Sparse Dynamic Programming and Aggregate Fluctuations

Xavier Gabaix*

December 31, 2013
Preliminary and Incomplete

Abstract

This paper proposes a way to model boundedly rational dynamic programming in a parsimonious and tractable way. It first illustrates the approach via a boundedly rational version of the consumption-saving life cycle problem. The consumer can pay attention to the variables such as the interest rate and his income, or replace them, in his mental model, by their average values – this way using a "sparse" model of the world. Endogenously, the consumer pays little attention to the interest rate but pays keen attention to his income. This helps resolve some extant puzzles in consumption behavior, especially the tenuous link between interest rates and consumption growth. The model is then applied to a Merton-style portfolio choice problem. This problem is usually quite complex and formidable. We see how a sparse agent will handle the problem, and will have a simpler solution to it: the agent may for instance pay limited or no attention to the varying equity premium and hedging demand terms.

Finally, the paper studies the impact of bounded rationality on macroeconomic outcomes, in a prototypical DSGE model with one variable, capital. We find that in general equilibrium, bounded rationality leads to more persistent shocks, and to larger aggregate fluctuations.

*xgabaix@stern.nyu.edu. I thank Jerome Williams for very good research assistance. For useful comments I thank Jennifer La'O, Ali Lazrak, Thomas Sargent, Alp Simsek and seminar participants at Chicago, the Federal Reserve Board, the Minnesota macro theory conference, NBER, NYU, Penn and Yale. I am grateful to INET and the CGEB for financial support.

1 Introduction

This paper proposes a way to do dynamic programming, but with an element of bounded rationality. This is first illustrated in microeconomic contexts, in canonical consumption-investment problems. Then, the framework is used to study some macroeconomic implications of bounded rationality. One conclusion is that with bounded rationality, macroeconomic fluctuations are larger and more persistent.

Before the macro consequences, let us study the micro motivation.

Modelling bounded rationality: first, microeconomics. The issue of rationality is important. One of the criticisms of traditional economic models is the potential unrealism of the infinitely forward-looking agent who computes the whole equilibrium in her own head. This lack of realism has long been suspected to be the cause of some empirical misfits that we will review below. Behavioral economics aims to provide an alternative. However, the greatest successes of behavioral economics change the agents' tastes (e.g. prospect theory or hyperbolic discounting) or their beliefs (e.g. overconfidence), but keeps the assumption of rationality. When tackling the rationality assumption, there is much less agreement and no dynamic alternative to the traditional model has really emerged. This paper proposes a compromise that keeps much of the generality of the rational approach and injects some of the wisdom of the behavioral approach, mostly inattention and simplification. It does so by proposing a way to insert some bounded rationality into a large class of problems, the "recursive" contexts, i.e. with dynamic programming in some stochastic steady state.

To illustrate these ideas, let us consider a canonical consumption-savings problem. The agent maximizes utility from consumption, subject to a budget constraint, with a stochastic interest rate and stochastic income. In the rational model, the agent solves a complex DP problem with three state variables (wealth, income and the interest rate). This is a complex problem that requires a computer to solve it.

How will a boundedly rational agent behave? I assume that the agent starts with a much simpler model, where the interest rate and income are constant – this is the agent's "default" model. Only one state variable remains, his wealth. He knows what to do then, but what will he do in a more complex environment, with stochastic interest rate and stochastic income? In the sparse version, he considers parsimonious enrichments to the value function, as in a Taylor expansion. He asks, for each component, whether it will matter enough for his decision. If a given feature (say, the interest rate) is small enough compared to some threshold (taken to be a fraction of standard deviation of consumption), then he drops the feature, or partially attenuates it. The result is a consumption policy that pays partial attention to income, and possibly no attention at all to the interest rate. This does seem realistic.

The result is a sparse version of the traditional permanent-income model. We see that it is often *simpler* than the traditional model. Indeed, the agent typically ends up using a rule which is simpler (e.g., not paying attention to the interest rate). Hence, the framework can avoid the pitfall of some behavioral models, which often lead to agents solving problems that are more complex than those of the traditional model. Arguably, the reason those models are more complex is indeed their maintained assumption of some form of hyper-rationality.

One application is a Merton-style dynamic portfolio choice problem, i.e. allocating one's wealth between stocks and bonds when the expected returns are stochastic and correlated with past returns. This is a notoriously complicated problem for a rational agent. I study how a sparse agent would handle it. The sparse agent first anchors his action by imagining he's facing a simpler problem – a world with a constant equity premium. Then, he can sparsely enrich his model to take into account the more complex features (the stochasticity of the equity premium, its correlation with past returns, which creates a hedging demand). Hence the agent will take these complex features into account only partially, or not at all. This may be a more satisfying description than the hyper-rational model of how people behave in a complex environment. At the very least, it is important to have a concrete alternative to that hyper-rational model.

Let us now turn to macro consequences of this approach. The main conclusion is: with bounded rationality, macroeconomic fluctuations are larger and more persistent.

I illustrate this proposition, and qualify it, as it appears to hold for most reasonable values of the parameters, but can be overturned for extreme values.

To see the idea, which is fundamentally quite simple, imagine first an economy with only one state variable, capital. It starts with a steady state amount of capital. Then, there is a positive shock to the endowment of capital. In a rational economy, agents would consume a certain fraction of it, say 6%, every period. That will lead the capital stock to revert quickly to its mean. However, in a an economy with sparse agents, investors will not pay full attention to the additional capital. They will consume less of it than a rational agent would. Hence, capital will be depleted more slowly and will mean-revert more slowly. The shock has more persistent effects.

Given that shocks are more persistent, past shocks accumulate more. Mechanically, this leads to larger average deviations of capital from its trend. As a consequence, the interest rate and GDP also have larger, and more persistent, deviations from trend.

The model allows us to express those ideas in simple, quantitative ways. It allows us to explore them in richer environments, e.g. with shocks to both productivity and the capital stock. The proposition, "BR leads to larger and more persistent fluctuations," still holds true for most parameter values.

Literature review. Besides behavioral lit, cite Krusell Smith (but accent here is on the consequences of BR), Caballero (1995), Campbell-Mankiw (1989), Sims, Maćkowiak & Wiederholt, Mankiew-Reis, Veldkamp, Woodford. One difference: no entropy, so much more simplicity.

The rest of the paper is as follows. Section 2 studies the best response of an agent. The leading example is a consumption-savings problem. Once it is well understood, I formulate the more general notation of sparse DP, in section 3, which also treats the Merton portfolio problem. Section 4 uses the framework to study a general equilibrium situation. If formulates and illustrates the amplifying effect of sparsity on aggregate fluctuations. Section 5 presents an application to the (failure of) Ricardian equivalent. Section 7 concludes. The Appendix contains the more technical material and derivations.

2 Partial equilibrium: Sparse dynamic programming in a consumption-savings problem

2.1 Bounded Rationality in a 2-Period Problem

Consider for concreteness the following decision problem with just two periods: the agent's value function is:

$$V(c, \widehat{r}_t, \widehat{y}_t) = u(c) + \beta v\left((1 + \overline{r} + \widehat{r}_t)(w - c) + \overline{y} + \widehat{y}_t\right),$$

and he wishes to solve $\max_c V(c, \widehat{r}_t, \widehat{y}_t)$. That is, the consumer starts from an initial wealth w, and picks his consumption c in order to maximize his utility, given that next period's consumption will be next period's income, $y_t = \overline{y} + \widehat{y}_t$, plus today's savings, $w - c$, compounded by the interest rate, $r_t = \overline{r} + \widehat{r}_t$. Here $\overline{r}$ is the average value of the interest rate (I take the default value to be the average), and $\widehat{r}_t$ is the (mean-zero) deviation of the interest rate from its average; the same holds for $\overline{y}$, the average income, and $\widehat{y}_t$, the deviation of income from its average.

A rational consumer will solve: $\max_c V(c, \widehat{r}_t, \widehat{y}_t)$. What will a sparse consumer do? Using a mix of psychological and economic reasoning, I propose in Gabaix (2013) a reasonably systematic way of handling this. The agent trades off the cost of having an imperfect decision against the benefits of saving on "thinking costs" . This leads to an algorithm that boils down to the following procedure in our consumption-investment case.

First, the consumer knows what to do under a "default model" where $(\widehat{r}_t, \widehat{y}_t) = (0, 0)$, i.e., all variables are at their average values. Then, the consumer has cognitive access to $\partial c / \partial r = -V_{cr}/V_{cc}$ at the default model, i.e., by how much consumption should change if the interest rate goes up by a small amount. It may seem a bit strange that the consumer might know so much, but this assumption captures parsimoniously the fact that people do have a sense that some quantities (e.g., their income) matter a lot, while others (e.g., the volatility of the 1-year interest rate and, perhaps, that interest rate itself) do not matter very much.

Step 1. Replace the interest rate $\widehat{r}_t$ (to be more precise, the deviation of the interest rate from its average) with its truncated version: the interest rate perceived by a sparse agent is (very shortly I will motivate and explain this particular formula):

$$\widehat{r}_t^s = \tau\left(1, \frac{\kappa \sigma_c}{\frac{\partial c}{\partial r} \sigma_r}\right) \widehat{r}_t, \tag{1}$$

where $\partial c / \partial r$ is taken at the default model, and the truncation function

$$\tau(\mu, k) = \mu \max\left(1 - \frac{k^2}{\mu^2}, 0\right) \tag{2}$$

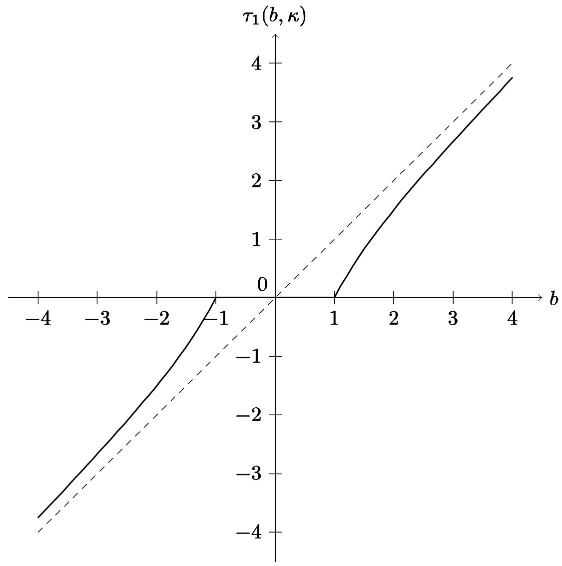

Figure 1: The truncation function τ. A sparse agent replaces a slope b by a truncated slope $\tau(b,\kappa)$, where κ is a context-dependent threshold.

is represented in Figure 1. $\widehat{r}_t^s$ is the deviation of the interest rate from its default perceived by a sparse consumer.

Likewise, the perceived income innovation is: $\widehat{y}_t^s = \tau\left(1, \frac{\kappa\sigma_c}{\frac{\partial c}{\partial y}\sigma_y}\right)\widehat{y}_t$.

Step 2: Then, the sparse agent solves $\max_c V\left(c, \widehat{r}_t^s, \widehat{y}_t^s\right)$.

Step 2 is unproblematic: given the perceived interest rate and income, the agent optimizes consumption. The heart of the model is in Step 1. To interpret rule (1), note that it implies: "Replace the interest rate with 0 if taking the interest rate into account changes consumption by less than κ standard deviations, i.e., if $\left|\frac{\partial c}{\partial r}\sigma_r\right| < \kappa\sigma_c$."

This means: on average, a one-standard-deviation change in the interest rate makes the sparse agent change his consumption by only $\frac{\partial c}{\partial r}\frac{\sigma_r}{\sigma_c}$ standard deviations of consumption. If that ratio is small enough (I calibrate the model to $\kappa = 0.3$, so that features which account for less than $\kappa^2 = 9\%$ of the variance are eliminated), then replace the interest rate by 0.[1]

The penalty for lack of sparsity, κ, is akin to an index of bounded rationality: if $\kappa = 0$, the agent is fully rational.

Take the case where $\left|\frac{\partial c}{\partial r}\frac{\sigma_r}{\sigma_c}\right| < \kappa$, so that $\widehat{r}_t^s = 0$ and the agent proceeds as if the interest rate was the average interest rate $\overline{r}$ rather than the true interest rate r_t. We have the picture of a sensible agent: he does not pay attention to the interest rate all the time, he saves (so he is not "myopic" in the sense of heavily discounting the future), but he does not obsess about smoothing his consumption given all fluctuations in the interest rate. This agent is arguably more sensible and realistic than the traditional agent (below I will offer some empirical evidence for that intuition).

[1] The main paper provides a microfoundation based on the welfare loss from a suboptimal answer.

Here, we use the "average values" for the interest rate and income shocks. In a one-shot problem, we would use the above rule, replacing $|\sigma_r|$ by $|\widehat{r}_t|$, so that instead of (1) we obtain $\widehat{r}_t^s = \tau\left(\widehat{r}_t, \frac{\kappa\sigma_c}{\frac{\partial c}{\partial r}}\right)$. Then, the rule becomes: "Replace the interest rate by 0 iff taking it into account makes consumption change by less than κ standard deviations." Indeed, the agent does not respond to the interest rate at all if $\left|\frac{\partial c}{\partial r} \times \widehat{r}_t^s\right| < \kappa\sigma_c$. Thus, most of the time, the agent will not take the small fluctuations in the interest rate into account, but will pay attention to changes in the interest rate only when changes are very large (e.g., if there is a large, one-time discount of, say, cars).

The truncation rule embodies the idea that an agent who seeks "sparsity" (uncluttering his mind from lots of small things) should sensibly drop relatively unimportant features: if they account for less than κ standard deviations of the actions, they are dropped entirely. In addition, if the features are larger than that cutoff, they are still dampened: in Figure 1, $\tau(\mu, \kappa)$ is below the 45 degree line (for positive μ; in general, $|\tau(\mu, \kappa)| < |\mu|$). This reflects Kahneman and Tversky's "anchoring-and-adjustment" process, in which there is an anchor in the default model, and then a partial adjustment toward the truth. This feature could be abandoned, using the "hard-thresholding" function $\tau_0(\mu, \kappa) = \mu 1_{|\mu| > \sqrt{2}|\kappa|}$, see Appendix A. However, the above function τ has the advantage of yielding continuous demand curves, which are likely in practice. For many cases, the smooth adjustment makes more empirical sense than the "all-or-nothing" adjustment, which predicts discontinuities that we are unlikely to see empirically.

I hope that the reader has gotten a sense of the intuition for the model in a (quasi-)static context. Let us now see how to proceed in more dynamic contexts.

2.2 Infinite-Horizon Problem

One important payoff from the framework is that it allows for boundedly rational dynamic programming (BRDP). This is important because many models in macroeconomics and finance take the form of dynamic programming (Ljungqvist and Sargent 2004). The outcome will be a model that is often *simpler* than the traditional model, because agents pay attention to fewer things and, in particular, do not react to all future variables.

In addition, it is well-known that an important conceptual and practical problem when dealing with dynamic programming is the curse of dimensionality. Strictly speaking, there are perhaps over 1,000 state variables that should matter in our decisions, but solving dynamic-programming problems with more than a few state variables (let alone 1,000 state variables) is extremely hard in practice because of the combinatorial explosion of the problem's complexity. Even the most powerful computers cannot handle such complexity and solve the problems exactly. Given that, how would a boundedly rational agent proceed?

I illustrate the approach in a canonical consumption-investment problem. The agent has utility $\mathbb{E}\sum_{t=0}^{\infty} \beta^t c_t^{1-\gamma}/(1-\gamma)$. We assume he has solved the life-cycle problem in a simple model, where the interest rate is constant at $\overline{r}$ (for simplicity, assume here that

$\overline{R} \equiv 1 + \overline{r} = \beta^{-1}$) and his income is constant at $\overline{y}$: his wealth w_t evolves according to

$$w_{t+1} = (1 + \overline{r})(w_t - c_t) + \overline{y}$$

(that is, wealth at $t+1$ is savings at t, $w_t - c_t$, invested at rate $\overline{r}$, plus current income, $\overline{y}$). Then, the optimal consumption is $c^d(w_t) = (\overline{r} w_t + \overline{y}) / \overline{R}$, and the value function is $V^d(w_t) = A(\overline{r} w_t + \overline{y})^{1-\gamma}$ for a constant A.

Now, the agent is told that the world is more complicated: the interest rate is actually $\overline{r} + \widehat{r}_t$ and his income is $\overline{y} + \widehat{y}_t$, where $\widehat{r}_t$ and $\widehat{y}_t$ are deviations of the interest rate and income from their means, respectively, and follow AR(1) processes:

$$y_{t+1} = \rho_y \widehat{y}_t + \varepsilon^y_{t+1}, \qquad r_{t+1} = \rho_r \widehat{r}_t + \varepsilon^r_{t+1}$$

$\varepsilon^{\cdot}_{t+1}$ are independent disturbances with mean zero. Hence, wealth follows:

$$w_{t+1} = (1 + \overline{r} + \widehat{r}_t)(w_t - c_t) + \overline{y} + \widehat{y}_t.$$

What will the consumption function $c(w_t, \widehat{y}_t, \widehat{r}_t)$ of a sparse agent be? It is difficult, because this is a dynamic-programming problem with 3 state variables, and has no closed-form solution. Under the previous approach, one might think that one should solve for the value function $V(w_t, \widehat{y}_t, \widehat{r}_t)$; but that would be a very difficult task in general: dynamic programming with 3 or more (and in practice perhaps 20) state variables is very difficult. However, we obviate this difficulty by using the following algorithm.

Step A (Taylor expansion around the simple, default model with just one state variable). We observe that a rational agent would consume, for small disturbances $\widehat{y}_t$ and $\widehat{r}_t$:

$$\begin{aligned} \ln c^r(w_t, \widehat{y}_t, \widehat{r}_t) &= \ln c^d(w_t) + b_y \widehat{y}_t + b_r \widehat{r}_t \\ &\quad + \text{2nd-order terms.} \end{aligned} \tag{3}$$

Importantly, the terms b_y and b_r are easy to derive by a local expansion of the simple, one-dimensional value function $V^d(w_t)$ (i.e., without solving for the full function $V(w_t, \widehat{y}_t, \widehat{r}_t)$). Indeed, by perturbation arguments, detailed in the Appendix, we find:

$$b_y = \frac{\overline{r}}{\overline{R}\left(\overline{R} - \rho_y\right) c^d_t}, \qquad b_r = \frac{\overline{r}\left(\frac{w_t}{c^d_t} - 1\right) - 1/\gamma}{\overline{R} - \rho_r}. \tag{4}$$

Then, we assume that the sparse agent somehow has cognitive access to b_r and b_y: while it may seem counterintuitive, this merely represents that the sparse agent senses that, for instance, the interest rate is not a very important decision for his consumption ($|b_r|$ is small).

Step B (Simplification of the reaction function). The agent does a sparse truncation of (3), according to equation (1). Hence, we obtain the following.

Proposition 1 *A sparse agent has the following consumption policy, up to second order terms:*

$$\ln c_t^s = \ln c^d(w_t) + b_y^s \widehat{y}_t + b_r^s \widehat{r}_t \tag{5}$$

where (for $x = y, r$) $b_x^s := \tau\left(b_x, \frac{\kappa\sigma_{\ln c}}{\sigma_x}\right)$ and b_x are in (4).

Equation (5) shows a "feature-by-feature" truncation. It is useful because it embodies in a compact way the policy of a sparse agent in quite a complicated world. Note that the agent can solve this problem without solving the 3-dimensional (and potentially 21-dimensional, say, if there are 20 state variables besides wealth) problem. Only local expansions and truncations are necessary.

In this manner, we arrive at a quite simple way to do sparse dynamic programming. There is just one continuously-tunable parameter, κ. When $\kappa = 0$, the agent is (to the leading order) the traditional rational agent. When κ is large enough, the agent is fully sparse, and does not react to any variable. Hence, we have a simple, smooth way to parametrize the agent, from very sparse to fully rational.

2.3 Application: Insensitivity to the Interest Rates and Low Measured Intertemporal Elasticity of Substitution

To get a feel for the effects, consider a calibration with (using annual units): $\gamma = 1$, $r = 5$, $\overline{w} = 2\overline{c}$, $\overline{c} = 1$, $\sigma_r = 0.8\%$, $\sigma_y = 0.2\overline{c}$, $\rho_y = 0.95$, $\sigma_{\ln c} = 5\%$, and $\rho_r = 0.7$: as income shocks are persistent, they are important to the consumer's welfare.

Then, Figure 2 shows the impact of a change in the interest rate and income on consumption. Consider the left panel, b_r^s. If the cost of rationality is $\kappa = 0$, then the agent reacts like the rational agent: if interest rates go up by 1%, then consumption falls by 2.8% (the agent saves more). However, for a sparsity parameter $\kappa \simeq 0.5$, the agent essentially does not respond to interest rates. Psychologically, he thinks "the interest rate is too unimportant, so let me not take it into account." Hence, the agent does not react much to the interest rate, but will react more to a change in income (right panel of Figure 2), which is more important: the sensitivity of consumption to income remains high even for a high cognitive friction κ. Note that this "feature-by-feature" selective attention could not be rationalized by just a fixed cost to consumption, which is not feature-dependent.

The same reasoning holds in every period. The above describes a practical way to do sparse dynamic programming. In some cases, this is simpler than the rational way (as the agent does not need to solve for the equilibrium), and it may also be more sensible.

Consequence. A behavioral solution to puzzles and controversies regarding the intertemporal elasticity of substitution For many finance applications (e.g., Bansal and Yaron 2004, Barro 2009, Gabaix 2012), a high intertemporal elasticity of substitution (IES, denoted $\psi = 1/\gamma$) is important ($\psi > 1$). However, micro studies point to an IES of less than 1 (e.g., Hall 1988). I show how this may be due to the way econometricians

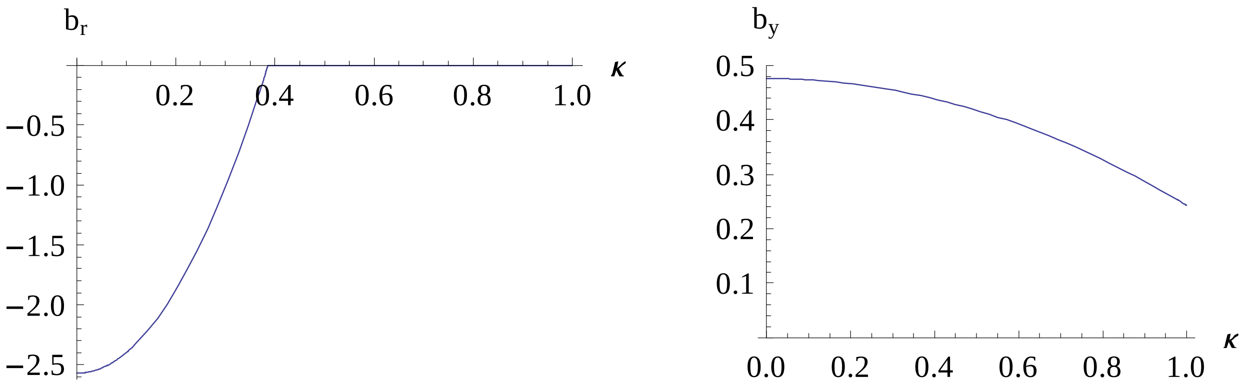

Figure 2: Impact of a change in the interest rate (resp. income) on consumption, as the function of weight on sparsity, κ. $\kappa = 0$ is the rational-agent model.

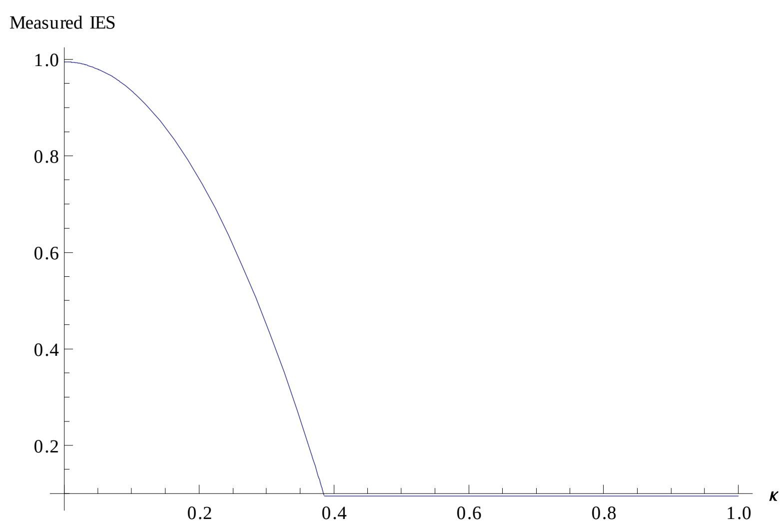

Figure 3: Measured intertemporal elasticitiy of substitution (IES), $\widehat{\psi}$, if the consumer is sparse with cost κ, while the econometrician assumes he is fully rational. The true IES is $\psi = 1$.

proceed, by fitting the Euler equation, which yields $\ln c_{t+1} - \ln c_t = \frac{\widehat{\psi}}{R} r_t +$ constant, where $\widehat{\psi}$ is the measured IES. If the consumer "under-reacts to the interest rate," the measured IES will be biased towards 0. Using the above model, we can more precisely calculate that if consumers are boundedly rational (in the sense laid out above), the estimated IES will be: $\widehat{\psi} = \overline{r}\left(w_t/c_t^d - 1\right) - b_r^s \overline{R}\left(\overline{R} - \rho_R\right)$. This is a point prediction that goes beyond Chetty (forth.)'s prediction of an interval bound. Hence we obtain:

Proposition 2 *An econometrician fitting an Euler equation even though the agent is sparse will estimate a downwardly-biased IES (intertemporal elasticity of substitution):*

$$\widehat{\psi} = \psi - \overline{R}\left(\overline{R} - \rho_R\right)\left(b_r^s - b_r\right) < \psi$$

where $\widehat{\psi}$ is the estimated IES, ψ the true IES and $b_r^s - b_r$ is the difference between the sparse agent's and the traditional agent's interest-rate sensitivity of consumption.

The above calibration yields Figure 3, which plots the measured IES $\widehat{\psi}$ if the consumer is sparse with sparsity cost κ. If $\kappa = 0$, the consumer is the traditional, frictionless rational agent. We see that as κ increases, the IES becomes more and more biased. Hence, inattention may explain why while macro-finance studies require a high IES, microeconomic studies find a low IES.[2]

2.4 Application: Source-dependent Marginal Propensity to Consume

The agent has initial wealth w, future income y, can consume c at time 1, and invest the savings at a rate R. Hence, the problem is as follows. Given an initial wealth w, solve $\max_c V = u(c) + \mathbb{E}\left[v\left(y + R\left(w - c\right)\right)\right]$, where income is $y = y_* + \sum_{i=1}^n y_i$: there are n sources of income y_i with mean 0. Let us study the solution of this problem with the algorithm. The agent observes the income sources sparsely: he uses the model $y(m) = y_* + \sum_{i=1}^n m_i y_i$, with m_i to be determined. Applying the model, we obtain (assuming exponential utility with absolute risk aversion γ for simplicity)

Proposition 3 *Time-1 consumption is: $c = \frac{1}{1+R}(Rw + \delta/\gamma - \gamma\sigma_\varepsilon^2/2 + y_* + \sum_i m_i y_i)$, $m_i = \tau(1, \frac{\kappa^m \sigma_{c_2}}{\sigma_{y_i}})$. The marginal propensity to consume (MPC) at time 1 out of income source i is:*

$$MPC_i^s = MPC_i^r \cdot m_i, \tag{6}$$

where $MPC_i^s = (\partial c/\partial y_i)^s$ is the MPC under the sparse model, and $MPC_i^r = (\partial c/\partial y_i)^r$ is the MPC under the traditional rational-actor model. Hence, in the sparse model, unlike in the traditional model, the marginal propensity to consume is source-dependent.

---

[2]This is in the spirit of Gabaix and Laibson (2002)'s analysis of the biases in the estimation of the coefficient of risk aversion with inattentive agents, in a different context and a more tractable model. See also Fuster, Laibson and Mendel (2010) for a model where agents' use of simplified models leads to departures from the standard aggregate model.

Different income sources have different marginal propensities to consume – this is reminiscent of Thaler (1985)'s mental accounts. Equation (6) makes another prediction, namely that consumers pay more attention to sources of income that usually have large consequences, i.e., have a high σ_{y_i}. Slightly extending the model, it is plausible that a shock to the stock market does not affect the agent's disposable income much – hence, there will be little sensitivity to it: the MPC out of wage income will be higher than the MPC to consume out of portfolio income.

This model shares similarities with models of inattention based on a fixed cost of observing information. Those models are rich and relatively complex (they necessitate many periods, or either many agents or complex, non-linear boundaries for the multidimensional s, S rules, or signal extraction as in Sims 2003), whereas the present model is simpler and can be applied with one or several periods. As a result, the present model, with an equation like (6), lends itself more directly to empirical evaluation. Some interesting "low-complexity" models include Bordalo, Gennaioli, and Shleifer (2011) and Koszegi and Szeidl (2011). A distinctive feature of the model presented in this note is its ability to handle continuous choices (e.g., a consumption level) rather than the discrete choice between distinct goods.

3 General Framework

Here we present the more general procedure underlying the model of the previous section. First, we present the smax operator, an operator representing sparse maximization. Then, we state the dynamic programming problem, and some results that make its computation easy.

3.1 Previous results on the sparse max operator

In Gabaix (2013), I defined a sparse max or smax operator, which is a sparse version of the max operator. The agent faces a maximization problem which is, in its rational version, $\max_a u(a, x)$. The x_i are viewed by the agent as being drawn with a standard deviation σ_i, and covariance σ_{ij}. There is a nonnegative parameter κ, which is a taste for sparsity. When $\kappa = 0$, the agent is the traditional agent.

Definition 1 *(Sparse max operator, Gabaix 2013) The sparse maximum, u^s, and maximand, a^s, of a function $u(a, x)$ written:*

$$u^s := \operatorname*{smax}_a u(a, x), \qquad a^s := \arg \operatorname*{smax}_a u(a, x)$$

are defined by the following procedure.

Step 1: choose the attention vector m^*:

$$m^* = \arg\min_m \frac{1}{2} \sum_{i,j=1\ldots n} (1 - m_i) \Lambda_{ij} (1 - m_j) + \kappa \sum_{i=1\ldots n} |m_i|^\alpha \tag{7}$$

with $\Lambda_{ij} = -\sigma_{ij} a_{x_i} u_{aa} a_{x_j}$, and $a_{x_i} := -u_{aa}^{-1} u_{ax_i}$.

Define $x_i^s = m_i^* x_i$, the sparse representation of x.

Step 2: Choose the action

$$a^s = \arg\max_a u\left(a, x^s\right) \tag{8}$$

and set the resulting utility to be $u^s = u\left(a^s, x\right)$. In the expressions above, derivatives are evaluated at $x = 0$ and $a^d = \arg\max_a u\left(a, 0\right)$.

In other terms, the agent solves for the optimal m^* that trades off a proxy for the utility losses (the first term in the right-hand side of equation (7)) and a psychological penalty for deviations from a sparse model (the second term on the left-hand side of equation (7)). Then, the agent maximizes over the action a, as if m^* were the true model.

The following proposition derives the main case.

Proposition 4 *(Gabaix 2013) When variables are perceived to be uncorrelated, the* smax *operator can be equivalently formulated as:*

$$a^s = \arg\max_a u\left(a, m_1^* x_1, ..., m_n^* x_n\right)$$

with

$$m_i^* = \mathcal{A}_\alpha\left(1, \frac{\kappa^{1/2}}{\sigma_i \left|a_{x_i} u_{aa} a_{x_i}\right|^{1/2}}\right) \tag{9}$$

and $\partial a/\partial x_i = -u_{aa}^{-1} \cdot u_{a,x_i}$.

The intuition is that the x_i's are truncated. If $\left|\partial a/\partial x_i\right|$ is small enough, so that x_i shouldn't matter much any way, then $m_i = m_i^d$, and the agent doesn't pay attention to x_i (if $m_i^d = 0$).

For instance, in the first part of this paper, equation (1) came from the above proposition, with $m_r = \tau\left(1, \frac{\kappa\sigma_c}{\frac{\partial c}{\partial r}\sigma_r}\right)$.

The following proposition gives a more explicit version of the action:[3]

Proposition 5 *If the rational action is:*

$$a^s\left(x\right) = a^d + \sum_i b_i x_i + O\left(\|x\|^2\right)$$

then the sparse action is:

$$a^s\left(x\right) = a^d + \sum_i \tau\left(b_i, \frac{\kappa_a}{\sigma_{x_i}}\right) x_i + O\left(\|x\|^2\right) \tag{10}$$

[3] This proposition suggests a potential generalization of the SparseBR algorithm: just postulate the procedure in the propositions, with a potentially different truncation function τ. For instance, we could have $\tau\left(\mu, \kappa'\right) = \mu 1_{|\mu| \geq |\kappa'|}$, or some smoother function.

with

$$\kappa_a := (\kappa / |u_{aa}|)^{1/2}.$$

For a quadratic utility function $u = -\left(a - \sum_i b_i x_i\right)^2$*, the above expressions are exact (i.e. hold without the* $O\left(\|x\|^2\right)$ *terms).*

We see the contrast. In the first procedure, the slope is chosen before seeing x_i. Hence, the policy is still linear in x_i. In the second policy, the truncation is chosen after seeing the x_i. The policy is now non-linear in x_i. The linearity of policies make the first procedure useful for macro. Equipped with this piece of machinery, we turn to dynamic problems.

3.2 Sparse Dynamic Programming

We consider a stationary environment. The rational version of the DP problem is:

$$\begin{aligned} V(w,x) &= \max_a u(a,w,x) + \beta \mathbb{E} V(w',x') \\ w' &= F^w(w,x,a), \qquad x' = F^x(w,x,a) \end{aligned}$$

where F^w and F^x are potentially random functions, i.e. functions of some noise.

In the sparse version, the vector w is always considered (it's in the default model). However, the vector x represents variables that may not be considered by the sparse agent.

We define the value function as follows:

Definition 2 *The DP value function is the solution (provided it exists) of:*

$$\begin{aligned} V^s(w,x) &= \operatorname*{smax}_a u(a,w,x) + \beta \mathbb{E} V^s(w',x') \\ w' &= F^w(w,x,a), \qquad x' = F^x(w,x,a) \end{aligned}$$

where the smax operator for sparse maximization is defined in Definition 1.

Slightly more explicitly, this is:

$$\begin{aligned} V^s(w,x) &= \operatorname*{smax}_a U(a,w,x) \\ U(a,w,x) &:= u(a,w,x) + \beta \mathbb{E} V^s\left(F^w(w,x,a), F^x(w,x,a)\right) \end{aligned}$$

The notion is recursive. However, the problem is actually quite simple to solve, at least to the first order.

Indeed, we have:

Proposition 6 *For small* x*, we have:*

$$V^s(w,x) = V(w,x) + x\phi(w,x)x$$

where ϕ *is continuous in* (w,x) *and twice differentiable at* $x=0$*, with* $\phi(w,0)$ *negative semi-definite. In other words, the sparse value function and the rational value functions differ only by second order terms in* x*.*

This basically generalizes the envelope theorem. It implies that, at $x = 0$:

$$V_w^s = V_w, \qquad V_{ww}^s = V_{ww}, \qquad V_x^s = V_x \qquad V_{wx}^s = V_{wx} \tag{11}$$

However, in most situations we have $V_{xx}^s \neq V_{xx}$, and indeed $V_{xx}^s < V_{xx}$

This leads to a simple proposition to calculate the value function.

Proposition 7 *(Calculation of the optimal sparse policy). Suppose that $F_{a|x=0}^x = 0$ (which is usually satisfied in models). Consider the first order expansion of the optimal policy for small x,*

$$a^r(w,x) = a^d(w) + \sum_i b_i(w)\, x_i + O\left(x^2\right)$$

Then, the sparse policy is:

$$a^s(w,x) = a^d(w) + \sum_i \tau\left(b_i(w), \frac{\kappa_a}{\sigma_{x_i}}\right) x_i + O\left(x^2\right) \tag{12}$$

This proposition will be quite useful. To derive policies, first we can simply do a Taylor expansion of the rational policy around the default model, and then truncate term by term.

I conclude with a remark which will be useful later, drawing again on Gabaix (2013). As κ has the units of utils, it cannot be a primitive parameter. One can make it more endogenous with the primitive, unitless parameter $\overline{\kappa}$, by setting:

$$\kappa = \overline{\kappa} \sum_{ij} \Lambda_{ij} \tag{13}$$

3.3 Application: Dynamic Portfolio Choice

I now study a Merton problem with dynamic portfolio choice. The agent's utility is: $\mathbb{E}\left[\frac{1}{1-\gamma}\int_0^\infty e^{-\rho s} c_s^{1-\gamma} ds\right]$, and his wealth w_t evolves according to:

$$dw_t = \left(-c_t + rw_t\right) dt + w_t\theta_t\left(\pi_t dt + \sigma dZ_t\right)$$

where θ_t is the allocation to equities. The equity premium $\pi_t = \overline{\pi} + \widehat{\pi}_t$ has a variable part $\widehat{\pi}_t$, which follows

$$d\widehat{\pi}_t = -\phi\widehat{\pi}_t dt - \chi_t \sigma dZ_t + \sigma_\pi' dB_t$$

The parameter $\chi_t \geq 0$ indicates that equity returns mean-revert: good returns today lead to lower returns tomorrow. That will create a hedging demand term – a term that's quite complex.

The agent's problem is to find the policies c_t and θ_t to maximize expected utility under the constraints. Hence, the value function for the agent is $V\left(w_t, \widehat{\pi}_t, \chi\right)$.

We have the following (using the notation $\psi = 1/\gamma$ for the IES):

Proposition 8 *(Behavioral dynamic portfolio choice) The fraction of wealth allocated to equities is, with* $\overline{\theta} := \frac{\overline{\pi}}{\gamma\sigma^2}$

$$\theta_t^s = \overline{\theta} + \tau\left(\frac{\widehat{\pi}_t}{\gamma\sigma^2}, \kappa\sigma_\theta\right) + \tau\left(B\chi_t, \kappa\sigma_\theta\right)$$

while consumption is:

$$c_t^s = \mu w_t \left[1 + \tau\left(\frac{1-\psi}{\mu+\phi_\Lambda}\overline{\theta}\widehat{\pi}_t, \kappa\sigma_{\ln c}\right) + \tau\left(B\left(1-\psi\right)\frac{\overline{\pi}\chi}{\mu+\phi_\Lambda}, \kappa\sigma_{\ln c}\right)\right]$$

using the notations:

$$B = \frac{\left(1-\frac{1}{\gamma}\right)\overline{\theta}}{\mu+\phi}, \qquad \mu := \psi\rho + \left(1-\psi\right)\left(r + \overline{\theta}\overline{\pi}\right).$$

Proposition 8 predicts the choice of a sparse agent. When $\kappa = 0$, it is the policy of a fully rational agent, e.g. as worked out by Campbell and Viceira (2002). When $\kappa > 0$, it is the policy of a sparse agent. When κ is larger, portfolio choice becomes insensitive to the change in the equity premium, $\widehat{\pi}_t$, and the agent thinks less about the mean-reversion of asset, the $B\chi$ terms.

In addition, the agents' consumption function pays little attention to the mean-reversion of assets. [Next iteration should have a calibration, and the proof.]

4 Bounded Rationality in General Equilibrium

The raison d'être of this model is the tractability that allows us to study GE effects. We start with a basic question. Suppose that agents are sparse; what will the impact in general equilibrium be? The answer is:

Bounded rationality leads to more persistent and larger aggregate fluctuations.

I will illustrate this thesis in a very basic model first, with just one state variable. Then, we'll move on to more complex models. We shall see it holds for many (but not all) values of parameters.

4.1 A simple example with shocks to capital

Let us start with a simple example. The utility function is still $\mathbb{E}\sum_t \beta^t C_t^{1-\gamma}/\left(1-\gamma\right)$. In the aggregate, the capital stock follows:

$$K_{t+1} = F\left(K_t, L\right) + \left(1-\delta\right)K_t - C_t + \varepsilon_{t+1} \tag{14}$$

where ε_{t+1} are mean-zero shocks, whose distribution we'll specify later. This way, there is just one state variable in the economy, the capital stock.

The question is: how will a sparse economy react compared to a traditional (i.e., rational-agent) economy?

This is a textbook example: and can be found in Acemoglu (2009, Chapter 8), Blanchard-Fischer (1989, Chapter 2), Romer (2012, Chapter 2); it introduces generations of students to macroeconomics. However, it looks somewhat odd (in my opinion), with these infinitely-rational forward looking agents that calculate the whole macroeconomic equilibrium in their heads. I present here an alternative to that presentation.

4.1.1 The essence of the argument

I present first the essence of the argument, sweeping under the rug several specifics that will be made explicit in the more special models.

If there were no shocks, the economy would be at the steady state, with capital stock K^*. I use the hat notation for the deviation (not in logs) from the mean, e.g. $\widehat{K}_t = K_t - K^*$. The law of motion for capital (14) is, in linearized form:

$$\widehat{K}_{t+1} = (1 + r)\widehat{K}_t - \widehat{C}_t + \varepsilon_{t+1} \tag{15}$$

where r is the steady state interest rate, $r = \beta^{-1} - 1$.

Given there is one state variable, the policy function of the agent (rational or not) will take the form of a deviation of consumption from trend:

$$\widehat{C}_t = b\widehat{K}_t$$

for some positive b.

Plugging this into (15) we obtain: $\widehat{K}_{t+1} = (1 + r - b)\widehat{K}_t + \varepsilon_{t+1}$, i.e.

$$\widehat{K}_{t+1} = (1 - \phi)\widehat{K}_t + \varepsilon_{t+1} \tag{16}$$

with a speed of mean-reversion:

$$\phi = b - r. \tag{17}$$

Generally, sparse consumers are less attentive than rational consumers, hence their policy will take the form:

$$\widehat{C}_t = b'\widehat{K}_t$$

for some b' :

$$0 < b' < b$$

Hence, *the speed of mean reversion in the sparse economy will be less than in the rational economy*:

$$\phi' = b' - r < \phi.$$

Therefore, fluctuations mean-revert less quickly in the sparse economy. This is because consumers respond less to shocks.

Finally, squaring equation (16), we obtain: $var\widehat{K}_{t+1} = (1-\phi)^2 var\widehat{K}_t + \sigma_\varepsilon^2$. As in the steady state, $var\widehat{K}_{t+1} = var\widehat{K}_t$,

$$var\widehat{K}_t = \frac{\sigma_\varepsilon^2}{1-(1-\phi)^2}$$

When shocks mean-revert more slowly (lower ϕ), the average deviation of the stock price from trends is higher (shocks "pile up" more). Hence, *the variance of shocks will be larger in the sparse economy than in the rational economy.*

The above argument gives the qualitative essence of what is going on. However, we need to flesh it out more to obtain more quantitative answers. Let us do that now.

4.1.2 The more detailed argument

The rational economy Formally, the rational agent has a value function $V(K_t)$, which satisfies:

$$V(K) = \max_c u(c) + \beta \mathbb{E}\left[V(K')\right]$$
$$K' = f(K,L) + (1-\delta)K - c + \varepsilon_t$$

where $f(K,L)$ is gross output, and $F(K,L) := f(K,L) - \delta K$ is output net of depreciation.

The solution is that small deviations of the capital stock mean revert at a speed ϕ ($\widehat{K}_t = e^{-\phi t}\widehat{K}_0$) that we will characterize soon.

It comes from the following policy function of the representative agent (see Appendix):

$$\widehat{C}_t = b\widehat{K}_t$$
$$b = r + \frac{\xi}{r+\phi}$$
$$\xi = -C^* F''(K^*)/\gamma > 0$$

By the argument above, this leads to a speed of mean-reversion:

$$\phi = b - r = \frac{\xi}{r+\phi} \tag{18}$$

Hence, solving via rational expectations imposes:

$$\phi = \frac{-r+\sqrt{r^2+4\xi}}{2}. \tag{19}$$

The boundedly rational version The agent has wealth k_t (and we normalize the population to be 1, so that in equilibrium will be equal to K_t, the aggregate wealth). It evolves as:

$$k_{t+1} = (1+r_t)(k_t + y_t - c_t)$$

where $y_t = F(K_t) - K_t F'(K_t)$ is labor income, and $r_t = F'(K_t)$ is the interest rate. We have:

$$\widehat{y}_t = -K^* F''(K^*) \widehat{K}_t$$
$$\widehat{r}_t = F''(K^*) \widehat{K}_t$$

This leads to the optimal policy:

$$\begin{aligned}
\widehat{c}_t &= r\widehat{k}_t + \frac{r}{r+\phi}\widehat{y}_t + \frac{rk^* - c^*\psi}{r+\phi}\widehat{r}_t \\
&= r\widehat{k}_t + \frac{(-rK^*F''(K^*) + (rk^* - c^*\psi) F''(K^*))}{r+\phi}\widehat{K}_t \\
\widehat{c}_t &= r\widehat{k}_t + \frac{\xi}{r+\phi}\widehat{K}_t
\end{aligned}$$

The sparse version is:

$$\widehat{c}_t = r\widehat{k}_t + \tau\left(\frac{\xi}{r+\phi}, \frac{\kappa_c}{\sigma_K}\right)\widehat{K}_t \tag{20}$$

We need to solve for the equilibrium. The speed of mean-reversion ϕ affects the importance of aggregate fluctuations on the agent's policy, and hence affects the attention that the consumer brings to it. In turns, this attention affects the MPC, hence the speed of mean-reversion. Hence, we have a fixed-point to solve, featuring speed of mean-reversion, and attention.

To solve it, we use the "scale-free" version of κ, equation (13).

$$\begin{aligned}
\frac{\kappa_c}{\sigma_K} &= \left(\frac{\kappa}{-u_{cc}}\right)^{1/2}\frac{1}{\sigma_K} = \left(\frac{-\overline{\kappa}u_{cc}\sigma_c^{r2}}{-u_{cc}}\right)^{1/2}\frac{1}{\sigma_K} \\
&= \sqrt{\overline{\kappa}}\frac{\sigma_c^r}{\sigma_K} = \sqrt{\overline{\kappa}}\frac{\left(\frac{\partial c}{\partial K}\right)^r \sigma_K}{\sigma_K} = \sqrt{\overline{\kappa}}\left(\frac{\partial c}{\partial K}\right)^r \\
&= \sqrt{\overline{\kappa}}\left(r + \frac{\xi}{r+\phi}\right)
\end{aligned}$$

Hence, we obtain:

$$\widehat{c}_t = r\widehat{k}_t + \tau\left(\frac{\xi}{r+\phi}, \sqrt{\overline{\kappa}}\left(r + \frac{\xi}{r+\phi}\right)\right)\widehat{K}_t$$

so that in equilibrium, $\widehat{c}_t = b^s \widehat{K}_t$

$$b^s = r + \tau\left(\frac{\xi}{r+\phi}, \sqrt{\overline{\kappa}}\left(r + \frac{\xi}{r+\phi}\right)\right)$$

Hence, the equilibrium mean-reversion ϕ is: $\phi = b^s - r$:

$$\phi = \tau\left(\frac{\xi}{r+\phi}, \sqrt{\overline{\kappa}}\left(r + \frac{\xi}{r+\phi}\right)\right) \tag{21}$$

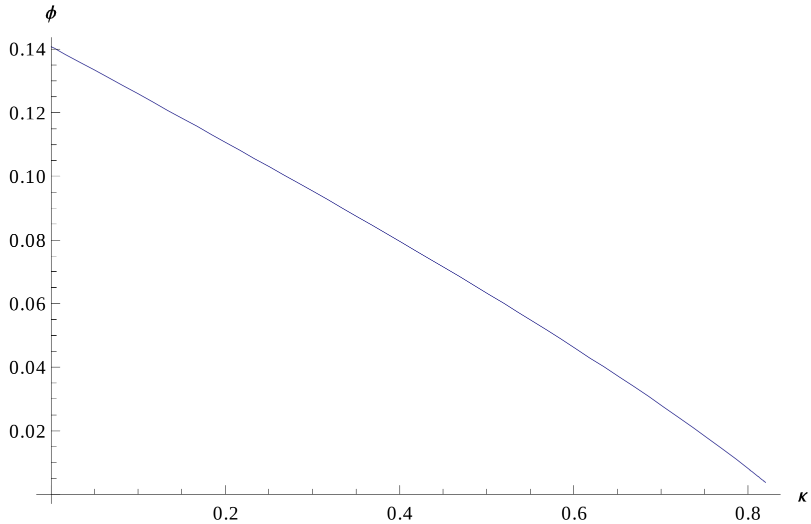

Figure 4: This Figure plots the speed of mean-reversion of fluctuations, ϕ, as a function of the cost of rationality, κ.

Hence, ϕ is a solution of:

$$\phi = \left(\frac{\xi}{r+\phi} - \overline{\kappa} \left(r + \frac{\xi}{r+\phi} \right)^2 \frac{r+\phi}{\xi} \right)^+ .$$

Proposition 9 *Shocks are more persistent in the sparse economy: $\phi < \phi^r$ if $\overline{\kappa} > 0$. More precisely, the speed of mean-reversion is given by*

$$\phi = \frac{\left[-\left(2\overline{\kappa}r^2 + 2\xi\overline{\kappa} + \xi\right) r + \xi\sqrt{r^2 + 4\xi(1-\overline{\kappa})} \right]^+}{2(\xi + \overline{\kappa}r^2)} \tag{22}$$

In particular, ϕ is decreasing in $\overline{\kappa}$, $\phi\left(\overline{\kappa}=0\right) = \phi^r$. We have $\phi > 0$ iff $\overline{\kappa} \leq \overline{\kappa}^* := \left(1 + r^2/\xi\right)^{-2}$.

The following proposition studies the variance of the stocks, $var\widehat{K}_t = \frac{\sigma_\varepsilon^2}{1-(1-\phi)^2}$.

Proposition 10 *Shocks are larger in the sparse economy. More precisely, the quadratic deviation from trend in capital, interest rate and GDP is multiplied by $M\left(\overline{\kappa}\right) = \frac{1-(1-\phi^r)^2}{1-(1-\phi(\overline{\kappa}))^2}$, where ϕ is given by (22). The effect can be unboundedly large: $\lim_{\overline{\kappa}\uparrow\overline{\kappa}^*} M\left(\overline{\kappa}\right) = \infty$.*

Calibration The parametrization is conventional, $F\left(K\right) = K^{1-\alpha}/\left(1-\alpha\right) - \delta K$, $r = 5\%$, $\delta = 8\%$, a capital share $1-\alpha = 1/3$, log utility ($\gamma = 1$). This yields the following graphs:

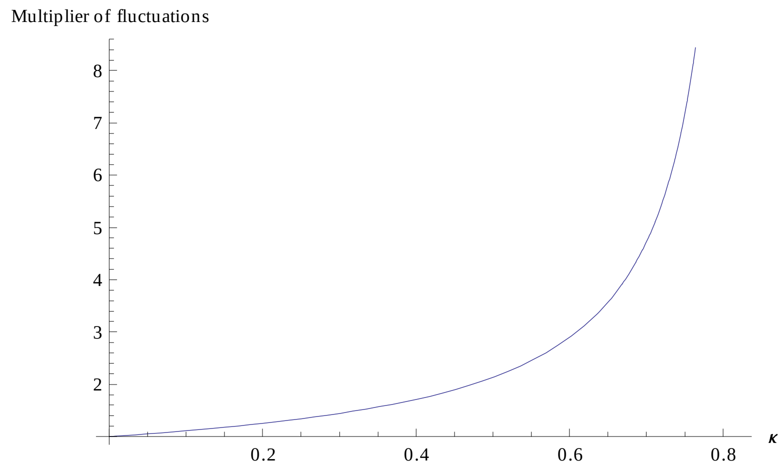

Figure 5: This Figure plots the "multiplier of fluctuations" as a function of the cost of rationality, κ. This is capital's average squared deviation from its mean under the boundedly rational model, divided by the same quantity under the rational model.

Figure 4 plots the speed of mean-reversion, ϕ, of fluctuations, as a function of the cost of rationality, κ. At $\kappa = 0$, we have the rational persistence level. We see that the impact can be quite high.

Figure 5 plots the "multiplier of fluctuations," $M(\kappa)$. We see that the impact can be substantial indeed.

5 Partial Failure of Ricardian equivalence

Intuitively, a sparse agent will violate Ricardian equivalent. I study more the dynamics of that violation.

For simplicity, we use continuous time. The interest rate is $r = -\ln \beta$. The government needs to collect a present value of G/r. This could be done by taxing the population (of size normalized to 1) by $H = Ge^{rT}$, starting at a period T.[4] Hence, the path of taxes is: 0 for $t < T$, and H for $t \geq T$.

What is a consumer's response at time $t < T$? If the consumer is perfectly attentive, then he should start saving at time 0. However, a sparse agent might not pay attention to those future taxes increases, and start cutting on consumption only later, or indeed perhaps just when the tax cuts are enacted.

Let us analyze this more in detail. At T, the tax H is enacted, so that for $t \geq T$, the agent is aware of it. This yields: $\widehat{c}_t = r\widehat{w}_t - H$.

Before the enactment of taxes ($t < T$), will the consumer think of the tax H? That tax

[4] If taxes are collected later, then to guarantee the same present value, they need to be larger by a factor e^{rT}.

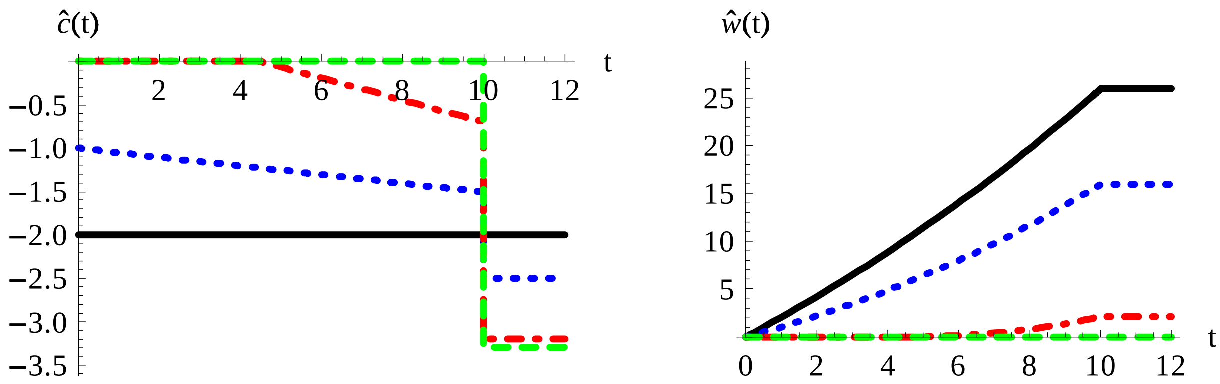

Figure 6: Reaction of consumption and wealth to an increase of future taxes, for different level of κ. *Notes.* At time 0, it is announced that taxes will be paid start at time $T = 10$. This Figure plots the change in consumption and wealth. The solid line is the prediction of the rational model (i.e.$\kappa = 0$), the other lines the reaction for different value of κ ($\kappa = 0.01$ (blue, dotted), $\kappa = 0.025$ (red, dashed-dotted), $\kappa = .1$ (green, dashed)). The very BR agents does not react at first, but starts reacting when he is closer to T. He reacts even more when taxes are in effect. As he delayed his savings, he needs to cut more on consumption when taxes start. Units are percentage points of previous steady state consumption. The amount is $G = 2\%$ of permanent income.

lowers the present value of his income by $He^{-r(T-t)}$, so the consumer's response is:

$$\widehat{c}_t = r\widehat{w}_t - \tau\left(He^{-r(T-t)}, \kappa\right)$$

Hence, the consumer will not think about the tax increase H when $He^{-r(T-t)} \leq \kappa$. Call $s \in [0, T)$ the first moment when he thinks about them (if it exists, i.e. if $H > \kappa$), otherwise we set $s = T$.

The next Proposition details the dynamics.

Proposition 11 (Myopic behavior and failure of Ricardian equivalence) *Suppose that taxes will go up at time T. While a rational agent would cut consumption at time 0, a sparse agent cuts consumption later, at a time $s = \max\left(0, \min\left(T, \frac{1}{r}\ln\frac{\kappa}{He^{-rT}}\right)\right)$. His consumption path is:*

$$\widehat{c}_t = \begin{cases} 0 & \text{for } t < s \\ -He^{-r(T-s)} + \kappa\left(1 - r\left(t-s\right)\right) & \text{for } s \leq t < T \\ r\widehat{w}_T - H & \text{for } t \geq T \end{cases}$$

with $\widehat{w}_T = \frac{H}{r}\left(1 - e^{-t(T-s)}\right) - \kappa\left(T - s\right).$

Let us take an example illustrated in Figure 6, with $r = 5\%$, $G = 2\%$, $T = 10$ years. This Figure plots the change in consumption and wealth for the rational actor $\kappa = 0$ (black, solid), and progressively less rational agents: $\kappa = 0.01$ (blue, dotted), $\kappa = 0.025$ (red, dashed-dotted), $\kappa = .1$ (green, dashed). The traditional Ricardian consumer ($\kappa = 0$) immediately decreases his consumption by 2%, which leads to wealth accumulation at until time T. In constrast the very BR consumer ($\kappa = 0.1$) doesn't react at all until $T = 10$ (hence he doesn't accumulated any wealth), and then cuts a lot on consumption. The value $\kappa = 0.01$ and $\kappa = 0.025$ display an intermediary behavior. For $\kappa = 0.025$, the consumer initially doesn't pay attention to the future tax. However, at a time $s = 4.5$ years, (i.e., when there are 3.6 years remaining until the taxes are effective), he starts paying attention, and starts savings for the future taxes. As the tax looms larger, the agent saves more. As the agent delayed his savings, he ends up cuttings down on consumption more drastically when taxes are in effect.

Smaller taxes generate a more delayed reaction. Controlling for the PV of taxes, consumers are better off with early rather than delayed taxes (as this allows them to smooth more).

6 Discussion

6.1 Active decision: Consumption or Savings?

Here we assume that the active decision was one of consumption. One could imagine that it would be in savings. Does this matter? First, for many variables, it does matter: the impact of interest rates, future taxes, future income shocks etc are the same whether a sparse agent uses the the consumption frame or saving frame. However, the frame does matter for one variable: current income. Indeed, take the permanent-income setup. [5]

Which frame does the agent use? Here, we'll use the working hypothesis that the agent takes the frame that yields the higher expected utility. We use the following Proposition.

Proposition 12 (Welfare under the consumption vs savings frame) *The consumption frame yields greter utility than the savings frame if and only if $\phi_y > r$, i.e. if income shocks mean-revert not too slowly. More precisely, under the "active consumption" frame, the utility loss*

[5] Recall that $\widehat{c}_t^r = \frac{r}{r+\phi}\widehat{y}_t$, so

$$\widehat{c}_t^s = \frac{r}{r+\phi} m\widehat{y}_t \text{ under the consumption frame}$$

However, if the consumer choose savings, S_t, and then consumes $c_t = wy_t - S_t$, the rational amount is $\widehat{S}_t^r = \widehat{y}_t - \widehat{c}_t^r$, i.e. $\widehat{S}_t^r = \frac{\phi}{r+\phi}\widehat{y}_t$. Hence, the savings of a sparse agent is $\widehat{S}_t^s = \frac{\phi}{r+\phi} m\widehat{y}_t$, and the deviation of consumption is: $\widehat{c}_t^s = \widehat{y}_t - \widehat{S}_t^s$, i.e.

$$\widehat{c}_t^s = \left(1 - \frac{m\phi}{r+\phi}\right)\widehat{y}_t \text{ under the savings frame}$$

which is generally not the same as $\widehat{c}_t^s$ under the consumption frame.

from a BR policy is, to the leading order in σ_y^2, $L^C = A(1-m)^2 \phi_y^2$, for $A = \frac{u''(c^d(w))\sigma_y^2}{2(r+2\phi_y)(r+\phi_y)^2}$, while under the "active savings" frame, they are $L^S = A(1-m)^2 r^2$.

When $\phi > r$ (which is probably the relevant case, if business-cycle fluctuations partly mean-revert), the "consumption" frame is indeed better for the agent, at least most of the time. This make sense: savings are there to absorb transitory income shocks, and consumption should be smooth. When the agent chooses consumption in an inattentive manner, it makes consumption quite smooth indeed. However, if the agent chooses savings inattentively, he makes savings smooth, but consumption needs to absorb the shocks, hence is quite volatile. Hence, generally, to keep consumption smooth, choosing consumption inattentively is better than choosing savings inattentively.

However, when income shocks are a random walk ($\phi = 0$), the savings frame is better. An inattentive agent will keep a constant savings, and let consumption react one for one to income shock, which is the normatively correct behavior when income shocks are completely persistent.

It may be useful to see the effect in a simpler context. Take a 3 period model with $\beta = R = 1$, and an income shock with persistence ρ: $\widehat{y}_t = \rho^{t-1}\varepsilon$ for $t = 0, 1, 2$, with ε a mean-0 shock. Normatively, that should induce the change $\widehat{c} = (\widehat{c}_t)_{t=0,1,2} = (1,1,1)\frac{1+\rho+\rho^2}{3}\varepsilon$ (indeed, the total value of income has increased by $(1+\rho+\rho^2)\varepsilon$). Let us not consider a BR agent with $m = 0$. However, under the consumption frame, $\widehat{c}^C = \left(0, \frac{1}{2}, \frac{1}{2} + \rho + \rho^2\right)\varepsilon$ (as there is no reaction of c_0, so that time-1 wealth increases by $\widehat{w}_1 = \varepsilon$, of which 1/2 is consumed at time 1, so $\widehat{c}_1^C = \frac{\varepsilon}{2}$). Under the savings frame, we get $\widehat{c}^S = (1, \rho, \rho^2)\varepsilon$ (savings doesn't change, consumption absorbs all the shocks). It is easy to verify that for ρ small, the utility is higher under the consumption frame, while the opposite for large ρ.[6] Indeed, when $\rho = 0$, $\widehat{c}^C = \left(0, \frac{1}{2}, \frac{1}{2}\right)\varepsilon$ and $\widehat{c}^S = (1,0,0)\varepsilon$, so there is more smoothing under the consumption frame. Other the other hand, with $\rho = 1$, $\widehat{c}^C = \left(0, \frac{1}{2}, \frac{5}{2}\right)\varepsilon$ and $\widehat{c}^S = (1,1,1)\varepsilon$, and there is more smoothing under the savings frame.

7 Conclusion

I presented a practical way to do boundedly rational dynamic programming. It is portable and to the first order has just one free continuous parameter, κ, the penalty for lack of sparsity, which can also be interpreted as a cost of complexity.

It allows us to revisit canonical models in economics, and give them a behavioral flavor.

From the micro point of view, we obtain inattention and delayed response. Those are not necessarily very surprising features – however, it is useful to have clean model that generates those things and can be calibrated. The model could be empirically evaluated, but that would take us too far away.

[6] To the leading order, $\widehat{u} = \frac{1}{2}u''(c^d)E\sum_t \widehat{c}_t^2$, so $\widehat{u}^C = \frac{1}{2}u''(c^d)\sigma_\varepsilon^2\left(\frac{1}{4} + \left(\frac{1}{2}+\rho+\rho^2\right)^2\right)$ and $\widehat{u}^S = \frac{1}{2}u''(c^d)\sigma_\varepsilon^2(1+\rho^2+\rho^4)$. This yields $\widehat{u}^C \geq \widehat{u}^S$ iff $\rho < \rho^* \simeq 0.32$.

From the macro point of view, the model allows us to think about bounded rationality in general equilibrium. The upshot is that compared to the rational model, sparsity leads to larger and more persistent fluctuations. The reason is that rational actors tend to "dampen" fluctuations. For instance, they consume more when more capital is available. This channel is muted with sparse agents. Hence, fluctuations are more persistent, innovations have a longer-lasting effect, and the average fluctuations (deviations from the mean) are larger.

Given that it seems easy to use and sensible, we can hope that this model may be useful for other extent issues in macroeconomics and finance.

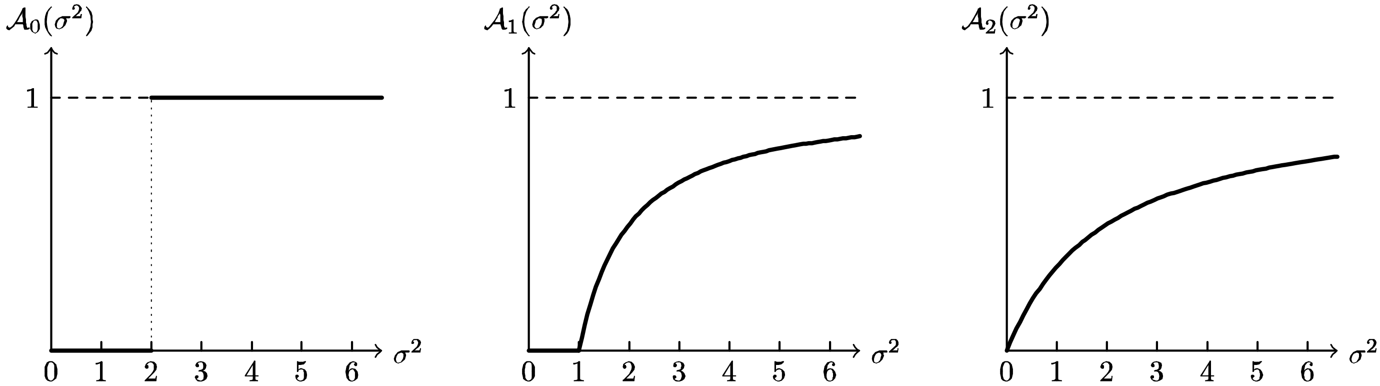

Figure 7: Three attention functions $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$, corresponding to fixed cost, linear cost and quadratic cost respectively. We see that $\mathcal{A}_0$ and $\mathcal{A}_1$ induce sparsity – i.e. a range where attention is exactly 0. $\mathcal{A}_1$ and $\mathcal{A}_2$ induce a continuous reaction function. $\mathcal{A}_1$ alone induces sparsity and continuity.

A Appendix: Attention and Truncation Functions

Here are some good truncation functions. In Gabaix (2013), I study attention functions $\mathcal{A}_\alpha(\sigma^2)$: they are weakly increasing, from 0 (complete inattention) to 1 (full attention). Here I defined the related truncation function τ_α :

$$\tau_\alpha(b,k) := b\mathcal{A}_\alpha\left(\frac{b^2}{k^2}\right)$$

It is the coefficient b, times the attention to the coefficient, divided by the scaled cognition cost k. For instance, for the values $\alpha = 0, 1, 2$, we have (Gabaix 2013):

$$\mathcal{A}_0\left(\sigma^2\right) = 1_{\sigma^2 \geq 2}, \qquad \mathcal{A}_1\left(\sigma^2\right) = \max\left(1 - \frac{1}{\sigma^2}, 0\right), \qquad \mathcal{A}_2\left(\sigma^2\right) = \frac{\sigma^2}{2+\sigma^2} \tag{23}$$

hence the truncation functions $\tau_\alpha(b,k)$:

$$\tau_0(b,k) = b \cdot 1_{b^2 \geq 2k^2}, \qquad \tau_1(b,k) = b\max\left(1 - \frac{k^2}{b^2}, 0\right), \qquad \tau_2(b,k) = \frac{b^3}{b^2+k^2}$$

Figure 7 plots the attention functions, and Figure 8 the corresponding truncation functions.

B Appendix: Tools to Expand a Simple Model Into a More Complex one

Here I develop the method to derive the Taylor expansion of a richer model, when starting from a simpler one. Here the methods are entirely paper and pencil. They draw from the

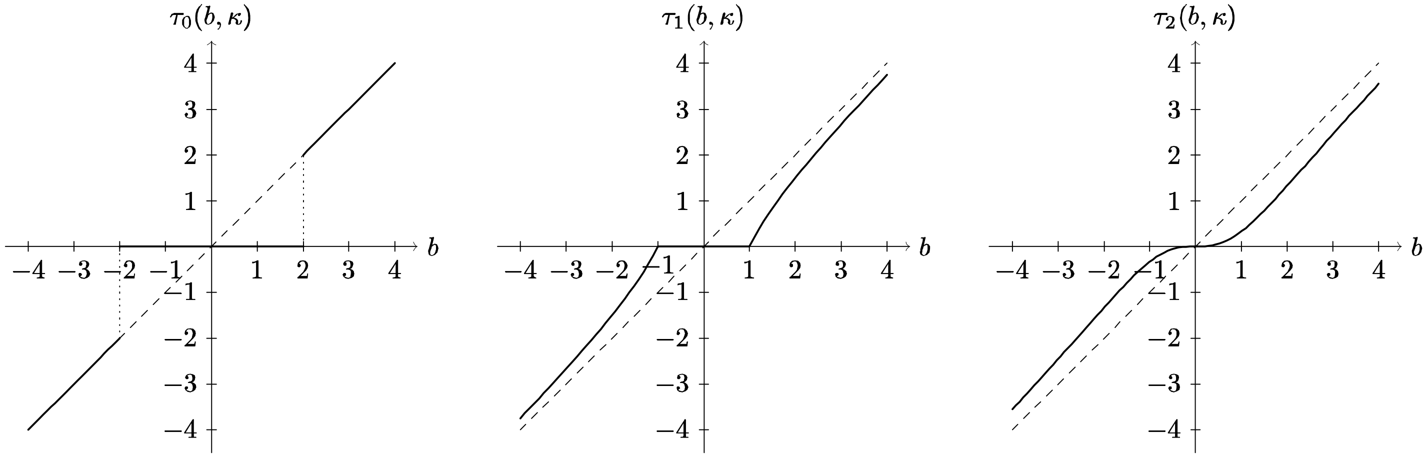

Figure 8: Three truncation functions. Because it gives sparsity and continuity, the τ_1 function is recommended.

techniques surveyed by Judd (1998, Chapter 14), who has a more computer-based perspective.

B.1 General Situation

Consider the fully rational model:

$$V(w,x) = \max_a u(w,x,a) + \beta \mathbb{E} V(w'(a), x')$$

The state variables are w and x, and the decision variable is a. The state variables evolve according to:

$$w' = G^{w'}(w,x,a)$$
$$x' = G^{x'}(x)$$

We start with a simpler model, where $x \equiv 0$, i.e.

$$V^d(w) = \max_a u(w,x,a) + \beta \mathbb{E} V^d(w'(a))$$

where $w' = G^{w'}(w,0,a)$.

Using the notation $D_w f = \partial_w f + \frac{da}{dw}\partial_a f$, which is the total derivative with respect to w (e.g. the full impact of a change in w, including the impact it has on a change in the action a). Differentiating the Bellman equation (first with respect to the new variable x, then with respect to the default variable w), we obtain:

$$V_x(w,x) = u_x + \beta V'_{w'} G_x^{w'}(w,x,a) + \beta V'_{x'} G_x^{x'}$$
$$V_{w,x}(w,x) = D_w u_x + \beta D_w \left[V'_{w'} G_x^{w'}(w,x,a) \right] + \beta G_x^{x'} V'_{w',x'} D_w w'$$

so

$$V_{w,x}(w,0) = \frac{D_w u_x + \beta D_w \left[G_x^{w'}(w,0,a)\, V'_{w'}(w',0)\right]}{1 - \beta G_x^{x'} D_w w'} \tag{24}$$

Proposition 13 *The impact of a change x on the value function is:*

$$V_{w,x}(w,0) = \frac{D_w u_x + \beta D_w \left[G_x^{w'}(w,0,a)\, V'_{w'}(w')\right]}{1 - \beta G_x^{x'} D_w w'} \tag{25}$$

The impact of a change x on the optimal action is:

$$da = -\Psi_a^{-1}\Psi_x dx$$

$$\begin{aligned}
\Psi(a,x) &= u_a(w,a) + \beta V'_{w'} G_a^{w'} \\
\Psi_a &= u_{aa} + \beta G_a^{w'} V'_{w'w'} G_a^{w'} + \beta V'_{w'} G_{aa}^{w'} \\
\Psi_x &= u_{ax} + \beta V'_{w'x} G_a^{w'} + \beta V'_{w'} G_{ax}^{w'}
\end{aligned}$$

They depend only on the transition functions and the derivatives of the simpler baseline value function $V'_{w'}(w')$.

The same procedure can be followed when $x' = G^{x'}(w,x,a)$, with more complex algebra.

B.2 Life-cycle example

We start from the simple life-cycle example. We assume, for simplicity, a stationary environment with no trend growth. The Bellman equation is:

$$V(w,r) = \max_c u(c) + \beta V'((R+r)(w-c) + y', r') \tag{26a}$$

I suppress the expectation operator, as the shocks are assumed to be small. We assume a law of motion:

$$r' = \rho r + \varepsilon'$$

Call next-period wealth w':

$$w' = (R+r)(w-c) + y'$$

We assume that the agent knows the simple model where the interest rate is always at its average, $r \equiv 0$. As is well-known, the optimal policy is $c = rw + y$, and, with $\overline{R} = 1 + \overline{r}$,

$$V(w) = A\left(w + w^H\right)^{1-\gamma} / (1-\gamma),\ w^H = Y/\overline{r}, A = \left(\overline{r}/\overline{R}\right)^{-\gamma}$$

First, we differentiate the Bellman equation with respect to the new variable:

$$\begin{aligned}
V_r(w,r) &= \beta V'_{w'}(w',r') \frac{\partial w'}{\partial r} + \beta V'_{r'}(w',r') \frac{\partial r'}{\partial r} \\
V_r(w,r) &= \beta V'_{w'}(w',r')(w-c) + \beta V'_{r'}(w',r')\,\rho
\end{aligned} \tag{27}$$

Evaluating at $r = 0$, this leads to:

$$V_r(w,0) = V_w^d(w)\frac{\beta(w-c)}{1-\beta\rho}$$

We now take the total derivative with respect to w , $D_w f = \partial_w f + \frac{da}{dw}\partial_a f$, e.g. the full impact of a change in w, including the impact it has on a change in the consumption c. The baseline policy is $c(w) = \overline{r}w/\overline{R} + \overline{y}$, so $D_w c = \overline{r}$, and $D_w w' = d\left(\overline{R}(w-c)\right)/dw = \overline{R} - \overline{R}\overline{r}/\overline{R} = 1$.

$$\begin{aligned} D_w c &= \overline{r}/\overline{R} \\ D_w w' &= 1 \end{aligned}$$

This means that one extra dollar of wealth received today translates into exactly one dollar of wealth next period: its interest income, r, is entirely consumed.

So differentiate (using the total derivative) equation 27. We obtain:

$$\begin{aligned} \beta^{-1}V_{wr}(w,r) &= V'_{w'w'}(w',r')(D_{w'}w)\cdot(w-c) + V'_{w'}(w',r')D_w(w-c) + V'_{w'r'}(w',r')\rho D_w w' \\ &= V'_{w'w'}(w',r')(w-c) + V'_{w'}(w',r')(1-\frac{\overline{r}}{\overline{R}}) + V'_{w'r'}(w',r')\rho \end{aligned}$$

so, using

$$\begin{aligned} V'_{w'w'}(w',r') &= -\gamma V'_w \cdot \frac{1}{w+w^H} = -\gamma V'_w \cdot \frac{\overline{r}}{\overline{R}c} \\ V_{w,r} &= \frac{\beta\frac{V'_{w'}}{\overline{R}}\left(1-\gamma\overline{r}\left(\frac{w-c}{c}\right)\right)}{1-\rho\beta} \end{aligned}$$

Finally, let's derive the impact of a change in r on c : We have

$$V_w = \beta\left(\overline{R}+r\right)V'_{w'} = u'(c)$$

so

$$\begin{aligned} \frac{dc}{dr} &= \frac{V_{wr}}{u''(c)} = \frac{-1}{u''(c)}\frac{V_w}{\overline{R}}\frac{1-\gamma\overline{r}\left(\frac{w}{c}-1\right)}{\overline{R}-\rho_r\beta} \\ &= \frac{-1}{\gamma u'(c)c}\frac{V_w}{\overline{R}}\frac{1-\gamma\overline{r}\left(\frac{w}{c}-1\right)}{\overline{R}-\rho_r} \\ \frac{dc}{c} &= \frac{1}{\overline{R}}\frac{\overline{r}\left(\frac{w}{c}-1\right)-1/\gamma}{\overline{R}-\rho_r}dr \end{aligned}$$

B.3 Continuous time

Calculations are typically cleaner in continuous time, so we develop the continuous-time version of the machinery. We take for now problems without stochastic terms (those should be added later).

The laws of motion are:

$$\dot{w}_t = F^w(w, x, a)$$
$$\dot{x}_t = F^w(w, x)$$

and the Bellman equation is:

$$\rho V(w, x) = \max_a u(w, x, a) + V_w(w, x) F^w(w, x, a) + V_x(w, x) F^x(w, x, a)$$

In the more complex case $\dot{x}_t = F^w(w, x, a)$, we need to solve for a matrix Ricatti equation – but not here.

Call $D_w = \partial_w + a_w \partial_a$ the "total impact" of a change in w. Then:

$$\rho V_x = u_x + V_w F^w_x + V_x F^x_x + V_{xx} F^x \tag{28}$$

Now, we differentiate and evaluated at $x = 0$:

$$\rho V_{wx} = D_w(u_x + V_w F^w_x) + V_{wx} F^x_x + V_x F^x_{wx}$$

so

$$V_x = (\rho - F^x_x)^{-1} [u_x + V_w F^w_x] \tag{29}$$
$$V_{wx} = (\rho - F^x_x)^{-1} [D_w(u_x + V_w F^w_x) + V_x F^x_{wx}] \tag{30}$$

As a satisfies $\Psi = 0$ with

$$\Psi(a, w, x) = u_a + V_w F^w_a$$

Hence, the impact of x on the optimal action is

$$a_x = -\Psi_a^{-1} \Psi_x$$

$$\Psi_a = u_{aa} + V_w F^w_{aa}$$
$$\Psi_x = u_{ax} + V_{wx} F^w_a + V_w F^w_{ax}$$

Calculation of V_{xx}. We now turn to the more difficult case of V_{xx}. Using $D_x = \partial_x + a_x \partial_a$ the "total impact" of a change in x, we have:

$$\begin{aligned}\rho V_x &= D_x u + V_w D_x F^w + V_x F^x_x + V_{xx} F^x \\ &= a_x(u_a + V_w F^w_a) + u_x + V_w F^w_x + V_x F^x_x + V_{xx} F^x\end{aligned}$$

Next, differentiating at $x = 0$,

$$\begin{aligned}\rho V_{xx} &= a_x D_x\left(u_a + V_w F_a^w\right) + D_x\left[u_x + V_w F_x^w + V_x F_x^x\ \right] + V_{xx} F_x^x \\ &= a_x\left[u_{ax} + u_{aa} a_x + V_{wx} F_a^w + V_w F_{ax}^w + V_w F_{aa}^w a_x\right] \\ &\quad + u_{xx} + u_{xa} a_x + V_{xw} F_x^w + V_w D_x F_x^w + 2 V_{xx} F_x^x + V_x F_{xx}^x\end{aligned}$$

hence

$$\begin{aligned}(\rho - 2F_x^x) V_{xx} &= a_x\left[u_{ax} + u_{aa} a_x + V_{wx} F_a^w + V_w F_{ax}^w + V_w F_{aa}^w a_x\right] \\ &\quad + u_{xx} + u_{xa} a_x + V_{xw} F_x^w + V_w D_x F_x^w + V_x F_{xx}^x\end{aligned}$$

This is a bit of a complicated expression. Let us note it can be written

$$\left(\rho - 2F_x^x\right)\left(V_{xx}^s - V_{xx}^r\right) = a_x A + a_x B a_x + C$$

with $B = u_{aa} + V_w F_{aa}^w$.

We use the following elementary Lemma:

Lemma 1 *Let $f(a) = Aa + a'Ba + C$, for B symmetric negative definite. Let $a^* = \arg\max_a f(a)$, so $a^* = -\frac{1}{2}B^{-1}A$. Then, for any a,*

$$f(a) - f(a^*) = (a - a^*) B (a - a^*).$$

Let's compare V_{xx} under the sparse vs rational model: the difference is just in the D_x^r vs D_x^s term. Indeed,

$$D_x^s - D_x^r = \left(a_x^s - a_x^r\right)\partial_a$$

so, using the previous Lemma,

$$V_{xx}^s - V_{xx}^r = \left(\rho - 2F_x^x\right)^{-1}\left(a_x^s - a_x^r\right)\left(u_{aa} + V_w F_{aa}^w\right)\left(a_x^s - a_x^r\right) \tag{31}$$

We gather the results.

Proposition 14 (Difference in value functions) *Consider the value function V^r under the optimal policy and V^s under a potentially suboptimal policy, and $V^\delta(w,x) = V^s(w,x) - V^r(w,x)$. Then, evaluating at $x = 0$, we have:*

$$V^\delta = 0, V_w^\delta = 0, V_{ww}^\delta = 0, V_x^\delta = 0, V_{wx}^\delta = 0$$

and

$$V_{xx}^\delta = \left(\rho - 2F_x^x\right)^{-1}\left(a_x^s - a_x^r\right)\left(u_{aa} + V_w F_{aa}^w\right)\left(a_x^s - a_x^r\right) \tag{32}$$

Equation (32) has an intuitive interpretation. At a point in time, as a function of a, present and continuation utility is $v(a) = u(a, w_t)\,dt + (1 - \rho dt)\,V(w_t + F^w(w_t, a_t)\,dt)$. Hence (omitting for the dt to remove the notational clutter), $v'(a) = u_a + V_w F_a^w$ and $v''(a) = u_{aa} + V_w F_{aa}^w$. Hence, reacting imperfectly to a small x_t (with $a_t^\delta = a_t^s - a_t^r$) creates an instantaneous utility loss of $L_t = -\frac{1}{2} x_t a_x^\delta v_{aa} a_x^\delta x_t$. The full utility loss is the present discounted value of that, i.e.

$$
\begin{aligned}
2L &= \int_0^\infty e^{-\rho t} 2L_t dt = -\int_0^\infty e^{-\rho t} x_t a_x^\delta v_{aa} a_x^\delta x_t \text{ with } x_t = e^{-\phi t} x_0 \\
&= -\int_0^\infty e^{-\rho t} e^{-2\phi t} x_0 a_x^\delta v_{aa} a_x^\delta x_0 = \frac{1}{\rho + 2\phi} x_0 a_x v_{aa} a_x x_0 \\
&= -x_0 \left(\rho - 2F_x^x\right)^{-1} a_x^\delta \left(u_{aa} + V_w F_{aa}^w\right) a_x^\delta x_0 \text{ as } F_x^x = -\phi \\
&= -x_0 V_{xx}^\delta x_0.
\end{aligned}
$$

It is enough to study the "static" utility losses to derive the dynamic utility losses.

C Appendix: Proofs

Proof of Proposition 5 The rational reaction function satisfies:

$$a^r(x) = a^d + \sum_i b_i x_i + \lambda(x)$$

for a function $\lambda(x) = O\left(\|x\|^2\right)$.

So, $\partial a / \partial x_i = b_i$ and:

$$m_i^* = \tau\left(1, \frac{\kappa_a}{\sigma_{x_i} \cdot \partial a / \partial x_i}\right) = \tau\left(1, \frac{\kappa_a}{\sigma_{x_i} \cdot b_i}\right)$$

We shall use the notation $\overline{\lambda}(x) := \lambda\left((m_i^* x_i)_{i=1\ldots n}\right)$, which also satisfies $\overline{\lambda}(x) = O\left(\|x\|^2\right)$. The sparse reaction function is:

$$
\begin{aligned}
a^s(x) &= \arg\max_a u(a, m_1^* x_1, \ldots., m_n^* x_n) \\
&= a^s(m_1^* x_1, \ldots., m_n^* x_n) \\
&= a^d + \sum_i b_i m_i^* x_i + \lambda\left((m_i^* x_i)_{i=1\ldots n}\right) \\
&= a^d + \sum_i b_i \tau\left(1, \frac{\kappa_a}{\sigma_{x_i} \cdot b_i}\right) x_i + \overline{\lambda}(x) \\
&= a^d + \sum_i \tau\left(b_i, \frac{\kappa_a}{\sigma_{x_i}}\right) x_i + \overline{\lambda}(x) \\
&= a^d + \sum_i \tau\left(b_i, \frac{\kappa_a}{\sigma_{x_i}}\right) x_i + O\left(\|x\|^2\right)
\end{aligned}
$$

Proof of Proposition 7 Let us consider two functions U and u^s

$$U^*(a,w,x) := u(a,w,x) + \beta \mathbb{E} V(F^w(w,x,a), F^x(w,x,a))$$
$$U^{**}(a,w,x) := u(a,w,x) + \beta \mathbb{E} V^s(F^w(w,x,a), F^x(w,x,a))$$

and define the associated optimal actions:

$$a^*(w,x) := \arg\max_a U^*(a,w,x), \qquad a^{**}(w,x) := \arg\max_a U^{**}(a,w,x)$$

In U^{**}, there is no inattention: however, the continuation policy V^s is used: the agent will be inattentive in the future.

First, we will prove:

Lemma 2 *Suppose that $F_a^x = 0$. We have, at $x = 0$, $\frac{\partial a^*(w,x)}{\partial x} = \frac{\partial a^{**}(w,x)}{\partial x}$*

Proof. The key fact comes from Proposition 6, and is:

$$V_w(w,0) = V_w^s(w,0)$$
$$V_{ww}(w,0) = V_{ww}^s(w,0)$$
$$V_x(w,x)_{|x=0} = V_x^s(w,x)_{|x=0}$$
$$V_{wx}(w,x)_{|x=0} = V_{wx}^s(w,x)_{|x=0}$$

and

$$U_a^* = u_a(a,w,x) + \beta \mathbb{E}\left[V_w \cdot F_a^w(w,x,a) + V_x \cdot F_a^x(w,x,a)\right]$$

$$\begin{aligned} U_{ax}^* &= u_{ax} + \beta \mathbb{E}\left[F_x^w \cdot V_{ww} \cdot F_a^w + V_w \cdot F_{ax}^w\right] \\ &\quad + \beta \mathbb{E}\left[V_x \cdot F_{ax}^x + F_x^x \cdot V_{xx} F_a^x\right] \end{aligned}$$

Likewise, for U^{**},

$$\begin{aligned} U_{ax}^{**} &= u_{ax} + \beta \mathbb{E}\left[F_x^w \cdot V_{ww}^s \cdot F_a^w + V_w^s \cdot F_{ax}^w\right] \\ &\quad + \beta \mathbb{E}\left[V_x^s \cdot F_{ax}^x + F_x^x \cdot V_{xx}^s F_a^x\right] \end{aligned}$$

Hence, we have

$$U_{ax}^{**} = U_{ax}^* \text{ at } x = 0$$

Note that we used $F_a^x = 0$. This is necessary, because in general $V_{xx} \neq V_{xx}^s$.

Likewise,

$$\begin{aligned} U_{aa}^* &= u_{aa}(a,w,x) + \beta \mathbb{E}\left[F_a^w(w,x,a) \cdot V_{ww} \cdot F_a^w(w,x,a) + V_w \cdot F_{aa}^w(w,x,a)\right] \\ &\quad + 2\beta \mathbb{E}\left[F_a^x(w,x,a) \cdot V_{xw} \cdot F_a^w(w,x,a)\right] \\ &\quad + \beta \mathbb{E}\left[F_a^x(w,x,a) \cdot V_{xx} \cdot F_a^x(w,x,a) + V_x \cdot F_{aa}^x(w,x,a)\right] \end{aligned}$$

and a similar expression for U_{aa}^{**}, which leads to:

$$U_{aa}^{**} = U_{aa}^{*} \text{ at } x = 0$$

Finally, we have:

$$\begin{aligned}
\frac{\partial a^{**}(w,x)}{\partial x}_{|x=0} &= -U_{aa}^{**-1} \cdot U_{ax|x=0}^{**} = -U_{aa}^{*-1} \cdot U_{ax|x=0}^{*} \\
&= \frac{\partial a^{*}(w,x)}{\partial x}_{|x=0}.
\end{aligned}$$

■

Given $a^r(w,x) = a^d(w) + \sum_i b_i(w)\, x_i + O(x^2)$, we have

$$\frac{\partial a^r(w,x)}{\partial x_i} = b_i(w)$$

Hence, the lemma gives:

$$\frac{\partial a^{**}(w,x)}{\partial x_i} = b_i(w)$$

so

$$a^{**}(w,x) = a^d(w) + \sum_i b_i(w)\, x_i + O\left(x^2\right)$$

Finally,

$$\begin{aligned}
a^s(x) &= a^{**}(m_i^* x_i) \\
&= a^d(w) + \sum_i b_i(w)\, m_i^* x_i + O\left(x^2\right) \\
&= a^d(w) + \sum_i b_i(w)\, \tau\left(1, \frac{\kappa_a}{b_i(w)\,\sigma_{x_i}}\right) x_i + O\left(x^2\right) \\
&= a^d(w) + \sum_i \tau\left(b_i(w), \frac{\kappa_a}{\sigma_{x_i}}\right) x_i + O\left(x^2\right).
\end{aligned}$$

Proof of Proposition 9 When $\phi > 0$, we saw that

$$\phi = \left(\frac{\xi}{r+\phi} - \overline{\kappa}\left(r + \frac{\xi}{r+\phi}\right)^2 \frac{r+\phi}{\xi}\right)$$

Let $\psi := \dfrac{r+\phi}{\xi} \neq 0$. Then

$$\phi = \psi^{-1} - \kappa(r + \psi^{-1})^2 \psi,$$

which is equivalent to

$$\begin{aligned}\psi(\xi\psi - r) = \psi\phi &= 1 - \kappa[(r + \psi^{-1})\psi]^2 \\ &= 1 - \kappa(r\psi + 1)^2 \\ &= 1 - \kappa(r^2\psi^2 + 2r\psi + 1).\end{aligned}$$

Rearranging yields

$$(\xi + \kappa r^2)\psi^2 + (2\kappa - 1)r\psi + (\kappa - 1) = 0.$$

The quadratic formula then gives

$$\psi = \frac{(1 - 2\kappa)r \pm \sqrt{\Delta}}{2(\xi + \kappa r^2)},$$

where

$$\begin{aligned}\Delta &= [(2\kappa - 1)r]^2 - 4(\xi + \kappa r^2)(\kappa - 1) \\ &= r^2\left[(2\kappa - 1)^2 - 4\kappa(\kappa - 1)\right] + 4\xi(1 - \kappa) \\ &= r^2\left[(4\kappa^2 - 4\kappa + 1) - (4\kappa^2 - 4\kappa)\right] + 4\xi(1 - \kappa) \\ &= r^2 + 4\xi(1 - \kappa).\end{aligned}$$

In the case $\kappa = 0$, the correct root is the higher one for ψ (i.e., it's the higher root of $\phi = \frac{\xi}{r+\phi}$, the one with the $+\sqrt{\Delta}$ sign). Hence, $\psi = \dfrac{(1 - 2\kappa)r + \sqrt{\Delta}}{2(\xi + \kappa r^2)}$

Finally,

$$\begin{aligned}\phi &= \xi\psi - r \\ &= \frac{\xi\left[(1 - 2\kappa)r + \sqrt{\Delta}\right] - 2(\xi + \kappa r^2)r}{2(\xi + \kappa r^2)} \\ &= \frac{[\xi(1 - 2\kappa) - 2(\xi + \kappa r^2)]\, r + \xi\sqrt{\Delta}}{2(\xi + \kappa r^2)} \\ &= \frac{-\left[2\kappa r^2 + 2\xi\kappa + \xi\right] r + \xi\sqrt{\Delta}}{2(\xi + \kappa r^2)} \\ &= \frac{-\left[2\kappa r^2 + 2\xi\kappa + \xi\right] r + \xi\sqrt{r^2 + 4\xi(1 - \kappa)}}{2(\xi + \kappa r^2)}\end{aligned}$$

Proof of Proposition 12 We use the content[7] and notations of Proposition 14. We set $x_t = \widehat{y}_t$. We have $F^w(w, x, c) = rw + x_t - c_t$ and $F^x(w, x) = -\phi x$.

[7] We could also draw on the results in Cochrane (1989), with a variety of adjustments. Proposition 14 extend Cochrane's results (derive for consumption) to general dynamic problems.

Under the *consumption frame*, $a_t = c_t$, and $F^w_{aa} = 0$, so by Proposition 14, noting $\left[V^\delta_{xx}\right]^C$ the value of $V^\delta_{xx}(w,0)$ under the consumption frame:

$$\left[V^\delta_{xx}\right]^C = \frac{u''(c)}{r+2\phi_y}\left(c^s_y - c^r_y\right)^2 \tag{33}$$

and as $c^s_y = mc^r_y$ with $c^r_y = \frac{r}{r+\phi}$,

$$\left[V^\delta_{xx}\right]^C = \frac{u''(c)}{r+2\phi}(1-m)^2\left(\frac{r}{r+\phi}\right)^2$$

and the expected losses are (with $\sigma^2_y = E\left[\widehat{y}^2_t\right]$):

$$\begin{aligned} L^C &= \frac{-1}{2}\left[V^\delta_{xx}\right]^C \sigma^2_y = \frac{-1}{2}\frac{u''(c)\,\sigma^2_y}{r+2\phi}(1-m)^2\left(\frac{r}{r+\phi}\right)^2 \\ &= A(1-m)^2 r^2 \end{aligned}$$

Under the savings frame, a_t is savings, so $F^w = a_t$, and $c_t = rw_t + x_t - a_t$. Hence:

$$\left[V^\delta_{xx}\right]^S = \frac{u''(c)}{r+2\phi}\left(S^s_y - S^r_y\right)^2$$

and as $S^s_y = mS^r_y$, with $S^r_y = 1 - c^r_y = \frac{\phi}{r+\phi}$,

$$\left[V^\delta_{xx}\right]^S = \frac{u''(c)}{r+2\phi}(1-m)^2\left(\frac{\phi}{r+\phi}\right)^2$$

and expected losses are:

$$L^S = \frac{-1}{2}\left[V^\delta_{xx}\right]^S \sigma^2_y = A(1-m)^2\phi^2$$

Losses from a general variable x. Using the same reasoning, the losses from not paying attention to a variable x is:

$$L^x = \frac{-u''(c)}{r+2\phi}\sigma^2_x\left(c^s_x - c^{rat}_x\right)^2 = \frac{-u''(c)}{r+2\phi}\sigma^2_x c^2_x(1-m_x)^2$$

We parametrize the losses by the "equivalent permanent tax" λ^x such that $L^x = E\int_0^\infty e^{-\rho t}\left[u(c_t) - u(c_t\right.$
Hence, using a Taylor expansions, $\lambda^x = \frac{L}{u'(c)c/r}$. This gives:

$$\lambda^x = \frac{1}{2}\frac{-\frac{u''(c)}{u'(c)c/r}}{r+2\phi}\sigma^2_x c^2_x(1-m_x)^2$$

i.e., using $\gamma = \frac{-cu''(c)}{u'(c)}$,

$$\lambda^x = \frac{1}{2}\frac{r\gamma}{r+2\phi}\left[\frac{c_x\sigma_x}{c}(1-m_x)\right]^2 \tag{34}$$

Proposition 15 *The losses from paying only attention m_x to variable x, expressed in terms of an "equivalent proportional losses in consumption", λ^x are:*

$$\lambda^x = \frac{1}{2}\frac{r\gamma}{r+2\phi_x}\left[\frac{c_x\sigma_x}{c}(1-m_x)\right]^2 \tag{35}$$

where σ_x is the standard deviation of x, and $c_x = \frac{\partial c}{\partial x}$.

The calibration gives:

$$\lambda^r = (1-m_r)^2 \times 0.03\%, \qquad \lambda^y = (1-m_y)^2 \times 3.0\% \tag{36}$$

Proof of Proposition 11 Taxes lower the present value of his income by $He^{-r(T-t)}$, so the consumer's response is:

$$\widehat{c}_t = r\widehat{w}_t - \tau\left(He^{-r(T-t)}, \kappa\right)$$

so wealth accumulation is: $\frac{d}{dt}\widehat{w}_t = r\widehat{w}_t - \widehat{c}_t = \tau\left(He^{-r(T-t)}, \kappa\right)$. The consumer starts thinking about it at a time s s.t. $He^{-r(T-s)} = \kappa$ (assuming that the solution is in $(0,T)$), i.e.

$$s = \max\left(0, \min\left(T, \frac{1}{r}\ln\frac{\kappa}{He^{-rT}}\right)\right) \tag{37}$$

First, consider the case: $s < T$.

Then, for $t \in [s, T)$,

$$\begin{aligned}
\frac{d}{dt}\widehat{w}_t &= \tau\left(He^{-r(T-t)}, \kappa\right) = He^{-r(T-t)} - \kappa \\
\widehat{w}_t &= \int_s^t \left(He^{-r(T-t')} - \kappa\right) dt' \\
\widehat{w}_t &= \frac{H}{r}e^{-rT}\left(e^{rt} - e^{rs}\right) - \kappa(t-s)
\end{aligned}$$

$$\begin{aligned}
\widehat{c}_t &= r\widehat{w}_t - \tau\left(He^{-r(T-t)}, \kappa\right) \\
&= r\left(\frac{H}{r}e^{-rT}\left(e^{rt} - e^{rs}\right) - \kappa(t-s)\right) - \left(He^{-r(T-t)} - \kappa\right) \\
\widehat{c}_t &= -He^{-r(T-s)} + \kappa(1 - r(t-s))
\end{aligned} \tag{38}$$

So at $t = T$

$$\widehat{w}_T = \frac{H}{r}\left(1 - e^{-r(T-s)}\right) - \kappa(T-t)$$

At T, the tax H is enacted, so that for $t \geq T$, the agent is aware of it. This yields:

$$\begin{aligned}
\widehat{c}_t &= r\widehat{w}_t - H \\
\frac{d}{dt}\widehat{w}_t &= r\widehat{w}_t - H - \widehat{c}_t \ = \text{investment income - taxes - consumption change} \\
&= 0
\end{aligned}$$

hence for $t > T$, $\widehat{w}_t = \widehat{w}_T$, and $\widehat{c}_t = r\widehat{w}_T - H$.

We conclude that consumption is:

$$\widehat{c}_t = \begin{cases} 0 & \text{for } t < s \\ -He^{-r(T-s)} + \kappa\left(1 - r\left(t - s\right)\right) & \text{for } s \leq t < T \\ r\widehat{w}_T - H & \text{for } t \geq T \end{cases}$$

and wealth is

$$\widehat{w}_t = \begin{cases} 0 & \text{for } t < s \\ \frac{H}{r}e^{-rT}\left(e^{rt} - e^{rs}\right) - \kappa\left(t - s\right) & \text{for } s \leq t \leq T \\ \frac{H}{r}\left(1 - e^{-t(T-s)}\right) - \kappa\left(T - s\right) = \widehat{w}_T & \text{for } t \geq T \end{cases}$$

References

Abel, Andrew, Janice C. Eberly, and Stavros Panageas, "Optimal Inattention to the Stock Market with Information Costs and Transactions Costs," *Econometrica*, forth.

Acemoglu, Daron. 2008. *Introduction to Modern Economic Growth*, Princeton U. Press.

Akerlof, George A., and Janet L. Yellen, "Can Small Deviations from Rationality Make Significant Differences in Economic Equilibria?" *American Economic Review*, 75 (1985), 708-720.

Alvarez, Fernando, Luigi Guiso and Francesco Lippi. 2012. "Durable consumption and asset management with transaction and observation cost," *American Economic Review*, 102: 2859-96.

Baker, Malcolm, Stefan Nagel and Jeffrey Wurgler, "The Effect of Dividends on Consumption," *Brookings Papers on Economic Activity*, 38 (2007), 231-292.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance*, 59: 1481–1509.

Barro, Robert. 2009. "Rare Disasters, Asset Prices, and Welfare Costs." *American Economic Review* 99: 243–64.

Blanchard, Olivier and Stanley Fischer. 1989. *Lectures on Macroeconomics.* MIT Press

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk," *Quarterly Journal of Economics*, 127: 1243-85.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. "Salience and Consumer Choice", Working paper, Harvard, 2013.

Caballero, R.J., 1995. Near-rationality, heterogeneity, and aggregate consumption. *Journal of Money, Credit and Banking* 27 (1), 29–48.

Campbell, John Y., and N. Gregory Mankiw. 1989. "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence." *NBER Macroeconomics Annual* 1989, 185-216.

Campell, John and Luis Viceira. 2012. *Strategic Asset Allocation*, Oxford University Press.

Candès, Emmanuel, and Terence Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, 52 (2006), 5406-5425.

Chetty, Raj. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica.*

Cochrane, John. 1989. "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-RationalAlternatives" *American Economic Review*, 79(3): 319-337.

Donoho, David, "Compressed Sensing," *IEEE Transactions on Information Theory*, 52 (2006), 1289-1306.

Fuster, Andreas, David Laibson, and Brock Mendel. 2010. "Natural Expectations and Macroeconomic Fluctuations," *Journal of Economic Perspectives*, 24: 67-84.

Gabaix, Xavier. 2012. "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127(2), 645-700.

Gabaix, Xavier. 2013. "A Sparsity-Based Model of Bounded Rationality, Applied to Basic Consumer and Equilibrium Theory," Working Paper, NYU.

Gabaix, Xavier, and David Laibson. 2002. "The 6D bias and the Equity Premium Puzzle." *NBER Macroeconomics Annual*, 16: 257–312.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics*, 121: 505–40.

Gennaioli, Nicola, and Andrei Shleifer, "What Comes to Mind," *Quarterly Journal of Economics*, 125 (2010), 1399-1433.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Financial Innovation and Financial Fragility," Working Paper, Harvard University, 2010.

Hall, Robert E. 1988. "Intertemporal Substitution in Consumption." *Journal of Political Economy* 96: 339–57.

Hellwig, Christian and Venky Venkateswaran. 2012. "Hayek Vs Keynes: Incomplete Information and Market Prices", Working Paper, Toulouse.

Judd, Kenneth. 1998. *Numerical methods in economics.* MIT Press.

Koszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics* 128 (1): 53-104.

Krusell, Per and Anthony Smith. 1996. "Rules of Thumb in Macroeconomic Equilibrium: A Quantitative Analysis." *Journal of Economic Dynamics and Control*, 1996.

Krusell, Per and Anthony Smith. 1998. "Income and wealth heterogeneity in the macroeconomy," *Journal of Political Economy*, 106, 867-896.

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

Maćkowiak, Bartosz and Mirko Wiederholt. 2012. "Inattention to Rare Events." Working Paper, Northwestern.

Mankiw, N. Gregory and Ricardo Reis "Pervasive Stickiness," *American Economic Review Papers and Proceedings*, 96 (2), 164-169, 2006.

Matejka, Filip and Christopher Sims, "Discrete actions in information-constrained tracking problems", Working Paper, Princeton, 2011.

Rabin, Matthew and Georg Weizsäcker, "Narrow Bracketing and Dominated Choices," *American Economic Review*, 99 (2009), 1508-1543

Reis, Ricardo, "Inattentive Consumers," *Journal of Monetary Economics*, 53 (2006), 1761-1800.

Romer, David. 2012. *Advanced Macroeconomics.* McGraw-Hill.

Sims, Christopher. 2003. "Implications of Rational Inattention," *Journal of Monetary Economics*, 50: 665–690.

Thaler, Richard H. 1985. "Mental Accounting and Consumer Choice," *Marketing Science*, 4: 199-214.

Veldkamp, Laura, *Information Choice in Macroeconomics and Finance* (Princeton: Princeton University Press, 2011).

Woodford, Michael. 2010. "Robustly Optimal Monetary Policy with Near-Rational Expectations," *American Economic Review* 100: 274-303.

Woodford, Michael. 2012. "Inattentive Valuation and Reference-Dependent Choice," Working Paper, Columbia.