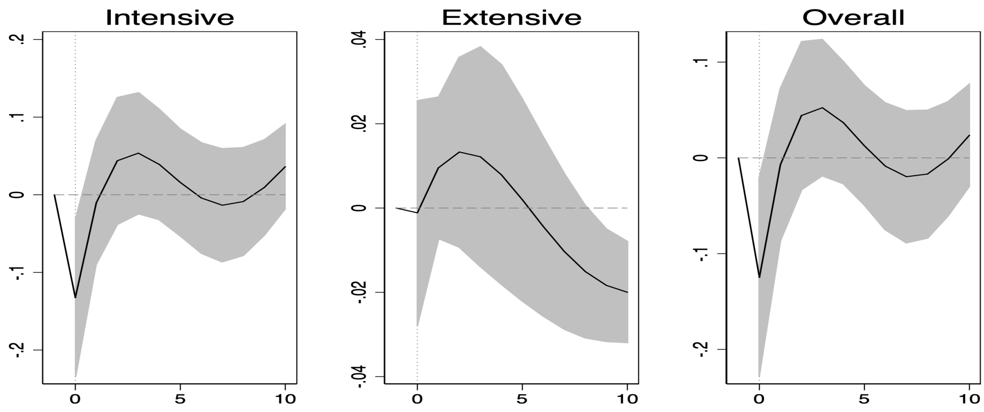
The estimated IRFs are reported in Figure 4.8 along with a confidence interval at the 90% level. Overall information rigidities are found to be state-dependent. For each forecast type, the dynamics of overall information rigidities is significantly affected when entering in a recession. These dynamics are however very different from one macroeconomic variable to another. In the aftermath of a recession, forecasters tend to pay significantly less attention to inflation. This increase in inflation information rigidities is only significant on impact and the dynamic effect barely persistent. For GDP growth, forecasters also tend to become less attentive when entering in a recession and this decrease persists for a year. In opposition, we observe a relatively long lasting increase in attention for unemployment forecasts during recessions. That is, while we observe a significant state dependent dynamics for each forecast type, these dynamics are highly heterogenous.

The observed decrease in information acquisition for inflation and GDP growth are counter-intuitive. Indeed, there is a common agreement in the literature that forecasters' information rigidities decrease during recessions (e.g. Loungani et al. (2013) and Coibion and Gorodnichenko (2015)). In particular, Coibion and Gorodnichenko (2015) reach this conclusion by pooling over all available forecast types in the US survey of professional forecasters (inflation, GDP growth, unemployment, industrial production and housing starts). As a comparison exercice, I define a new time series that averages (with equal weights) information rigidities for all three macroeconomic variables available from the European survey of professional forecasters. Using this novel time series, Figure 4.11 in appendix plots the estimated dynamics in the aftermath of a recession. As can be seen, we also conclude that, irrespective of the forecast type, information rigidities decrease during recessions.

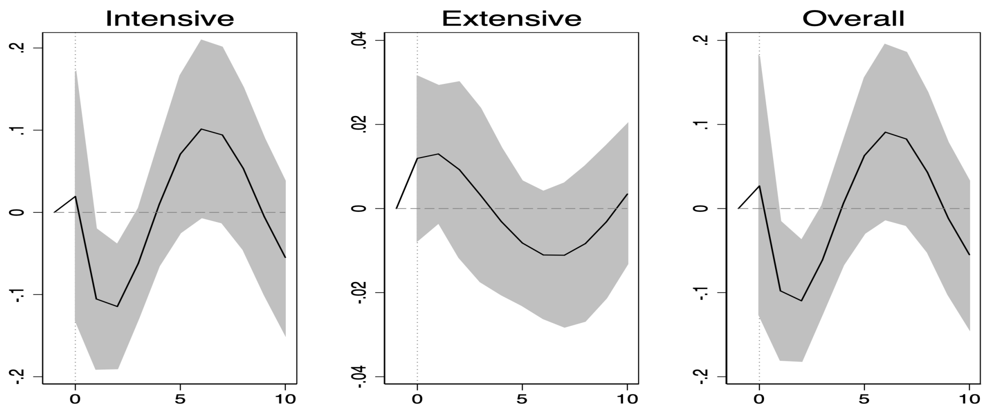
Decomposing the dynamics of overall information rigidities between the extensive and intensive margins, we again find that the former dynamics closely follows that at the intensive margin and is barely affected by the dynamics at the extensive margin.

Figure 4.8: Dynamics of Information Rigidities during Recessions

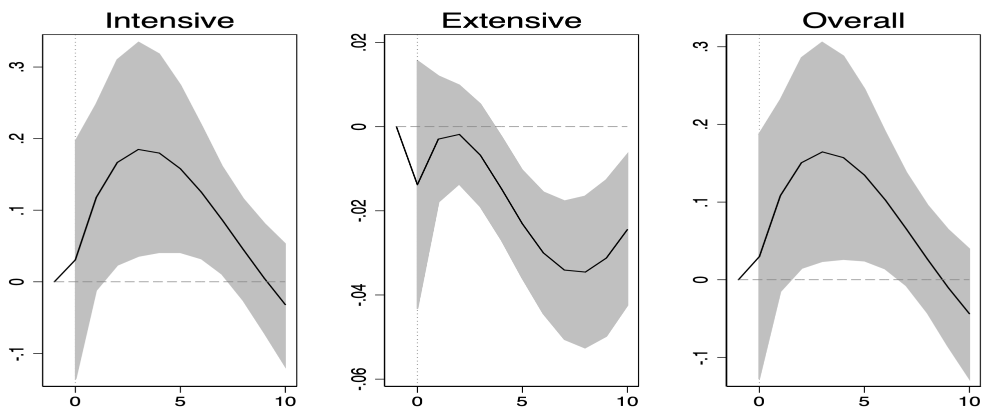
*Inflation*




*GDP Growth*




*Unemployment Rate*




NOTE: Impulse response functions for forecasters' attention parameters at the intensive margin alone $k_t$ (left panels), extensive margin alone $\lambda_t$ (center panels) and overall $\lambda_t k_t$ (right panels). The impulse responses are estimated using a polynomial distributed lag model for equation (4.18). The recession dates are defined according to the CEPR dating committee. The number of lags is chosen to minimize the AIC and the order of the polynomial is 4. Confidence intervals at the 90% level.

## 4.6 Conclusion

This paper estimates a two margin forecast formation process that allows for forecast rounding using individual data from the European survey of professional forecasters. It finds that information rigidities at the intensive margin are the main driver of the observed stickiness in consensus forecasts. Importantly, the paper proposes a novel approach in order to estimate quarterly variations in information frictions. While the trend in information rigidities has been globally constant for the last two decades, there is significant fluctuations around this trend. Most of the time, forecasters are relatively inattentive to new information releases. However, there are periods when they reach near full-information.

In future work, the time series derived in this paper can be used to analyze the determinants behind the variations in information frictions. The theoretical literature on rationally inattentive agents identified a few potential factors that should affect forecasters' attention. Among these factors, we should observe an increased level of attention when the variance of the economic variable being forecasted increases (Sims, 2003; Reis, 2006a). Similarly, one should expect the level of economic and political uncertainty to be positively correlated with attention. However, it is well-known that uncertainty has been increasing in Europe over the last decade (Baker et al., 2016). The time series developed in this paper however do not display an increasing pattern for information frictions during this same period. This result is puzzling and prompt for a better understanding of the determinants that prompt economic agents to become more attentive.

Furthermore, the paper proposes a simple way to incorporate the observed variations in the excess smoothness of average expectations into otherwise standard macroeconomic models. Understanding how these variations affect the dynamics of macroeconomic outputs and the transmission of monetary and fiscal policies is an important avenue for future work.

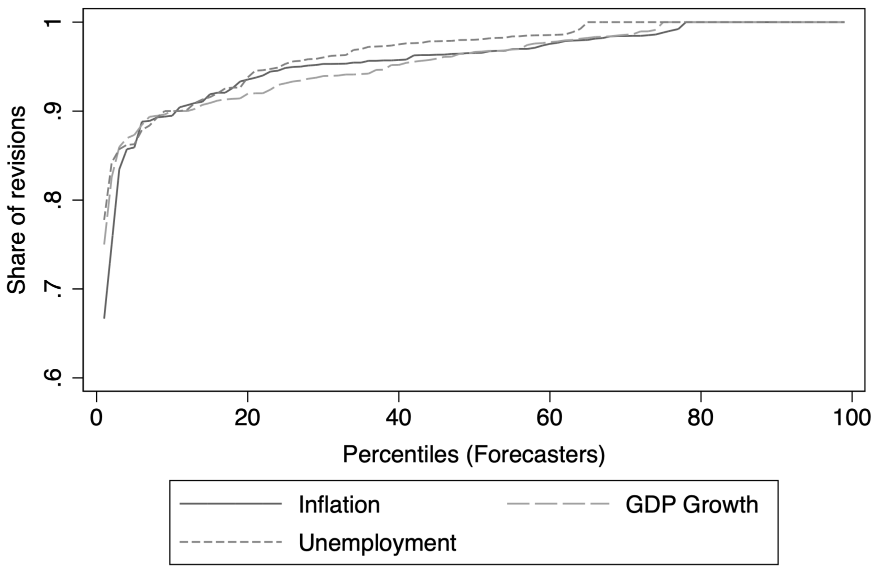
## 4.7 Complementary figures and tables

Table 4.5: Individual forecast errors and revisions

| | Pooled | | | One year ahead | | |
|---|---|---|---|---|---|---|
| | | 2-way | D. K. | | 2-way | D. K. |
| | coef. | s.e.<br>p-value | s.e.<br>p-value | coef. | s.e.<br>p-value | s.e.<br>p-value |
| Inflation | -0.298 | 0.208<br>0.152 | 0.221<br>0.180 | 0.051 | 0.211<br>0.809 | 0.204<br>0.803 |
| | $N$ | 6,425 | | $N$ | 856 | |
| GDP growth | -0.049 | 0.234<br>0.834 | 0.257<br>0.848 | 0.240 | 0.219<br>0.273 | 0.151<br>0.114 |
| | $N$ | 6,349 | | $N$ | 852 | |
| Unemployment | -0.532 | 0.265<br>0.045 | 0.289<br>0.069 | -0.825 | 0.133<br>0.000 | 0.132<br>0.000 |
| | $N$ | 6,497 | | $N$ | 883 | |

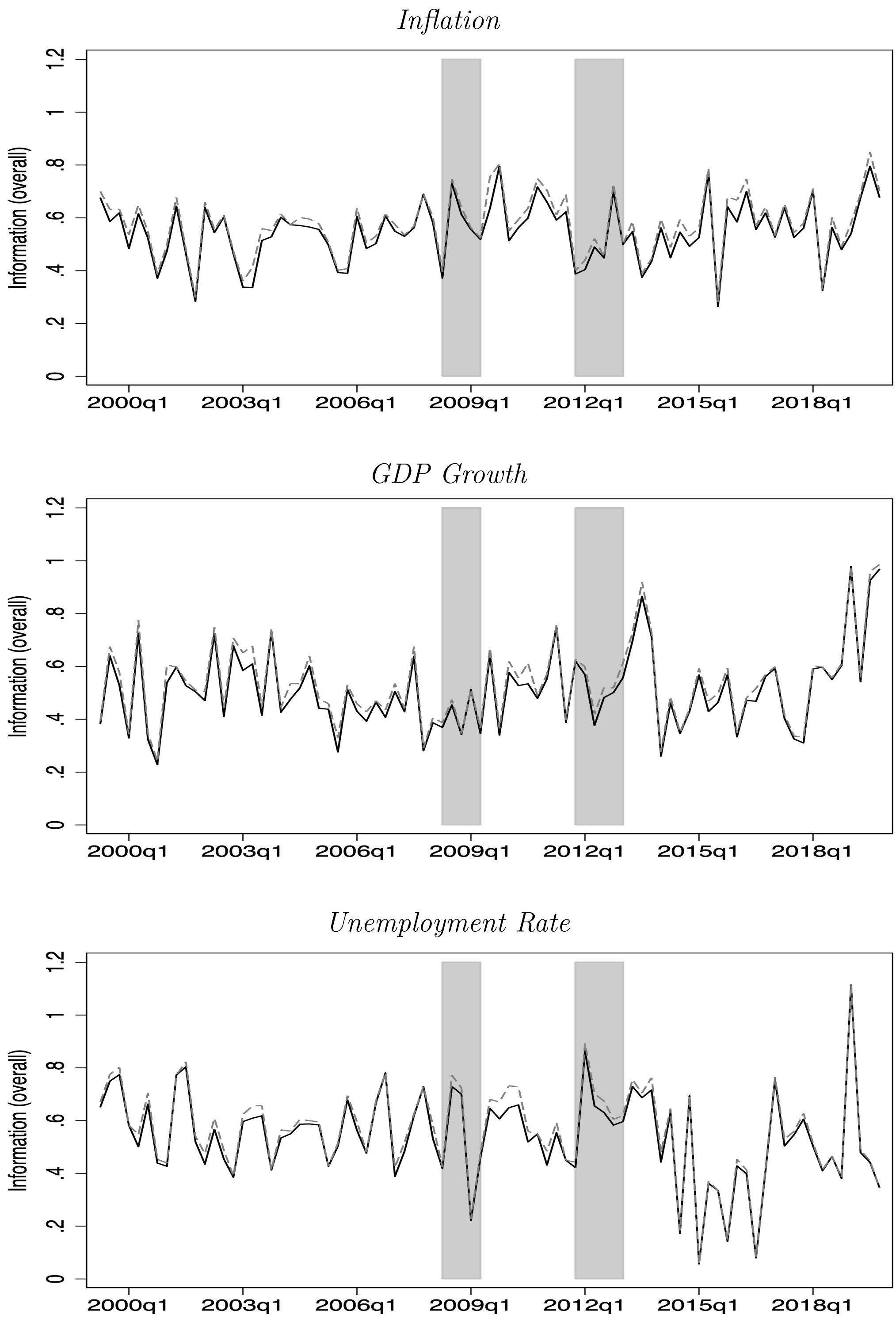
NOTE: Results from a forecaster fixed-effect regression of the forecast error on the forecast revision at the individual level (on the sample of participants at the intensive margin) with robust two-way cluster standard errors at the individual and date levels (2-way) as in Broer and Kohlhas (2018); Bordalo et al. (2020) or Driscoll and Kraay robust standard errors (D.K.). The pooled regression is for all forecast horizons and the right panel is for one year ahead forecasts. The methodology follows that in Broer and Kohlhas (2018); Bordalo et al. (2020) and, as they do, we measure forecast errors from the first vintage of actual data. The only methodological difference is that we rely on fixed-event calendar forecasts, while they rely on fixed-horizon (one year ahead) rolling forecasts.

Figure 4.9: Shares of Revisions per Forecaster



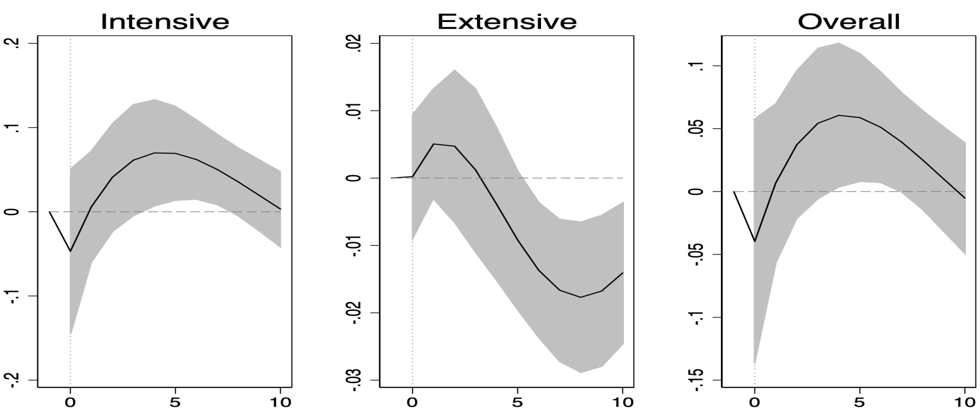

NOTE: Distribution of the forecasters' share of unrevised forecasts that may not be explained by rounding for inflation, GDP growth and unemployment. The identification of these non-revised forecasts follows from Algorithm 1.

Figure 4.10: Dynamics of Overall Information Rigidities ($\lambda_t k_t$)




NOTE: Evolution of overall information rigidities over time under the approximation $\bar{F}^{u(t)}_{t-1,h+1} \simeq \bar{F}^{nu(t)}_{t-1,h+1}$. The plain lines report the quarterly estimates. The dashed lines report the estimates at the intensive margin only. The grey areas are the recessions identified by the CEPR dating committee.

Figure 4.11: Dynamics of Information Rigidities (all forecast types)




NOTE: Impulse response functions for forecasters' attention parameters at the intensive margin alone $k_t$ (left panel), extensive margin alone $\lambda_t$ (center panel) and overall $\lambda_t k_t$ (right panel) for the time series that averages the attention parameters for inflation, GDP growth and unemployment (with equal weights). The impulse responses are estimated using a polynomial distributed lag model for equation (4.18). The recession dates are defined according to the CEPR dating committee. The number of lags is chosen to minimize the AIC and the order of the polynomial is 4.

## 4.8 Derivation of the within transformation (4.3)

Let $\eta_{h,t} \equiv \bar{F}^{u(t)}_{t,h} - \bar{x}_{t+h|t}$ be the error between the consensus forecast among revising forecasters and the full-information forecast. This error may result from a systematic biais in forecasts, a sampling error, the effect of common information, etc. Introducing this expression in equation (4.2), we get

$$r_{i,t,h} = k_{t,h}(\bar{F}^{u(t)}_{t,h} - \eta_{t,h} - F_{i,t-1,h+1}) + k_{t,h}\omega_{i,t,h} \tag{4.19}$$

where $\eta_{t,h}$ is a fixed effect correlated with $\bar{F}^{u(t)}_{t,h}$. To eliminate for this fixed effect, define the sample average revision among revising forecasters $\bar{r}^{u(t)}_{t,h}$. Then, averaging equation (4.19), we have

$$\bar{r}^{u(t)}_{t,h} = k_{t,h}(\bar{F}^{u(t)}_{t,h} - \eta_{t,h} - \bar{F}^{u(t)}_{t-1,h+1}) + k_{t,h}\bar{\omega}^{u(t)}_{t,h} \tag{4.20}$$

Therefore,

$$\begin{aligned} r_{i,t,h} - \bar{r}^{u(t)}_{t,h} &= k_{t,h}(\bar{F}^{u(t)}_{t,h} - \eta_{t,h} - F_{i,t-1,h+1}) + k_{t,h}\omega_{i,t,h} - k_{t,h}(\bar{F}^{u(t)}_{t,h} - \eta_{t,h} - \bar{F}^{u(t)}_{t-1,h+1}) - k_{t,h}\bar{\omega}^{u(t)}_{t,h} \\ &= k_{t,h}(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}) + k_{t,h}(\omega_{i,t,h} - \bar{\omega}^{u(t)}_{t,h}) \end{aligned} \tag{4.21}$$

**Bias from the pooled regression $\hat{k}^{pooled}_h$**

First, note that,

$$\begin{aligned} & Cov\Big(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}, r_{i,t,h} - \bar{r}_{t,h}\Big) \\ = \; & Cov(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}, r_{i,t,h}\Big) - Cov\Big(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}, \bar{r}_{t,h}\Big) \\ = \; & \lambda_h Cov^{u(t)}(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}, r_{i,t,h}\Big) - \lambda_h Cov\Big(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}, \bar{r}^{u(t)}_{t,h}\Big) \end{aligned} \tag{4.22}$$

Now, instead of specification (4.21), assume that we directly estimate

$$r_{i,t,h} - \bar{r}_{t,h} \;=\; a_h + k^{pooled}_h(\bar{F}^{u(t)}_{t-1,h+1} - F_{i,t-1,h+1}) + \varepsilon_{i,t,h} \tag{4.23}$$

on all observations. Then, it hold for all $h$ (averages and covariances are computed for observations with horizon $h$):

$$\hat{k}_h^{pooled} = \frac{Cov\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, r_{i,t,h} - \bar{r}_{t,h}\Big)}{Var\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}\Big)} \tag{4.24}$$

We can relate the pooled estimate $\hat{k}_h^{pooled}$ and the estimate $\hat{k}_h$ from the subsample of revised forecast by first realizing that

$$\begin{aligned}
& Cov\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, r_{i,t,h} - \bar{r}_{t,h}\Big) \\
= \ & \lambda_h Cov^{u(t)}\big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, r_{i,t,h}\Big) - \lambda_h Cov\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, \bar{r}_{t,h}^{u(t)}\Big) \qquad (4.25) \\
= \ & \lambda_h Cov^{u(t)}\big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, r_{i,t,h} - \bar{r}_{t,h}^{u(t)}\ \big) \\
& \quad -\lambda_h \Delta Cov\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, \bar{r}_{t,h}^{u(t)}\Big) \qquad (4.26)
\end{aligned}$$

with $\Delta Cov\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, \bar{r}_{t,h}^{u(t)}\Big) = Cov\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, \bar{r}_{t,h}^{u(t)}\Big) - Cov^{u(t)}\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, \bar{r}_{t,h}^{u(t)}\Big)$ the difference between the covariances obtained on each sample. Then, using equations (4.24) and (4.26), we get

$$\begin{aligned}
\hat{k}_h^{pooled} &= \lambda_h \frac{Cov^{u(t)}\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, r_{i,t,h} - \bar{r}_{t,h}^{u(t)}\ \Big) - Cov\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, \bar{r}_{t,h}^{u(t)}\Big)}{Var\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}\Big)} \\
&= \frac{\lambda_h}{Var\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}\Big)} \left(\hat{k}_h\ Var^{u(t)}\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}\Big) - \Delta Cov\right) \qquad (4.27)
\end{aligned}$$

with

$$\hat{k}_h = \frac{Cov^{u(t)}\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}, r_{i,t,h} - \bar{r}_{t,h}\Big)}{Var^{u(t)}\Big(\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}\Big)} \tag{4.28}$$

the estimate obtained from regressing (4.21). This formula simplifies when the selection at the extensive margin is independent from $\bar{F}_{t-1,h+1}^{u(t)} - F_{i,t-1,h+1}$. In this special case, the pooled estimate is equal to

$$\hat{k}_h^{pooled} = \lambda_h \hat{k}_h$$

It is an analogous formulation to the approximation made in Section 4.5.1. In particular, it shows that pooled regressions may not assert the relative contribution of each margin but estimate a byproduct of both.

# Bibliography

Aalen, O. O., Gjessing, H. K., et al. (2001). Understanding the shape of the hazard rate: A process point of view (with comments and a rejoinder by the authors). *Statistical Science*, 16(1):1–22.

Abe, N. and Ueno, Y. (2015). Measuring Inflation Expectations: Consumers' Heterogeneity and Nonlinearity. Technical report, Research Center for Economic and Social Risks, Institute of Economic Research, Hitotsubashi University.

Aghion, P., Akcigit, U., Lequien, M., and Stantcheva, S. (2017). Tax simplicity and heterogeneous learning. *NBER Working Paper*.

Allcott, H., Lockwood, B., and Taubinsky, D. (2018). Ramsey strikes back: Optimal commodity tax and redistribution in the presence of salience effects. In *AEA Papers and Proceedings*, volume 108, pages 88–92.

Andrade, P., Crump, R. K., Eusepi, S., and Moench, E. (2016). Fundamental disagreement. *Journal of Monetary Economics*, 83:106–128.

Andrade, P. and Le Bihan, H. (2013). Inattentive Professional Forecasters. *Journal of Monetary Economics*, 60(8):967–982.

Angeletos, G.-M., Huo, Z., and Sastry, K. A. (2020). Imperfect macroeconomic expectations: Evidence and theory. Technical report, National Bureau of Economic Research.

Arellano, M., Blundell, R., and Bonhomme, S. (2017). Earnings and consumption dynamics: a nonlinear panel data framework. *Econometrica*, 85(3):693–734.

Armantier, O., Nelson, S., Topa, G., Van der Klaauw, W., and Zafar, B. (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics*, 98(3):503–523.

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

Atkinson, A. B. and Stiglitz, J. E. (1976). The design of tax structure: direct versus indirect taxation. *Journal of public Economics*, 6(1-2):55–75.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636.

Baker, S. R., McElroy, T. S., and Sheng, X. S. (2020). Expectation formation following large, unexpected shocks. *Review of Economics and Statistics*, 102(2):287–303.

Barro, R. J. and Gordon, D. B. (1983). A positive theory of monetary policy in a natural rate model. *Journal of Political Economy*, 91(4):589–610.

Bec, F., Boucekkine, R., and Jardet, C. (2017). Why are inflation forecasts sticky? theory and application to france and germany. *Bank de France Working Paper*.

Bernheim, B. D. and Taubinsky, D. (2018). Behavioral public economics. In *Handbook of Behavioral Economics*, volume 1, pages 381–516. Elsevier.

Beveridge, W. H. (1909). *Unemployment: A problem of industry*. Longmans, London.

Binder, C. (2017). Consumer forecast revisions: Is information really so sticky? *Economics Letters*, 161:112–115.

Blesse, S., Buhlmann, F., and Doerrenberg, P. (2019). Do people really want a simple tax system? Evidence on preferences towards income tax simplification. *Evidence on Preferences Towards Income Tax Simplification*, pages 19–058.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–82.

Bourguignon, F. and Spadaro, A. (2012). Tax–benefit revealed social preferences. *The Journal of Economic Inequality*, 10(1):75–108.

Branch, W. A. (2007). Sticky information and model uncertainty in survey data on inflation expectations. *Journal of Economic Dynamics and Control*, 31(1):245–276.

Broer, T. and Kohlhas, A. (2018). Forecaster (mis-) behavior. *CEPR Discussion Paper No. DP12898*.

Browning, M. and Collado, M. D. (2001). The response of expenditures to anticipated income changes: panel data estimates. *The American economic review*, 91(3):681–692.

Burstein, A. T. (2006). Inflation and output dynamics with state-dependent pricing decisions. *Journal of Monetary Economics*, 53(7):1235–1257.

Caballero, R. J. (1993). Durable goods: An explanation for their slow adjustment. *Journal of Political Economy*, 101(2):351–384.

Caballero, R. J. (1995). Near-rationality, heterogeneity, and aggregate consumption. *Journal of Money, Credit and Banking*, 27(1):29–48.

Caballero, R. J. and 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(3):383–398.

Caplin, A., Dean, M., and Leahy, J. (2019). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3):1061–1094.

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *the Quarterly Journal of economics*, 118(1):269–298.

Carroll, C. D., Crawley, E., Slacalek, J., Tokuoka, K., and White, M. N. (2020). Sticky expectations and consumption dynamics. *American economic journal: macroeconomics*, 12(3):40–76.

Cheremukhin, A. and Tutino, A. (2016). Information rigidities and asymmetric business cycles. *Journal of Economic Dynamics and Control*, 73:142–158.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ.*, 1(1):451–488.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5):1–33.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

Chetty, R. and Szeidl, A. (2016). Consumption commitments and habit formation. *Econometrica*, 84(2):855–890.

Cochrane, J. H. et al. (1989). The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives. *American Economic Review*, 79(3):319–337.

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

Coibion, O. and Gorodnichenko, Y. (2015). Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review*, 105(8):2644–78.

Conseil Constitutionnel (2005). Décision n. 2005-530 DC du 29 décembre 2005.

Conseil Constitutionnel (2012). Décision n. 2012-662 DC du 29 décembre 2012.

Crowe, C. (2010). Consensus forecasts and inefficient information aggregation. *IMF Working Papers*, pages 1–43.

Diamond, P. A. (1998). Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review*, pages 83–95.

Dovern, J. (2015). A multivariate analysis of forecast disagreement: Confronting models of disagreement with survey data. *European Economic Review*, 80:16–35.

Dovern, J., Fritsche, U., Loungani, P., and Tamirisa, N. (2015). Information rigidities: Comparing average and individual forecasts for a large international panel. *International Journal of Forecasting*, 31(1):144–154.

Dovern, J., Fritsche, U., and Slacalek, J. (2012). Disagreement among forecasters in g7 countries. *Review of Economics and Statistics*, 94(4):1081–1096.

Dräger, L. and Lamla, M. J. (2012). Updating Inflation Expectations: Evidence from Micro-data. *Economics Letters*, 117(3):807–810.

Farhi, E. and Gabaix, X. (2020). Optimal taxation with behavioral agents. *American Economic Review*, 110(1):298–336.

Farhi, E. and Werning, I. (2019). Monetary policy, bounded rationality, and incomplete markets. *American Economic Review*, 109(11):3887–3928.

Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics*, 124(3):969–1010.

Fochmann, M., Kiesewetter, D., Blaufus, K., Hundsdoerfer, J., and Weimann, J. (2010). Tax perception - an empirical survey. *Arqus Discussion Paper*.

Fochmann, M. and Weimann, J. (2013). The effects of tax salience and tax experience on individual work efforts in a framed field experiment. *FinanzArchiv: Public Finance Analysis*, 69(4):511–542.

Friedman, M. (1957). Introduction to" a theory of the consumption function". In *A theory of the consumption function*, pages 1–6. Princeton university press.

Fuhrer, J. (2017). Expectations as a source of macroeconomic persistence: Evidence from survey expectations in a dynamic macro model. *Journal of Monetary Economics*, 86:22–35.

Fuhrer, J. C. (2018). Intrinsic expectations persistence: Evidence from professional and household survey expectations.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics, Vol 2, edited by D Bernheim, S DellaVigna and D Laibson, Forthcoming*. Elsevier.

Gabaix, X. (2020). A behavioral new keynesian model. *American Economic Review*, 110(8):2271–2327.

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

Galí, J. (1993). Variability of durable and nondurable consumption: Evidence for six oecd countries. *The Review of Economics and Statistics*, pages 418–428.

Gerritsen, A. (2016). Optimal taxation when people do not maximize well-being. *Journal of Public Economics*, 144:122–139.

Giacomini, R., Skreta, V., and Turen, J. (2020). Heterogeneity, inattention, and bayesian updates. *American Economic Journal: Macroeconomics*, 12(1):282–309.

Gideon, M. (2017). Do individuals perceive income tax rates correctly? *Public Finance Review*, 45(1):97–117.

Goldin, J. (2015). Optimal tax salience. *Journal of Public Economics*, 131:115–123.

Gorodnichenko, Y. (2008). Endogenous Information, Menu Costs and Inflation Persistence. NBER Working Papers 14184, National Bureau of Economic Research, Inc.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of political economy*, 86(6):971–987.

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

Hoopes, J. L., Reck, D. H., and Slemrod, J. (2015). Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy*, 7(3):177–208.

Hoover, K. D. and Young, W. (2011). Rational expectations: Retrospect and prospect. *Economic Research Initiatives at Duke (ERID) Working Paper*.

Hsieh, C.-T. (2003). Do consumers react to anticipated income changes? evidence from the alaska permanent fund. *The American Economic Review*, 93(1):397–405.

Jacobs, B., Jongen, E. L., and Zoutman, F. T. (2017). Revealed social preferences of dutch political parties. *Journal of Public Economics*, 156:81–100.

Jacquet, L. and Lehmann, E. (2017). Optimal income taxation with composition effects. *Journal of the European Economic Association*.

Jappelli, T. and Pistaferri, L. (2010). The consumption response to income changes. *Annual Review of Economics*, 2(1):479–506.

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

Jaskowski, M. and van Dijk, D. (2016). First-passage-time in discrete time and intra-horizon risk measures. *Forthcoming*.

Keynes, J. M. (1936). The general theory of employment, interest and money.

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

Kleven, H. (2020). Sufficient statistics revisited.

Kueng, L. (2018). Excess sensitivity of high-income consumers. *The Quarterly Journal of Economics*, 133(4):1693–1751.

Kumar, S. and Jia, P. (2019). Financial crisis and persistence: evidence from sticky expectations consumption growth model. *Applied Economics*, 51(17):1799–1807.

Kydland, F. E. and Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85(3):473–491.

Kydland, F. E. and Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370.

Larsen, V. H., Thorsrud, L. A., and Zhulanova, J. (2020). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*.

Liebman, J. B. and Zeckhauser, R. J. (2004). Schmeduling.

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

Lockwood, B. B. and Weinzierl, M. (2016). Positive and normative judgments implicit in us tax policy, and the costs of unequal growth and recessions. *Journal of Monetary Economics*, 77:30–47.

Loungani, P., Stekler, H., and Tamirisa, N. (2013). Information rigidity in growth forecasts: some cross-country evidence. *International Journal of Forecasting*, 29(4):605–621.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of economic theory*, 4(2):103–124.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, volume 1, pages 19–46.

Luo, Y. (2008). Consumption Dynamics under Information Processing Constraints. *Review of Economic Dynamics*, 11(2):366–385.

Luo, Y., Nie, J., Wang, G., and 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. and Young, E. R. (2014). Signal Extraction and Rational Inattention. *Economic Inquiry*, 52(2):811–829.

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

Maćkowiak, B. and Wiederholt, M. (2015). Business Cycle Dynamics under Rational Inattention. *The Review of Economic Studies*, 82(4):1504–1532.

Mackowiak, B. A., Matejka, F., Wiederholt, M., et al. (2018). Survey: Rational inattention, a disciplined behavioral model. *CEPR Discussion Papers*.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of econometrics*, 18(1):83–114.

Mankiw, N. G. and Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Mankiw, N. G. and Reis, R. (2006). Pervasive stickiness. *American Economic Review*, 96(2):164–169.

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

Matějka, F. and Tabellini, G. (2017). Electoral competition with rationally inattentive voters. *SSRN 3070204*.

McCloskey, D. N. (1998). *The rhetoric of economics*. Univ of Wisconsin Press.

Meghir, C. and Pistaferri, L. (2011). Earnings, consumption and life cycle choices. In *Handbook of labor economics*, volume 4, pages 773–854. Elsevier.

Meyler, A. and Rubene, I. (2009). Results of a special questionnaire for participants in the ecb survey of professional forecasters (spf). *MPRA Paper No. 20751.*

Miller, B. M., Mumford, K., et al. (2015). The salience of complex tax changes: Evidence from the child and dependent care credit expansion. *National Tax Journal*, 68(3):477–510.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208.

Molin, A. and 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. and Hirche, S. (2012). An iterative algorithm for optimal event-triggered estimation. In *4th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS'12).*

Morrison, W. and Taubinsky, D. (2019). Rules of thumb and attention elasticities: Evidence from under-and overreaction to taxes. Technical report, National Bureau of Economic Research.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, pages 315–335.

Nimark, K. P. (2014). Man-Bites-Dog Business Cycles. *American Economic Review*, 104(8):2320–67.

Ocal, N. and Osborn, D. R. (2000). Business Cycle Non-Linearities in UK Consumption and Production. *Journal of Applied Econometrics*, 15(1):27–43.

Pesaran, M. H. and Weale, M. (2006). Survey expectations. *Handbook of economic forecasting*, 1:715–776.

Pfajfar, D. and Santoro, E. (2013). News on inflation and the epidemiology of inflation expectations. *Journal of Money, Credit and Banking*, 45(6):1045–1067.

Pigou, A. (1927). Industrial fluctuations.

Piketty, T. and Saez, E. (2013). Chapter 7 - optimal labor income taxation. In Alan J. Auerbach, Raj Chetty, M. F. and Saez, E., editors, *Handbook of Public Economics, vol. 5*, volume 5 of *Handbook of Public Economics*, pages 391 – 474. Elsevier.

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

Reck, D. (2016). Taxes and mistakes: What's in a sufficient statistic? *SSRN 2268617*.

Rees-Jones, A. and Taubinsky, D. (2020). Measuring "schmeduling". *The Review of Economic Studies*, 87(5):2399–2438.

Reis, R. (2006a). Inattentive consumers. *Journal of Monetary Economics*, 53(8):1761–1800.

Reis, R. (2006b). Inattentive producers. *The Review of Economic Studies*, 73(3):793–821.

Reis, R. (2020). Imperfect macroeconomic expectations: Yes, but, we disagree.

Ryngaert, J. (2017). What do (and don't) forecasters know about us inflation. *PhD Thesis*.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1):205–229.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.

Saez, E. and Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, 106(1):24–45.

Sausgruber, R. and Tyran, J.-R. (2005). Testing the Mill hypothesis of fiscal illusion. *Public choice*, 122(1-2):39–68.

Schneeweiß, H. and Komlos, J. (2009). Probabilistic rounding and sheppard's correction. *Statistical Methodology*, 6(6):577–593.

Scholnick, B. (2013). Consumption smoothing after the final mortgage payment: Testing the magnitude hypothesis. *Review of Economics and Statistics*, 95(4):1444–1449.

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

Sims, C. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 317–356. Elsevier.

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

Sims, C. A. (2010). Rational inattention and monetary economics. In *Handbook of monetary economics*, volume 3, pages 155–181. Elsevier.

Stantcheva, S. (2019). Understanding economic policies: What do people know and how can they learn? Technical report, Harvard University Working Paper.

Stark, T. (2013). Spf panelists' forecasting methods: A note on the aggregate results of a november 2009 special survey. *Federal Reserve Bank of Philadelphia*.

Taubinsky, D. and Rees-Jones, A. (2017). Attention variation and welfare: theory and evidence from a tax salience experiment. *The Review of Economic Studies*, 85(4):2462–2496.

Taylor, J. (2001). How the rational expectations revolution has changed macroeconomic policy research. In *Advances in Macroeconomic Theory*, pages 79–96. Springer.

Tutino, A. (2013). Rationally Inattentive Consumption Choices. *Review of Economic Dynamics*, 16(3):421–439.

Woodford, M. (2009). Information-Constrained State-Dependent Pricing. *Journal of Monetary Economics*, 56(S):S100–S124.

Woodford, M. (2013). Macroeconomic analysis without the rational expectations hypothesis. *Annual Review of Economics*, 5(1):303–346.

# List of Figures

# List of Tables