Sparse Dynamic Programming and Aggregate Fluctuations

Xavier Gabaix*

March 14, 2013
Very 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. Endogenously, the consumer pays little attention to 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. The model is then applied to a Merton-style portfolio choice problem. This problem is usually quite complex and formidable. We see how a sparsity agent will handle the problem, and 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 larger aggregate fluctuations.

*xgabaix@stern.nyu.edu. I thank David Laibson for a great many conversations about bounded rationality. For useful comments I thank Ali Lazrak, Thomas Sargent and seminar participants at Chicago, the Federal Reserve Board, Penn, NYU, and Yale. I am grateful to INET 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 first in microeconomic contexts, in canonical consumption-investment problems. Then, this framework is used to study macroeconomic implications of bounded rationality. A main 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 against traditional economic models is the potential unrealism of the infinitely forward-looking agent who computes the whole equilibrium in her own head. This unrealism 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 tastes (e.g. prospect theory or hyperbolic discounting) or the beliefs (e.g. overconfidence), but keeps the rationality. When tackling the rationality, there is much less agreement and no dynamic alternative to the traditional model has really emerged. This paper attempts to propose 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 in 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 stochastic interest rate and income. In the rational model, he would solve a complex DP problem with three state variables (wealth, income, interest rate). This is a complex problem that requires a computer to solve it.

How will a BR agent do? I assume that the agent starts with a much simpler model, where interest rate and income are constant – that's his 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 BR version, he considers parsimonious enrichments to the value function, as in a Taylor expansion. He asks, for each component, if it will matter enough for his decision. If a given feature (say, the interest rate), is small enough compared to a 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 perhaps no attention at all to the interest rates. This does seem realistic.

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

One application is a Merton-style dynamic portfolio choice problem, i.e. allocating one's wealth between stocks and bonds when the equity is stochatic 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 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 only partially, or not at all, take into those complex features. This may be a more satisfying description of what people do in a complex environment than the hyperrational model. At the very least, it is important to have a concrete alternative to that hyperrational 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 parameters, but can be overturned for extreme parameters.

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 capital to revert pretty fast to its mean. However, in a BR economy, investors will not fully pay attention to that extra 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 from trend of capital. 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 both shocks to 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), 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 that of a consumption-saving problem. Once it is well understood, I formulate the more general notation of BRDP, in section 3, which also treats the Merton portfolio problem. Section 4 uses this framework to study a general equilibrium situation. If formulates and illustrates the amplifying effect of BR on aggregate fluctuations. Section 5 presents an application to the (failure of) Ricardian equivalent. Section 6 concludes. The Appendix contains the more technical material and derivations.

2 Partial equilibrium: BRDP 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 DM'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 $\max_c V(c, \widehat{r}_t, \widehat{y}_t)$. 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 do: $\max_c V(c, \widehat{r}_t, \widehat{y}_t)$. What will a BR consumer do? Using a mix of psychological and economic reasoning, I propose in Gabaix (2012) a reasonably systematic way of handling that. The DM 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) by its truncated version: the interest rate perceived by a BR 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

$$\tau(\mu, \kappa') = (|\mu| - |\kappa'|)_+ \, sign(\mu) \tag{2}$$

is represented in Figure 1. $\widehat{r}_t^s$ is the deviation of the interest rate from its default perceived by a BR 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 BR agent does $\max_c V(c, \widehat{r}_t^s, \widehat{y}_t^s)$.

Step 2 is unproblematic: given the perceived interest rate and income, the DM optimizes consumption. The nerve of the model is in Step 1. To interpret rule (1), note that it implies: "Replace the interest rate by 0 if taking the interest rate into account changes consumption

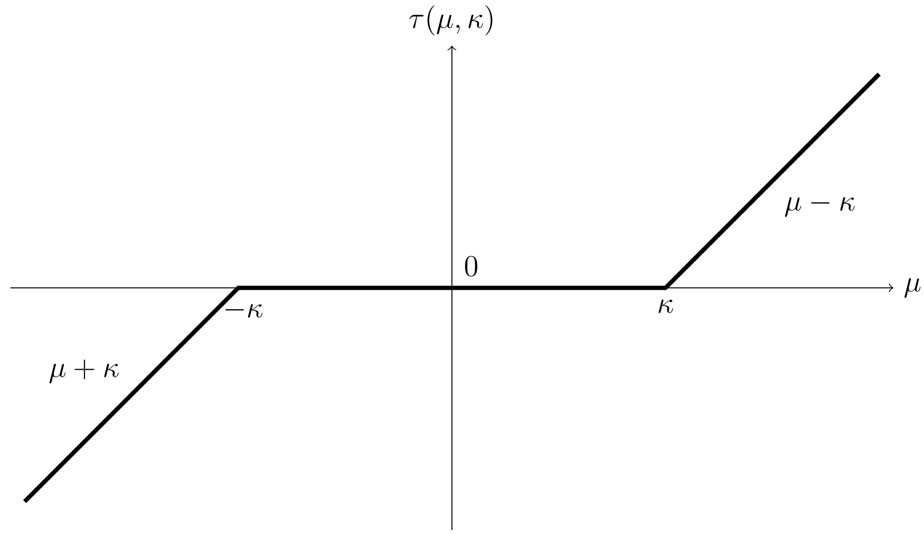

Figure 1: The anchoring-and-adjustment function τ

by less than κ standard deviations, i.e., if $\left|\frac{\partial c}{\partial r}\sigma_r\right| < \kappa\sigma_c$."

That means: on average, a one-standard-deviation change in the interest rate makes the BR 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 DM 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 wrinkles to 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).

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

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

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 wrinkles of 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 a DM 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 $\tau^H(\mu, \kappa) = \mu 1_{|\mu|>|\kappa|}$ instead: if $|\mu| > \kappa$, then this is the 45 degree line, $\tau^H(\mu, \kappa) = \mu$. 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 got 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 mean, respectively, and follow AR(1)s:

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

$\varepsilon_{t+1}^{\cdot}$ are mean-0 disturbances. Hence, wealth follows:

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

What will be the consumption function $c(w_t,\widehat{y}_t,\widehat{r}_t)$ of a BR agent? It is difficult, because this is a dynamic-programming problem with 3 state variables, and has no closed forms. 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: DP 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, 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, one finds:

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

Then, we assume that the BR agent somehow has cognitive access to b_r and b_y: while it may seem counterintuitive, this merely represents that the BR 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 DM does a BR truncation of (3), according to formula (1). Hence, we obtain the following.

Proposition 1 *A BR agent has the following consumption policy:*

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

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).

Formula (5) shows a "feature-by-feature" truncation. It is useful because it embodies in a compact way the policy of a BR agent in a quite complicated world. Note that the agent can do that 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 BRDP. 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 BR, and does not react to any variable. Hence, we have a simple, smooth way to parametrize the agent, from very BR to (essentially) 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 $\rho_r = 0.7$ with yearly units:

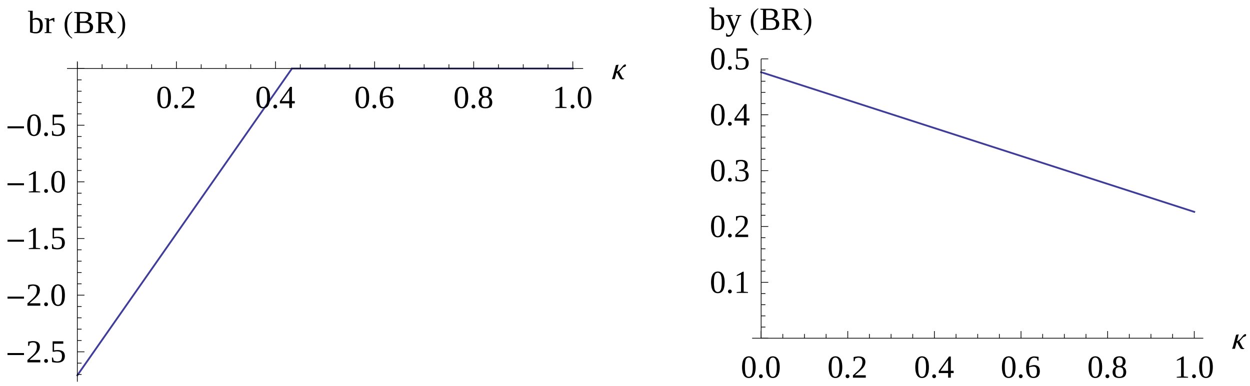

Figure 2: Impact of a change in the interest rate (resp. income) on consumption, as the function of weight on sparsity, κ. $\kappa = 0$ is the rational-agent model.

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 BR dynamic programming. In some cases, this is simpler than the rational way (as the agent does not need to solve for the equilibrium), and this may also be more sensible.

Consequence. A behavioral solution to Puzzles and controversies around the intertemporal elasticity of substitution (IES) 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 less than 1 (e.g., Hall 1988). I show how this may be due to the way econometricians proceed, by fitting the Euler equation, which yields $\ln c_{t+1} - \ln c_t = \frac{\widehat{\psi}}{R} r_t +$ constant, where $\widehat{\psi}$

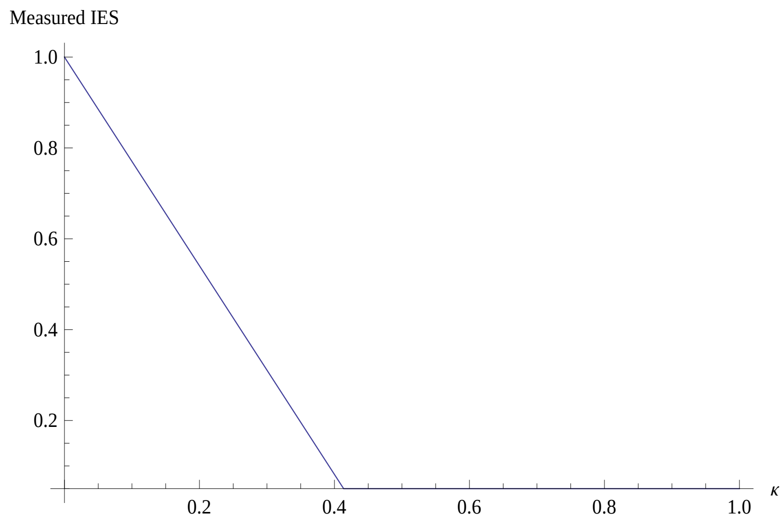

Figure 3: Measured IES $\widehat{\psi}$ if the consumer is boundedly rational with sparsity cost κ.

is the measured IES. If the consumer "underreacts 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 BR will estimate a downward biased IES (intertemporal elasticity of substitution):*

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

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

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

[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.

2.4 Application: Source-dependent Marginal Propensity to Consume

The agent has initial wealth w, future income y, he 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 $\max_c V = u(c) + \mathbb{E}[v(y + R(w-c))]$, 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 DM 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:*

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

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

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 A sparse Max operator in static contexts

One result from Gabaix (2012) is the sparse max or smax operator, which is a sparse version of the max operator. The agent faces a maximization problem which is, it its rational version, $\max_a u(a,\mu,x)$. We posit a procedure that the agent will follow. Its input are: σ_a for the normal modulus of a, and $\mathbb{P}$ for the distribution from which the x_i are drawn.

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

$$u^s := \underset{a|m^d,\kappa\eta_a,\mathbb{P}}{\text{smax}}\, u(a,\mu,x), \qquad a^s := \arg \underset{a|m^d,\kappa\eta_a,\mathbb{P}}{\text{smax}}\, u(a,\mu,x)$$

are defined by the following procedure: First, calculate the optimally sparse representation of the world:

$$m^* = \arg\max_m \frac{1}{2}(\mu-m)' \mathbb{E}_{\mathbb{P}}\left[\frac{\partial a}{\partial m} u_{aa} \frac{\partial a}{\partial m}\right](\mu-m) - \kappa \sum_i \left|m_i - m_i^d\right| \mathbb{E}_{\mathbb{P}}\left[\left(\frac{\partial a}{\partial m_i}\cdot u_{aa}\cdot \eta_a\right)^2\right]^{1/2} \tag{7}$$

Second, define:

$$a^s = \arg\max_a u(a,m^*,x), \qquad u^s = u(a^s,\mu,x). \tag{8}$$

In the expression above, $\partial a/\partial m := -u_{aa}^{-1}\cdot u_{am}$, and derivatives are evaluated at the default model m^d, $a^d = \arg\max_a u\left(a,m^d,x\right)$. If the arg max *have several elements, then, we pick the elements lead to the highest value of u^s.*

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 7). Then, the agent maximizes over the action a, taking m^* to be the true model.

In practice, the typical case applying the sparse max operator is the following: action a maximizes a utility function $u(a,x)$, where disturbances x might be neglected by the agent. To apply the algorithm, we form the surrogate utility function:

$$\overline{u}(a,m,x) = u(a, m_1 x_1, ..., m_n x_n)$$

form the ideal vector $\mu = (1,...,1)$, and calculate the $\text{smax}_{a|m^d,\eta_a,\mathbb{P}}\, \overline{u}(a,\mu,x)$. It is actually quite simple to calculate. There two versions for $\mathbb{P}$.

1. Attention chosen before seeing the variables. In the second procedure, the DM chooses the weight m_i before seeing the x_i. He simply takes into account their magnitude (as captured by their standard deviation). Formally, the probability $\mathbb{P}$ assumes that the X_i's are uncorrelated, and X_i has a standard deviation σ_i. Then, the solution is very simple to calculate.

2. Attention chosen after seeing the variables. This is the above procedure, but instead of using σ_{x_i}, the agent sets $\sigma_{x_i} = |x_i|$. Hence, the algorithm, when deciding what to truncate, sees the magnitudes of variables but not the sign.

The following two Propositions derive the resulting procedures.

Proposition 4 *When "attention is chosen before seeing the variables", the* smax *operator can be equivalently formulated as:*

$$a^s = \arg \underset{a|\sigma_a,\sigma_{x_i}}{\text{smax}}\ u(a,x), \qquad u^s = \underset{a|\sigma_a,\sigma_{x_i}}{\text{smax}}\ u(a,x)$$

where

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

with

$$m_i^* = \tau\left(1, \frac{\kappa\sigma_a}{\sigma_{x_i}\cdot \partial a/\partial x_i}\right) \tag{9}$$

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

Proposition 5 *When "attention is chosen after seeing the variables", the* smax *operator can be equivalently formulated as:*

$$a^s = \arg \operatorname*{smax}_{a|\sigma_a} u(a,x), \qquad u^s = \operatorname*{smax}_{a|\sigma_a} u(a,x)$$

where

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

with

$$x_i^* = \tau\left(x_i, \frac{\kappa\sigma_a}{\partial a/\partial x_i}\right) \tag{10}$$

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

The proof is in the appendix.

The intuition is the x_i's are truncated. If $|\partial a/\partial x_i|$ 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 6 *In the limit of small* x*, if the rational action is:*

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

then under the procedure with "attention chosen before seeing the variables", the BR action is:

$$a^s(x) = \sum_i \tau\left(b_i, \frac{\kappa\sigma_a}{\sigma_{x_i}}\right) x_i + O\left(x^2\right) \tag{11}$$

and under the procedure with "attention chosen after seeing the variables", the BR action is:

$$a^s(x) = \sum_i \tau\left(b_i x_i, \kappa\sigma_a\right) + O\left(x^2\right) \tag{12}$$

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

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 $O(x^2)$ 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 an 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. function of some noise.

In the BR 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 BR agent.

We define the value function as follows:

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

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

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

Slightly more explicitly, is

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

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

Indeed, we have:

Proposition 7 *For small x, we have:*

$$V^s(w,x) = V(w,x) + x^2\phi(w,x)$$

where $\phi(w,x) \leq 0$ is continuous in (w,x) and twice differentiable in w at $x=0$. In other words, the BR value function and the rational value functions differ only by second order terms in x.

This basically generalizes the envelope's theorem. It implies that

$$V_w^s = V_w, \qquad V_{ww}^s = V_{ww}, \qquad V_x = V_x^s \qquad V_{wx} = V_{wx}^s, \qquad \text{at } x = 0 \tag{13}$$

However, in most situations we have $V_{xx}^s < V_{xx}$.

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

Proposition 8 *(Calculation of the optimal BR 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,*

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

Then, the BR policy is:

$$a^s(w,x) = a^s(w) + \sum_i \tau\left(b_i(w), \frac{\kappa\sigma_a}{\sigma_{x_i}}\right) x_i + O\left(x^2\right) \tag{14}$$

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 terms by terms.

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 $\widehat{\pi}_t$ as, 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 equities return 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 $V(w_t, \widehat{\pi}_t, \chi)$.

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

Proposition 9 *(Behavioral dynamic portfolio choice) The fraction of wealth allocated in equities is, with* $\theta^s = \frac{\overline{\pi}}{\gamma\sigma^2}$

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

while consumption is:

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

using the notations:

$$B = \frac{\left(1 - \frac{1}{\gamma}\right)\theta^s}{\mu + \phi}, \qquad \mu := \psi\rho + (1-\psi)\left(r + \frac{(\overline{\pi}/\sigma)^2}{2\gamma}\right).$$

Proposition 9 predicts the choice of a BR agent. When $\kappa = 0$, it is to a first order 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 BR agent. When κ is larger, then the agent first first, the portfolio choice reflects less the change in the equity premium, $\widehat{\pi}_t$. Also, 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 BR; what will be the impact in general equilibrium? The answer will be:

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 $\mathbb{E}\sum_t \beta^t C_t^{1-\gamma}/(1-\gamma)$. In the aggregate, the capital stock follows:

$$K_{t+1} = F(K_t, L) + (1-\delta) K_t - C_t + \varepsilon_{t+1} \tag{15}$$

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: will a BR economy react compared to a traditional (i.e., Qrational-agent) economy?

This is a textbook example: and can be found in Blanchard Fischer (1989, Chapter 2) and 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 rugs 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 hats notation for the deviation (not in logs) from mean, e.g. $\widehat{K}_t = K_t - K^*$.

The law of motion for capital (15) is, in linearized form:

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

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

$$\widehat{K}_{t+1} = (1-\phi)\,\widehat{K}_t + \varepsilon_{t+1} \tag{17}$$

with a speed of mean-reversion:

$$\phi = b - r. \tag{18}$$

Generally, BR 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 BR economy will be less than in the rational economy*:

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

Therefore, fluctuations mean-revert less fast in the BR economy. That's because consumers respond less to shocks.

Finally, squaring equation (17), 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 BR economy than in the rational economy.*

The above argument give 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:

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

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 by 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, that leads to a speed of mean-reversion:

$$\phi = b - r = \frac{\xi}{r + \phi} \tag{19}$$

Hence, solving via rational expectations imposes:

$$\phi = \frac{-r + \sqrt{r^2 + 4\xi}}{2}. \tag{20}$$

The boundedly rational version The DM has wealth k_t (and we normalize the population to be one, 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:

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

This leads to the optimal policy:

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

This is the Taylor expansion of the rational policy. The BR version is:

$$\widehat{c}_t = r\widehat{k}_t + \tau\left(\frac{\xi}{r+\phi}, \kappa\frac{\sigma_C}{\sigma_K}\right)\widehat{K}_t \tag{21}$$

We need to solve for the equilibrium. Note that σ_C is here an endogenous variable. Call

$$\begin{aligned}
\chi &:= \frac{\xi}{r+\phi} \\
b' &:= r + \tau\left(\chi, \kappa\frac{\sigma_C}{\sigma_K}\right)
\end{aligned} \tag{22}$$

Then, we will have $\widehat{c}_t = b'\widehat{K}_t$. That implies $\sigma_C = b'\sigma_K$, i.e.

$$\frac{\sigma_C}{\sigma_K} = b'$$

Hence, from (22), b' satisfies:

$$b' = r + \tau(\chi, \kappa b') \tag{23}$$

This is a piecewise linear equation. Take the interior region, so that $\chi > \kappa b'$. Then, the equation writes: $b' = r + \chi - \kappa b'$, so

$$b' = \frac{r + \chi}{1 + \kappa} \tag{24}$$

In the interior region, $\chi \leq \kappa b'$, $b' = r$. So, we have:

$$b' = \max\left(\frac{r + \chi}{1 + \kappa}, r\right)$$

Finally, the speed of mean-reversion of capital is:

$$\phi = b' - r = \frac{\chi - \kappa r}{1 + \kappa}$$

$$\phi = \left(\frac{\frac{\xi}{r+\phi} - \kappa r}{1 + \kappa}\right)^{+} \tag{25}$$

Hence, ϕ is the solution of the above equation, which is a quadratic equation in the interior domain. We obtain:

Proposition 10 *Shocks are more persistent in the BR economy. More precisely, the speed of mean-reversion is given by*

$$\phi = \left[\frac{-r\left(1 + 2\kappa\right) + \sqrt{r^2 + 4(1+\kappa)\xi}}{2\left(1 + \kappa\right)}\right]^{+} \tag{26}$$

In particular, ϕ is decreasing in κ, $\phi\left(\kappa = 0\right) = \phi^r$, and $\phi > 0$ iff $\kappa > \xi/r^2$.

As above, the variance of the stocks is:

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

Proposition 11 *Shocks are larger in the BR economy. More precisely, the quadratic deviation from trend in capital, interest rate and GDP is multiplied by $\frac{1-\left(1-\phi^{Rat}\right)^2}{1-\left(1-\phi\right)^2}$, where ϕ is given by (26).*

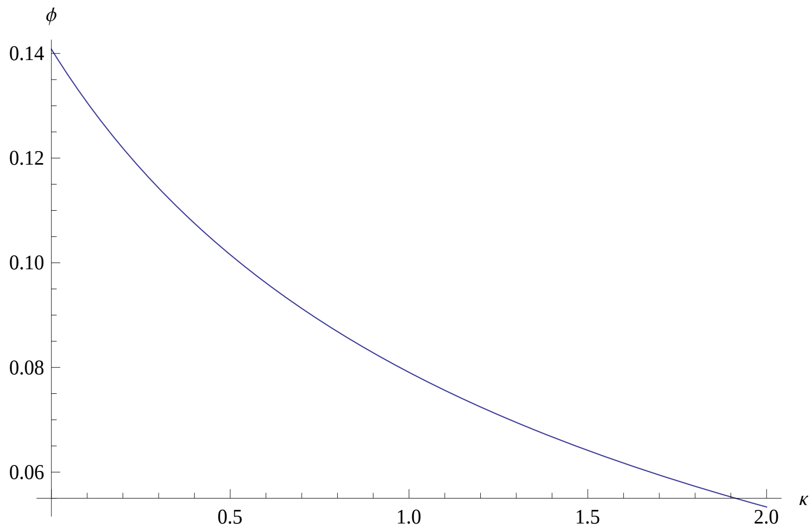

Figure 4: This Figure plots the speed of mean-reversion of fluctuations, ϕ, as a function of the cost of rationality, κ.

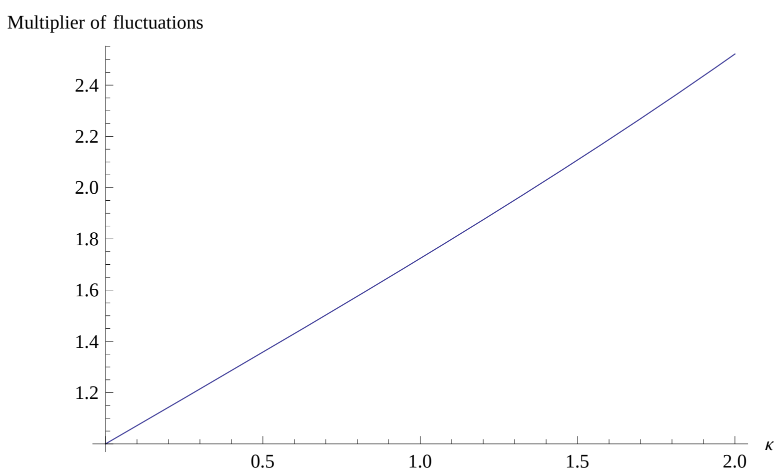

Figure 5: This Figure plots the "multiplier of fluctuations" as a function of the cost of rationality, κ. This is capital's average squared deviation from its mean under the boundedly rational model, divided by the same quantity under the rational model.

Calibration The parametrization is conventional, $F(K) = K^{1-\alpha}/(1-\alpha) - \delta K$, $r = 5\%$, $\delta = 8\%$, a capital share $1-\alpha = 2/3$, log utility ($\gamma = 1$). This yields the following graphs:

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," i.e. it plots $v(\kappa)/v(0)$, where with $v(\kappa) = \mathbb{E}\left[\widehat{K}_t^2\right]$ the dispersion of fluctuations under rationality cost κ. We see that the impacts can be substantial indeed.

5 Ricardian equivalence

Here I study how sparse BR predicts deviations from Ricardian equivalence. The gross interest rate is $R = 1 + r = \beta^{-1}$. The government needs to collect a present value of $G(1+r)/r$. This could be done by taxing the population (of size normalized to 1) by $g(t,T)$, starting at a period T :

$$\begin{aligned} g(t,T) &= 0 \text{ for } t < T \\ &= GR^T := H \text{ for } t \geq T \end{aligned}$$

If taxes are collected later, then to guaranty the same present value, they need to be larger by a factor R^T.

What is a consumer's response at time $t < T$?

If he's 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 the situation in more detail. Taxes lower the PV 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.

$$s = \max\left(0, \min\left(T, \frac{1}{r}\ln\frac{\kappa}{He^{-rT}}\right)\right) \tag{27}$$

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

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

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

$$\begin{aligned}
\widehat{c}_t &= r\widehat{w}_t - \tau\left(He^{-r(T-t)}, \kappa\right) \\
&= r\left(\frac{H}{r}e^{-rT}\left(e^{rt} - e^{rs}\right) - \kappa\,(t-s)\right) - \left(He^{-r(T-t)} - \kappa\right) \\
\widehat{c}_t &= -He^{-r(T-s)} + \kappa\,(1 - r\,(t-s))
\end{aligned} \tag{28}$$

So at $t = T$

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

After that, taxes are paid ever ytime, so the agent is aware of this permanent shock. This yields:

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

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

We summarize the results here:

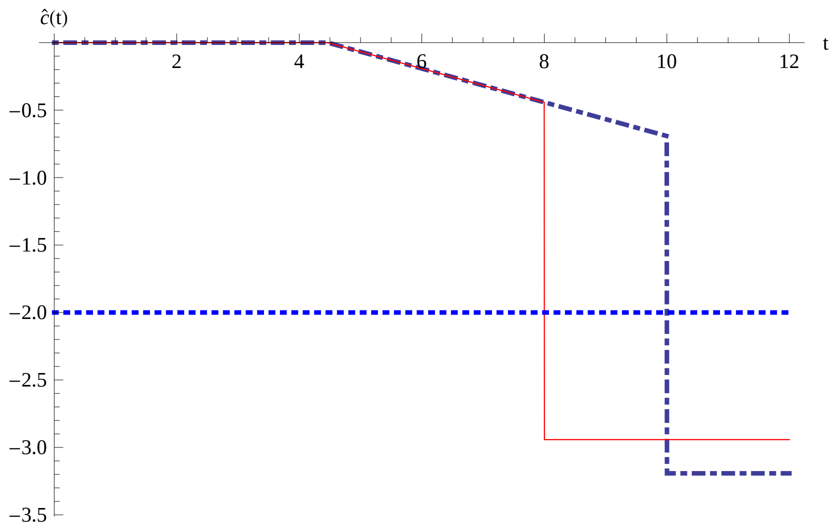

Figure 6: At time 0, it is announced that taxes will be paid start at time T (solid line, $T = 8$, dot-dashed line $T = 10$). This Figure plots the consumption path from $t = 0$ to 12. The dotted line is the prediction of the rational model. The BR agents does not react at first, but starts reacting 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.

Proposition 12 *Suppose that taxes will go up at time T. While a rational agent would cut on consumption at time 0, a sparse agent cuts on 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:*

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

and his wealth is

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

Let us take an example illustrated in Figure 6, with $r = 5\%$, $\kappa = 2.5\%$, and $G = 2\%$. Taxes will start being paid at $T = 8$ in one scenario and at $T = 10$ years. We see that initially, agents don't pay attention to future taxes: the normative impact if that they should reduce consumption by 2%, but that's below the threshold κ. However, at a time $s = 4.4$ years, (i.e., when there are 3.6 years remaining until the taxes are effective), then the impact is $Ge^{rs} = \kappa$, and agents do start paying attention: They starts savings for the future taxes. As taxes loom larger, they save more. When they're effective, they fully take them into account. If $T = 10$ years, the day of reckoning is just postponed. Note that, as agents delayed their savings, they will end up cutting down their consumption more 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 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.

From the macro point of view, the model allows us to think about BR in general equilibrium. The upshot is that compared to the rational model, BR 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 BR 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.

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

A.1 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:

$$V(w,r) = \max_c u(c) + \beta V'((R+r)(w-c) + y', r') \tag{29a}$$

I suppress any expectation, 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 interest is always as 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 w.r.t. the new variable:

$$V_r(w,r) = \beta V'_{w'}(w',r') \frac{\partial w'}{\partial r} + \beta V'_{r'}(w',r') \frac{\partial r'}{\partial r}$$
$$V_r(w,r) = \beta V'_{w'}(w',r')(w-c) + \beta V'_{r'}(w',r') \rho \tag{30}$$

Take the values are $r = 0$, this leads to:

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

We now take the total derivative w.r.t. 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 30. 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

$$V'_{w'w'}(w',r') = -\gamma V'_w \cdot \frac{1}{w+w^H} = -\gamma V'_w \cdot \frac{\overline{r}}{\overline{R}c}$$

$$V_{w,r} = \frac{\beta \frac{V'_{w'}}{\overline{R}}\left(1-\gamma\overline{r}\left(\frac{w-c}{c}\right)\right)}{1-\rho\beta}$$

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

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

so

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

A.2 General Situation

Suppose 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, 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 w.r.t. the new variable x, then w.r.t. 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

$$V_{w,x}(w,0) = \frac{D_w u_x + \beta D_w \left[G_x^{w'}(w,0,a) V'_{w'}(w',0) \right]}{1 - \beta G_x^{x'} D_w w'} \tag{31}$$

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

$$V_{w,x}(w,0) = \frac{D_w u_x + \beta D_w \left[G_x^{w'}(w,0,a) V'_{w'}(w') \right]}{1 - \beta G_x^{x'} D_w w'} \tag{32}$$

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

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

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

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 Appendix: Proofs

Proof of Proposition 4 We have

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial a}{\partial m_i} u_{aa} \frac{\partial a}{\partial m_j}\right] &= \mathbb{E}\left[X_i \frac{\partial a}{\partial x_i} u_{aa} \frac{\partial a}{\partial x_j} X_j\right] \\
&= \frac{\partial a}{\partial x_i} u_{aa} \frac{\partial a}{\partial x_j} \mathbb{E}[X_i X_j] \\
&= \frac{\partial a}{\partial x_i} u_{aa} \frac{\partial a}{\partial x_i} \sigma_i^2 \text{ if } i = j, \text{ or } 0 \text{ if } i \neq j
\end{aligned}$$

Hence,

$$(\mu - m)\, \mathbb{E}\left[\frac{\partial a}{\partial m} u_{aa} \frac{\partial a}{\partial m}\right](\mu - m) = \sum_i \frac{\partial a}{\partial x_i} u_{aa} \frac{\partial a}{\partial x_i} \sigma_i^2 (m_i - \mu_i)^2 .$$

So,

$$\begin{aligned}
m^* &= \arg\max_m \sum_i -\frac{1}{2} \frac{\partial a}{\partial x_i} u_{aa} \frac{\partial a}{\partial x_i} \sigma_i^2 (m_i - \mu_i)^2 - \kappa \sum_i \left|m_i - m_i^d\right| \mathbb{E}\left[\left(X_i \frac{\partial a}{\partial x_i} \cdot u_{aa} \cdot \eta_a\right)^2\right]^{1/2} \\
&= \arg\max_m \sum_i \left\{-\frac{1}{2} \frac{\partial a}{\partial x_i} u_{aa} \frac{\partial a}{\partial x_i} \sigma_i^2 (m_i - \mu_i)^2 - \kappa \left|m_i - m_i^d\right| \left|\sigma_a u_{aa} \frac{\partial a}{\partial x_i} \sigma_i\right|\right\}
\end{aligned}$$

We have n decoupled problems, which makes the problem easy to solve.

It is easy to verify that the solution of

$$\max_{m_i} -A (m_i - \mu_i)^2 - B \left|m_i - m_i^d\right|$$

is, for $A > 0$, $B \geq 0$,

$$m_i = m_i^d + \tau\left(\mu_i - m_i^d, \frac{B}{A}\right)$$

where function τ is given in (2).

Applying this result to $A = -\frac{1}{2}\frac{\partial a}{\partial x_i}u_{aa}\frac{\partial a}{\partial x_i}\sigma_i^2$ and $B = \kappa\left|\sigma_a u_{aa}\frac{\partial a}{\partial x_i}\sigma_i\right|$, and $\mu_i = 1$, we have the announced formula for m^*, with

$$\frac{B}{A} = \frac{\kappa\left|\sigma_a u_{aa}\frac{\partial a}{\partial x_i}\sigma_i\right|}{-\frac{\partial a}{\partial x_i}u_{aa}\frac{\partial a}{\partial x_i}\sigma_i^2} = \frac{\kappa\sigma_a}{\frac{\partial a}{\partial x_i}\sigma_i}$$

Proof of Proposition 6 The rational reaction function satisfies:

$$a^s(x) = a^d + \sum_i b_i x_i + \phi(x)$$

for a function $\phi(x)$ such that $|\phi(x)| \leq C\|x\|^2$ for $\|x\| \leq B$, for some positive B.

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

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

We shall use the notation $\overline{\phi}(x) := \phi\left((m_i^* x_i)_{i=1...n}\right)$ also satisfies that for $\|x\| \leq B$, $\left|\overline{\phi}(x)\right| \leq C\|x\|^2$, hence $\overline{\phi}(x) = O(x^2)$. The BR reaction function is:

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

Proof of Proposition 6 Let us consider two function U and u^s

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

and define $a^{**}(w,x)$ and $a^s(w,x)$ to be the optimal actions given the associated utility functions:

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

First, we will prove:

Lemma 1 *We have, at* $x=0$,

$$\frac{\partial a^s(w,x)}{\partial x}_{|x=0} = \frac{\partial a^{**}(w,x)}{\partial x}_{|x=0}$$

Proof. The key fact comes from Proposition 7, 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

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

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

Likewise, for U^s,

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

Hence, we have

$$U^*_{ax|x=0} = U^s_{ax|x=0}$$

Note that we used $F^x_a = 0$. It is necessarily, because in general $V_{xx} \neq V^s_{xx}$.

Likewise,

$$\begin{aligned} U_{aa} &= u_{aa}(a,w,x) + \beta \mathbb{E}\left[F^w_a(w,x,a) \cdot V_{ww} \cdot F^w_a(w,x,a) + V_w \cdot F^w_{aa}(w,x,a)\right] \\ &+ 2\beta \mathbb{E}\left[F^x_a(w,x,a) \cdot V_{xw} \cdot F^w_a(w,x,a)\right] \\ &+ \beta \mathbb{E}\left[F^x_a(w,x,a) \cdot V_{xx} \cdot F^x_a(w,x,a) + V_x \cdot F^x_{aa}(w,x,a)\right] \end{aligned}$$

and a similar expression for U^s_{aa}, which leads to:

$$U^*_{aa} = U^s_{aa} \text{ at } x = 0$$

Finally, we have:

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

■

Given $a^{Rat}(w,x) = a^s(w) + \sum_i b_i(w) x_i + O(x^2)$, we have

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

Hence, the lemma gives:

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

so

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

Finally,

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

References

Abel, Andrew, Janice C. Eberly, and Stavros Panageas, "Optimal Inattention to the Stock Market with Information Costs and Transactions Costs," Working Paper, University of Pennsylvania, 2010.

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, "Durable consumption and asset management with transaction and observation cost," *American Economic Review*, forth.

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. 2011. "Salience and Consumer Choice." Harvard University Working Paper.

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,* forthcoming.

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. "A Sparsity-Based Model of Bounded Rationality."

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

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. 2011. "A Model of Focusing in Economic Choice." Unpublished.

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. 2011. "Business Cycle Dynamics under Rational Inattention," WP, 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.