

Online Appendix for “Behavioral Macroeconomics Via Sparse Dynamic Programming”

Xavier Gabaix

28 September 2023

This online appendix presents provides various proofs and complements. Its main part is section C which gives other applications of the framework, and a behavioral version of: dynamic portfolio choice, linear-quadratic problems, precautionary savings, dynamic investment in general equilibrium, addictive behavior, Ricardian equivalence, a discussion of whether the active decision of a consumer is about consumption or savings, the measurement of intertemporal elasticities of substitution when consumers are partially inattentive to the interest rate, and a model with a source-dependent marginal propensity to consume.

Appendix A: Complements

A.1. Tools to Expand a Simple Model Into a More Complex one

Calculating the marginal impact of a new variable: $V_x, V_{w,x}$. Here I develop the method to derive the Taylor expansion of a richer model, when starting from a simpler one. The methods are entirely paper and pencil.⁶⁸

The state variables evolve according to:

$$w' = F^w(w, x, a), \quad x' = F^x(w, x, a)$$

where the $'$ denote next-period variables. Also, x is stable around 0: $F^x(w, 0, a) = 0$.

Consider the fully rational model:

$$V^r(w, x) = \max_a u(w, x, a) + \beta \mathbb{E} V^r(F^w(w, x, a), F^x(w, x, a)) \quad (\text{A.1})$$

We start with a simpler model, where x is always 0: $x \equiv 0, F^{x'} = 0$, i.e.

$$V^d(w) = \max_a u(w, 0, a) + \beta \mathbb{E} V^d(F^{w'}(w, 0, a))$$

68. They draw from the techniques surveyed by Judd (1998) (Chapter 14), who has a more computer-based perspective. Section F.9 extends this to continuous time.

We use the notation, for a function $f(w, a)$,

$$D_w f = \partial_w f + (\partial_a f) \partial_w a^d(w) \quad (\text{A.2})$$

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 .⁶⁹

Bellman method.

LEMMA A.A.1. *The impact of a change x on the value function and its derivative is, respectively:*

$$V_x = \frac{u_x + \beta V'_{w'} F_x^w(w, x, a)}{1 - \beta F_x^x} \quad (\text{A.3})$$

$$V_{w,x} = \frac{D_w u_x + \beta D_w [F_x^w(w, 0, a) V'_{w'}(w', 0)] + \beta V'_{x'} F_{xw}^x}{1 - \beta F_x^x D_w w'} \quad (\text{A.4})$$

where all derivatives are evaluated at $(a, x) = (a^d(w), 0)$.

Proof. The proof yields the general method of calculation. We shut down uncertainty and differentiate the rational Bellman equation (A.1), first with respect to the new variable x (using the envelope theorem):

$$V_x(w, x) = u_x + \beta V'_{w'} F_x^w + \beta V'_{x'} F_x^x$$

which yields the announced expression for V_x . Then we take the total derivative w.r.t. w :

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

so

$$V_{w,x}(w, 0) = \frac{D_w u_x + \beta D_w [F_x^w(w, 0, a) V'_{w'}(w', 0)] + \beta V'_{x'} D_w F_x^x}{1 - \beta F_x^x D_w w'}.$$

□

LEMMA A.A.2. *Assume the local autonomy condition (32), $F_a^x = 0$, and consider some value function $V(w, x)$. Then, the impact of a change x on the optimal action is (around $x = 0$):*

$$a_x = -\Psi_a^{-1} \Psi_x \quad (\text{A.5})$$

69. The notations are a bit awkward when variables are multidimensional, and can be made clearer by using tensors. To keep things elementary, I don't explicitly use tensors here.

with $\Psi(a, x) = u_a + \beta V'_{w'} F_a^w$, and

$$\Psi_a = u_{aa} + \beta F_a^w V'_{w'} F_a^w + \beta V'_{w'} F_{aa}^w \quad (\text{A.6})$$

$$\Psi_x = u_{ax} + \beta V'_{w'} F_{ax}^w + \beta F_x^w V'_{w'} F_a^w + \beta F_x^x V'_{w'x'} F_a^w. \quad (\text{A.7})$$

evaluated at $(a, x) = (a^d(w), 0)$. They depend only on the transition functions and the derivatives of the simpler baseline value function $V^d(w')$.

Proof. The first order condition for a is $\Psi(a, x) = 0$ with

$$\Psi(a, x) = u_a + \beta V'_{w'} F_a^{w'} + \beta V'_{x'} F_a^{x'}$$

The rest follows by the implicit function theorem: terms in V_{xx} drop out because $F_a^{x'} = 0$ at the default action. $\square$

When the local autonomy condition (32) doesn't hold, a term V_{xx} appears. Then, the situation is more complex, and requires solving for a fixed point, in the form of a matrix Ricatti equation.

Feynman-Kac method. In some cases, it is useful to do the same via a Feynman-Kac type of approach.⁷⁰ Here we view x_t as exogenous, i.e. assume $x_{t+1} = F^x(x_t)$. Calling w the initial condition for wealth, the Lagrangian is:

$$\begin{aligned} L = & \sum_{t=0}^{\infty} \beta^t u(w_t, x_t, a_t) \\ & + \sum_{t=1}^{\infty} \beta^t q_t^w (-w_t + F^{w_t}(w_{t-1}, x_{t-1}, a_{t-1})) + q_0^w (-w_0 + w) \end{aligned}$$

where $q_t^w = V_w(w_t)$ are the Lagrange multipliers. At the optimum, the agent solves $V(w) = \max_{(a_t, w_t)_{t \geq 0}} L$. This implies that $L_{a_t} = L_{w_t} = 0$. The envelope theorem gives:

$$V_{x_t} = L_{x_t} = \beta^t [u_x(w_t, x_t, a_t) + \beta q_{t+1}^w F_x^{w_{t+1}}(w_t, x, a_t)]$$

so that, using the total derivative notation (A.2),

$$\begin{aligned} V_{x_t, w_0} = & \beta^t \frac{Dw_t}{Dw_0} \Big|_{x=0} D_{w_t} \left[u_x(w_t, x_t, a_t^d(w_t)) \Big|_{x_t=0} \right. \\ & \left. + \beta V_{w_t}(w_{t+1}) F_x^{w_{t+1}}(w_t, x_t, a_t^d(w_t)) \Big|_{x_t=0} \right] \end{aligned}$$

or in short

$$V_{x_t, w_0} = \beta^t \frac{Dw_t}{Dw_0} D_{w_t} \left[u_x^t + \beta V_{w_{t+1}}^{t+1} F_x^{w_{t+1}} \right] \quad (\text{A.8})$$

70. I call it ‘‘Feynman-Kac’’ because this approach deals particularly well with stochastic problems.

Application. In the consumption problem with $\beta R = 1$, let us derive again at the impact of a one-time change of interest rate $\hat{r}_t$ from Lemma 7. Under the default model $c_t = c_0$ and $w_t = w_0$. So $\frac{Dw_t}{Dw_0} = 1$, $u_{x_t} = 0$, $V_w^d(w_t) = u'(c_t) = \left(\frac{\bar{r}w_t + \bar{y}}{R}\right)^{-\gamma}$, and given $F^{w_{t+1}} = (1 + r_t)(w_t - c_t) + \bar{y}$,

$$F_{r_t}^{w_{t+1}} = w_t - c_t = w_t - \frac{rw_t + \bar{y}}{R} = \frac{w_t - \bar{y}}{R} = \frac{w_t^-}{R}$$

using the notation $w_t^- := w_t - \bar{y}$ for the beginning of period wealth. Using (A.8) gives:

$$\begin{aligned} V_{x_t, w_0} &= \beta^t D_{w_t} \left[\beta \left(\frac{\bar{r}w_{t+1} + \bar{y}}{R} \right)^{-\gamma} \frac{w_t^-}{R} \right] = \frac{1}{R^{t+2}} D_{w_t} \left[\left(\frac{\bar{r}w_{t+1} + \bar{y}}{R} \right)^{-\gamma} w_t^- \right] \\ &= \frac{1}{R^{t+2}} \left(-\gamma \frac{\bar{r}}{R} c_{t+1}^{-\gamma-1} \frac{Dw_{t+1}}{Dw_t} w_t^- + c_t^{-\gamma} \right) = \frac{1}{R^{t+2}} c_0^{-\gamma-1} \left(-\gamma \frac{\bar{r}}{R} w_0^- + c_0 \right) \end{aligned}$$

as under the default model $c_t = c_0$. As time-0 consumption satisfies $u_{c_0} = V_{w_0}$, we have $u_{cc} \partial_{\hat{r}_t} c_0 = \partial_{\hat{r}_t} V_{w_0}$, and

$$\begin{aligned} \partial_{\hat{r}_t} c_0 &= \frac{\partial_{\hat{r}_t} V_{w_0}}{u_{cc}} \\ &= \frac{\frac{1}{R^{t+2}} c_0^{-\gamma-1} \left(-\gamma \frac{\bar{r}}{R} w_0^- + c_0 \right)}{-\gamma c_0^{-\gamma-1}} \\ &= \frac{1}{R^{t+2}} \left(\frac{\bar{r}}{R} w_0^- - \psi c_0 \right) \\ &= \frac{1}{R^{t+2}} \left(\frac{r}{R} (w_0 - \bar{y}) - \psi c_0 \right). \end{aligned}$$

which gives again Lemma 7 (the income part being easy as always).

A.2. Extension: Proportional Thinking and Scale-Free κ

Here are some notes for a practical use of the model.

A.2.1. Using the model without κ . First, one could imagine directly measuring m . For instance, take the basic life-cycle model of Section 2. Suppose we had data on retirement saving, one could measure the m_t of say high-educated people vs less educated people, at different points in time t . That would be interesting. This way, we could back out the correct attention function, or see its determinants over time – a useful research enterprise.

Second, suppose that we keep κ as a more basic parameter, and measure κ it – for one given type of decision. For instance, in the life-cycle model, one could study just one variable, consumption, fitting a κ for it, say over a lifetime.

One important difficulty is that κ is not “scale-free” – it depends on the units of utility. To remedy that problem, I discussed earlier a “scale-free κ ” (Gabaix (2014)), which I state and generalize here.

A.2.2. Scale-free κ and Proportional Thinking.

Scale-free parametrization with a one-dimensional action. Suppose first that there is just a one-dimensional action. The following formulation is easy to microfoundation, and proves sensible and convenient:⁷¹

$$\kappa = (\bar{\kappa} a^{\natural})^2 |v_{aa}(a^d(x, m^d))| \quad (\text{A.9})$$

where $\bar{\kappa}$ is a unitless parameter, which might portable from one situation to the next (Rabin (2013) advocates such portability of parameters). Here $a^{\natural}$ is the “natural” or “typical scale” of the action. In music, $\natural$ means the “natural” pitch of a note, so perhaps it’s a mnemonic symbol for the “natural scale” of variable).

What is a natural scale? In many cases, this is reasonably clear. If the action is a purchase c_k in dollar of good k , then $a^{\natural} = c^d$. If the action is a transform $f(c_k)$ of that purchase, then $a_k^{\natural} = |f'(c_k^d)| c_k^{\natural}$. As a result, if we consider log consumption, the natural scale is $a_k^{\natural} = 1$.⁷²

Formulation (A.9) implies that attention to dimension i is:⁷³

$$m_i = \mathcal{A} \left(\frac{\mathbb{E}[a_{m_i}^2]}{(\bar{\kappa} a^{\natural})^2} \right)$$

In turns, we obtain a rule (with $\mathcal{A} = \mathcal{A}_1$): “Keep only dimensions of reality that matter for more than $\bar{\kappa}$ of the typical scale $a^{\natural}$ of the decision”.

For instance, a reasonable parameter might be $\bar{\kappa} = 5\%$. People then eliminate all considerations that matter for less than 5% of the decision, whether it is for a small decision (e.g., buying at the supermarket) or a big decision (e.g., buying a car).

Some evidence consistent with that is presented by Samuelson and Zeckhauser (1988), who show similar percentage price dispersion between cheap and expensive goods, and by Tversky and Kahneman (1981) : people consider a \$5 discount more worthy of an extra shopping trip if it is for a \$15 calculator than for a \$125 jacket. Bordalo et al. (2013) and Bushong et al. (2021) have emphasized the importance of this proportional thinking.

I conjecture that this feature is a good benchmark which would be interesting to evaluate empirically (I do not claim it will work perfectly, but I

71. Section F.5 provides a microfoundation, which I outline here. Mental effort creates trembling in the action, with a standard deviation $\sigma_a = \sqrt{2\bar{\kappa} a^{\natural}} (\sum_i g(m_i))^{1/2}$ proportional to the typical scale of the action (as in much of psychophysics, e.g. in the Weber-Fechner law), and increasing in mental activity m . The resulting utility losses are then $\frac{-1}{2} v_{aa} \sigma_a^2 = \kappa \sum_i g(m_i)$ with the value of κ in (A.9).

72. In addition, if a is presented as belong to a natural range $a \in [\underline{a}, \bar{a}]$ (as in many lab experiments): then we can take $a^{\natural} = \frac{1}{2} (|\bar{a} + \underline{a}| + |\bar{a} - \underline{a}|)$, which reflects the absolute value of the midpoint, and the width of the range. In other contexts, we can take $a^{\natural} = \text{var}(a^r(x))^{1/2}$.

73. I use $m_i^* = \mathcal{A}(-\mathbb{E}[a_{m_i} u_{aa} a_{m_i}] / \kappa)$, from (10).

conjecture that it will hold more likely than the polar opposite prediction that people would be 100 times more precise for a good that costs 100 times more).

Scale-free parametrization with many actions. We can generalize the idea above to many actions. To do that, formally, I use a variant of sparse max, which specifies a different attention for each type of action (formally, for each coordinate of the action vector).⁷⁴

DEFINITION A.1. (*Sparse max: scale-free with action-specific attention*). When it is scale-free, with action-specific attention, the sparse max $\text{smax}_{a; m|a^\natural}^{SF} v(a, x, m)$ is defined as follows.

Step 1: For each dimension k of the action, the attention vector $m^k = (m_i^k)_{i=1\dots n}$ is given by:

$$m^* = \arg \min_{m \in [0,1]^n} \sum_i \left[\frac{1}{2} (a_{m_i}^k)^2 (1 - m_i)^2 + (\bar{\kappa} a^{k, \natural})^2 g(m_i) \right] \quad (\text{A.10})$$

Step 2: The k -th dimension of the action (a_k^*) is the optimal under the model m^k , given the other dimensions of actions (a_{-k}^*).

$$a_k^* \in \arg \max_{a_k} v((a_k, a_{-k}^*), x, m^k). \quad (\text{A.11})$$

Practically, Definition A.1 yields an explicit formula for attention:

$$m_i^k = \mathcal{A} \left(\frac{\mathbb{E} \left[(a_{m_i}^k)^2 \right]}{(\bar{\kappa} a^{k, \natural})^2} \right) \quad (\text{A.12})$$

and a first order condition for action is, for each k :⁷⁵

$$v_{a_k}(a^*, x, m^k) = 0 \quad (\text{A.13})$$

In conclusion: based on the psychological evidence on proportional thinking, it seems plausible that the above formulation will calibrate reasonable well, with one scale-free parameter $\bar{\kappa}$, which I conjecture to be around $\bar{\kappa} = 5\%$ – if people omit factors that matter by less than 5%.

For concreteness, here is an example. Suppose that there are two actions, a^1 , a^2 , and:

$$v(a, x) = -\frac{1}{2} \sum_{k=1}^2 w^k \left(a^k - \sum_i b_i^k x_i \right)^2$$

74. See Gabaix (2014) (Section XV.D of the online appendix)

75. If several actions are possible (which is not a generic situation) then for instance we take one closest to the default action by some metric.

where $w^k > 0$ is a scale, b^k is a vector. Then, the attention-augmented utility function is $v(a, x, m) = -\frac{1}{2} \sum_{k=1}^2 w^k (a^k - \sum_i b_i^k m_i x_i)^2$. The basic sparse max yields:

$$m_i = \mathcal{A} \left(-\mathbb{E} [a_{m_i} v_{aa} a_{m_i}] \frac{1}{\kappa} \right) = \mathcal{A} \left(\frac{1}{\kappa} \sum_{k=1}^2 w_k (b_i^k)^2 \sigma_{x_i}^2 \right)$$

and $a^k = \sum_i b_i^k m_i x_i$. In contrast, the scaled sparse max of Definition A.1 gives:

$$m_i^k = \mathcal{A} \left(-\mathbb{E} [a_{m_i}^k v_{a^k a^k} a_{m_i}^k] \frac{1}{\kappa} \right) = \mathcal{A} \left(\frac{(b_i^k)^2 \sigma_{x_i}^2}{(a^{k, \natural})^2} \frac{1}{\bar{\kappa}^2} \right)$$

and $a^k = \sum_i b_i^k m_i^k x_i$ for $k = 1, 2$. Here, the attention is independent of w^k : for action k , the agent makes sure to include all considerations i that makes him change his decision by more than $\bar{\kappa}$ times the typical scale $a^{k, \natural}$.

A.3. Complements: Life-cycle model with T periods

Here are some more precisions to see how the intuitive procedure of Section 2 is an outcome of the general formalism. The action is consumption c_t and the state vector is

$$z_t = (w_t, x, t)$$

Note that $x^* := (T - L) \hat{y}$ is the objective value, so $z_0 = (w_0, x^*, 0)$. But it is useful to consider x a variable.

The objective laws of motion are expressed in $F^z(c, z_t)$ ⁷⁶. So, the rational agent at time t solves $c_t^r = \arg \max_t v(c, z_t)$, with v defined in (3): $v(c, z_t) := u(c) + V^r(w_t + \bar{y} - c, x, t + 1)$.

For the behavioral agent, the attention-augmented law of motion is $F^z(z_t, m) := F^z(w_t, mx, t)$ before retirement (for $t < L$): only x is misperceived. After retirement, I assume correct perception, $F^z(z_t, m) := F^z(z_t)$ ⁷⁷. The proxy value function is $V^p = V^r$ (equation (2)). Then, Definition 2 gives:

$$c_t = \arg \text{smax } v(c, z_t, m) \text{ with } v(c, z_t, m) := v(c, w_t, mx, t) \quad (\text{A.14})$$

When we apply the sparse max of Definition 1. Step 1 gives the value of m_t in (6). Step 2 gives the value of c_t in (6).

76. Formally,

$$F^w(c, z_t) = w_t + \bar{y} + \frac{x}{T - L} 1_{t \geq L} - c_t, \quad F^x(c, z_t) = x, \quad F^t(c, z_t) = t + 1.$$

77. Formally, there is an element of arbitrariness: one could conceive a partial misperception of time t , or misperception income after retirement. A medium-run open question is a way to endogenize such “intuitively reasonable” choices. Still, the model produces a lot of reasonable behavior “on its own”, e.g. in Figure 1.

Appendix B: Proofs

Sections C to G contain most proofs.

Proof of Lemma 2. The rational reaction function satisfies: $a^r(x) = a^d + \sum_i b_i x_i + \lambda(x)$ for a function $\lambda(x) = O(\|x\|^2)$. So, $\partial a / \partial x_i = b_i$ and:

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

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

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

Proof of Proposition 7. Suppose first that the problem is deterministic. Given attention m_r, m_y , the agent's problem is identical to the rational agent's problem, in a world in which the true interest rate is $\bar{r} + m_r \hat{r}_t$, and true income is $\bar{y} + m_y \hat{y}_t$, and the transition function is the one perceived by the agent. Hence, calling $c_0^r(w_0, (\hat{r}_\tau, \hat{y}_\tau)_{\tau \geq 0})$ the rational consumption given flows $c_0^s(w_0, (\hat{r}_\tau, \hat{y}_\tau)_{\tau \geq 0}) = c_0^r(w_0, (m_r \hat{r}_\tau, m_y \hat{y}_\tau)_{\tau \geq 0})$. Using the Taylor expansion in Lemma 7 gives the result. When the problem is stochastic, as in the proof of Lemma 7, we take the certain equivalent at each period.

Proof of Proposition 10. If $\varphi^d \geq \varphi^r$, we have $\varphi_0 = \frac{\xi}{r + \varphi^d} \leq \frac{\xi}{r + \varphi^r} = \varphi^r$, hence $\varphi = m_K \varphi_0 < \varphi_0 \leq \varphi^r$. If $\varphi^d \geq \varphi$ then $\varphi^0 = \frac{\xi}{r + \varphi^d} \leq \frac{\xi}{r + \varphi}$, so that $\varphi = m_K \varphi_0 < \varphi_0 \leq \frac{\xi}{r + \varphi}$, and $\varphi(r + \varphi) < \xi$. Given that $\xi = \varphi^r(r + \varphi^r)$, that implies $\varphi < \varphi^r$.

By (60) $\varphi < \varphi^r$ implies that fluctuations in capital $\text{var}(\hat{K}_t)$ are more volatile in the sparse economy. Given $Y_t = f(K_t) + \delta K_t$, we have $\hat{Y}_t = (f'(K_*) + \delta) \hat{K}_t$. So the same properties holds for GDP fluctuations $\hat{Y}_t$ as capital $\hat{K}_t$, as the two are proportional.

For consumption have $\hat{C}_t = (r + \varphi) \hat{K}_t$, so $\sigma_C = (r + \varphi) \sigma_K = (r + \varphi) \frac{\sigma_\varepsilon}{\sqrt{2\varphi}}$ from (60). The latter function is increasing in φ for $\varphi \geq r$. Hence, σ_C is higher in the sparse economy ($r \leq \varphi \leq \varphi^r$). Next, $\hat{I}_t = \hat{Y}_t - \hat{C}_t = (f'(K_*) + \delta - (r + \varphi)) \hat{K}_t = (\delta - \varphi) \hat{K}_t$, so $\sigma_{\hat{I}_t} = |\delta - \varphi| \sigma_K = |\delta - \varphi| \frac{\sigma_\varepsilon}{\sqrt{2\varphi}}$ a

U-shaped function of φ with minimum at $\varphi = \delta$. So, if $\varphi^r \leq \delta$, investment is more volatile in the sparse economy.

Proof of Proposition 11. In continuous-time limit, the second derivative of the value normalized function

$$v(c) := [u(c)dt + \beta V(k_t + y_t - (1 + r_t dt)c_t dt, K_{t+1}) - V(k_t, K_t)]/dt$$

is $v_{cc} = u_{cc}$, this is a quite general phenomenon. Applying (10) with $c_K = \varphi_0$, we obtain:

$$m_K = \mathcal{A} \left(\frac{|u_{cc}| \varphi_0^2 \sigma_K^2}{\kappa} \right) = \mathcal{A} \left(\frac{|u_{cc}| \varphi_0^2 \sigma_\varepsilon^2}{2\varphi \kappa} \right) = \mathcal{A} \left(\frac{\varphi_0^2}{B\varphi} \right), \quad B := \frac{2\kappa}{|u_{cc}| \sigma_\varepsilon^2}$$

using (60), $\sigma_K^2 = \frac{\sigma_\varepsilon^2}{2\varphi}$. Using $\varphi = \varphi_0 m_K = \varphi_0 \max \left(1 - \frac{B\varphi}{\varphi_0^2}, 0 \right)$, so $\varphi = \frac{\varphi_0}{1 + \frac{B}{\varphi_0}}$.

Appendix C: Some Other Applications

To make sure that the model is widely applicable, I developed a behavioral version of a few other important machines of dynamic economics.

C.1. Dynamic Portfolio Choice

I now study a Merton (1971) 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:

$$dw_t = (-c_t + r w_t) dt + w_t \theta_t (\pi_t dt + \sigma dZ_t)$$

where π_t is the equity premium and θ_t the allocation to equities.

I start by describing the rational problem, and then the behavioral solution. I call $\psi = \frac{1}{\gamma}$ the IES. Although for simplicity I use a CRRA utility function, I try to write the expressions in a way that involves both γ and ψ , a way that would generalize correctly to Epstein-Zin utility, where the two notions are disentangled.

C.1.1. Taylor expansions of the value function: rational case. We examine the problem in the rational case first, with a reminder of notions of portfolio choice. In a deterministic context with interest rate r_t , the SDF is simply $M_t = e^{-\int_0^t r_s ds}$. Next, suppose that there is a stochastic opportunity set: set of assets with risk premium π_t and covariance matrix Σ_t . In a static maximization, the optimal portfolio of the certainty equivalent is a return: $R_t(\theta_t) = r_t + \theta_t \pi_t - \frac{\gamma}{2} \theta_t \Sigma_t \theta_t$, so that the (static) optimal portfolio choice is $\theta_t = \arg \max_\theta R_t(\theta)$, i.e. $\theta_t = \frac{1}{\gamma} \Sigma_t^{-1} \pi_t$, and the certainty equivalent is finally: $R_t = \max_{\theta_t} R_t(\theta_t)$

$$R_t = r_t + \frac{1}{2\gamma} \Lambda_t \tag{C.1}$$

where

$$\Lambda_t = \pi_t' \Sigma_t^{-1} \pi_t \quad (\text{C.2})$$

the “squared Sharpe ratio” of the investment opportunity set. Suppose that the process is driven by a Brownian motion B_t (which may be multidimensional) - if the price of risk is λ_t (so that $\Lambda_t = \|\lambda_t\|^2$), the stochastic discount factor can be represented as:

$$M_t = \exp \left[- \int_0^t \left(\left(r_s + \frac{\Lambda_t}{2} \right) ds - \lambda_s dB_s \right) \right] \quad (\text{C.3})$$

The value function is as follows.

LEMMA C.1. (Value function, traditional case) *Suppose that the interest rate r_t and the price of risk λ_t are deterministic, and that the agent is the traditional rational agent. The value function derivative is*

$$V_w(w_t, x_t) = (\mu_t w_t)^{-\gamma}$$

and the optimal policy is to consume $c_t = \mu_t w_t$ (μ_t is the MPC to consume out of wealth), where:

$$\mu_t^{-1} = \mathbb{E}_t \left[\int_0^\infty e^{-\psi \rho s} \left(\frac{M_{t+s}}{M_t} \right)^{1-\psi} ds \right] = \mathbb{E}_t \left[\int_0^\infty e^{-\int_t^s (\psi \rho_u + (1-\psi) R_u) du} dt \right]$$

where

$$R_t = r_t + \frac{1}{2\gamma} \Lambda_t.$$

is the certainty equivalent of expected portfolio returns (comprising stocks and bonds), with $\Lambda_t = \|\lambda_t\|^2$ is the square Sharpe ratio of the investment opportunity set.

When the opportunity set is constant, we have $R_t = R_*$ and $\mu_t = \mu_*$ with

$$\mu_* = \psi \rho + (1 - \psi) R_*. \quad (\text{C.4})$$

When it is not constant, we have, up to second order terms:

$$\mu_t = \psi \rho + (1 - \psi) R_t \quad (\text{C.5})$$

where $R_t = \mu_* V_t^R$ is the average future portfolio returns, and V_t^R is the present value of future portfolio returns.

$$V_t^R := \mathbb{E}_t \left[\int_t^\infty e^{-\mu_*(s-t)} R_s ds \right] \quad (\text{C.6})$$

Here R_t is the future average return of the portfolio (including stocks and bonds). Hence, the marginal propensity to consume is a weighted average (with

weights ψ and $1 - \psi$) of the pure rate of time preference ρ and the average future return of the portfolio.

Lemma C.1 summarizes and somewhat generalizes well-known notions, particularly from the work of Campbell and Viceira (2002). It indicates that what matters is the risk-adjusted rate of return of the portfolio, R_t : it is the safe short-term rate r_t , plus the square Sharpe ratio Λ_t , divided by two times the risk aversion. The future average return R_t is key to capture the (leading order of) the value function. Related ideas are found in Basak and Chabakauri (2010) and Malamud and Vilkov (2018)).

To structure the problem, suppose that the vector of asset returns $d\tilde{r}_t$ (where $d\tilde{r}_{it}$ is the return of asset i) :

$$\begin{aligned} d\tilde{r} &= (r + \pi_* + \hat{\pi}_t) dt + \sigma dZ_t \\ \hat{\pi}_t &= f' X_t \end{aligned}$$

where X_t is a vector of factors, following an AR(1):⁷⁸

$$dX_t = -\Phi X_t dt + \sigma^X dZ_t$$

and f is a matrix of weights. We call

$$\Sigma^{r,X} = \text{cov}(d\tilde{r}, dX'_t) / dt = \sigma \sigma^{X'}.$$

the matrix of covariance, i.e. $\Sigma_{ij}^{r,X} = \text{cov}(d\tilde{r}_{it}, dX_{jt}) / dt$. We define $\theta_* = \frac{1}{\gamma} \Sigma_*^{-1} \pi_*$ as the portfolio choice in the model with constant variance and expected returns.

Then, the portfolio return is

$$\begin{aligned} R_t &= \frac{1}{2\gamma} (\pi_* + \hat{\pi}_t)' \Sigma_t^{-1} (\pi_* + \hat{\pi}_t) = \frac{1}{2\gamma} \pi_*' \Sigma_*^{-1} \pi_* + \theta_*' \hat{\pi}_t + O(\|X_t\|^2) \\ &= R_* + \theta_*' \hat{\pi}_t \\ &= R_* + \theta_*' f' X_t = R_* + b' X_t \end{aligned}$$

i.e. the return is augmented by $\theta_*' \hat{\pi}_t$, with

$$b := f \theta_*.$$

Then, the present value of returns (C.6) is

$$V_t^R = \frac{R_*}{\mu_*} + b' (\mu_* I + \Phi)^{-1} X_t \quad (\text{C.7})$$

where I is the identity matrix of the X 's dimension.

78. Or X_t could be a linearity-generating twisted-AR(1), so that the derivations below can be exact (Gabaix (2009)).

For instance, if X_t is one-dimensional, then $bX_t = \hat{R}_t := R_t - R_*$, and $R_t := R_* + \frac{\mu_*}{\mu_* + \Phi} \hat{R}_t$.

$$\mu_t = \mu_* + (1 - \psi) \frac{\mu_*}{\mu_* + \Phi} \hat{R}_t \quad (\text{C.8})$$

Hence, we obtain a tractable representation of the value function to the leading order.

C.1.2. The hedging demand. We can calculate the hedging demand.

LEMMA C.2. (Hedging demand, rational) *The stock demand is*

$$\theta_t = \frac{1}{\gamma} \Sigma_t^{-1} (\pi_t + H_t) \quad (\text{C.9})$$

where H_t is the hedging demand premium, equal to (up to second order terms):

$$H_{it} = (1 - \gamma) \text{cov} (d\tilde{r}_i, dV_t^R) \quad (\text{C.10})$$

i.e. H_{it} is $(1 - \gamma)$ times the covariance between asset i 's return ($d\tilde{r}_i$) and the present value of future returns V_t^R (equation C.6).

In the AR(1) framework above,

$$H_t = (1 - \gamma) \Sigma^{r,X} (\mu_* I + \Phi')^{-1} b. \quad (\text{C.11})$$

Suppose that returns mean-revert, i.e. $\text{cov} (d\tilde{r}_{it}, d\frac{R_t}{\mu_*}) < 0$. So, if $\psi < 1$, then investors load more on stocks because of the hedging demand.

We next state the modification of the value function.

LEMMA C.3. (Value function with hedging demand, rational) *In the hedging demand context, we have:*

$$\mu_t = \psi \rho + (1 - \psi) (R_t + \theta' H_t) \quad (\text{C.12})$$

where $R_t = \mu_* V_t^R$ is the expected present value of returns, and H_t is the hedging demand term; they are explicit in (C.7) and (C.11).

The intuition for (C.9) is that H_{it} is a risk-adjusted risk premium of asset i . This intuition carries over to (C.12). Compared to (C.5), the expression for $\mu(X_t)$ offers one more term, the term $(1 - \psi) \theta' H_t$.

A tractable case. The equity premium $\pi_t = \bar{\pi} + \hat{\pi}_t$ has a variable part $\hat{\pi}_t$, which follows

$$d\hat{\pi}_t = -\varphi_R \hat{\pi}_t dt - \chi_t \sigma dZ_t^1 + \sigma'_\pi dZ_t^2$$

where the return is $d\tilde{r}_t = (r_t + \pi_t) dt + \sigma dZ_t^1$. 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.

We call $\theta_* := \frac{\bar{\pi}}{\gamma \sigma^2}$ the standard, myopic demand for stocks.

C.1.3. The sparse agent's investment and consumption. We can calculate the sparse agent's demand. Recall that $\psi = 1/\gamma$ is the IES. We state again the proposition.

PROPOSITION C.1. (*Behavioral dynamic portfolio choice*) The fraction of wealth allocated to equities is, with $\theta_* := \frac{\bar{\pi}}{\gamma\sigma^2}$,

$$\theta_t^s = \theta_* + \tau \left(\frac{\hat{\pi}_t}{\gamma\sigma^2}, \kappa \right) + \tau \left(\frac{H_t}{\gamma\sigma^2}, \kappa_\theta \right)$$

while consumption is $c_t^s = \mu_t^s w_t$ with

$$\mu_t^s = \mu_* + \tau \left((1 - \psi) \frac{\mu_*}{\mu_* + \Phi} \theta_* \hat{\pi}_t, \kappa_{c/w} \right) + \tau \left((1 - \psi) \theta_* H_t, \kappa_{c/w} \right)$$

where H_t is the hedging demand term (C.13)

$$H_t = (1 - \gamma) \text{cov} (dr_t, dV_t^R) = - (1 - \gamma) \theta_* \frac{1}{\mu_* + \Phi} \sigma^2 \chi_t$$

Proof. We first calculate the rational values. In that case

$$R_t = r_* + \frac{\Lambda_*}{2\gamma} + \theta_* \frac{\mu_*}{\mu_* + \Phi} \hat{\pi}_t$$

$$H_t = (1 - \gamma) \text{cov} \left(dr_t, d \left(\frac{R_t}{\mu_*} \right) \right) = - (1 - \gamma) \theta_* \frac{1}{\mu_* + \Phi} \sigma^2 \chi_t \quad (\text{C.13})$$

so that

$$\theta_t = \frac{\pi_* + \hat{\pi}_t + H_t}{\gamma\sigma^2}$$

In addition

$$\mu_t = \psi\rho + (1 - \psi) (R_t + \theta'_* H_t) = \mu_* + (1 - \psi) \left(\theta_* \frac{\mu_*}{\mu_* + \Phi} \hat{\pi}_t + \theta'_* H_t \right)$$

As in Proposition 2, with ex-post attention, the BR agent just truncates those terms.

□

Proposition C.1 predicts the choice of a sparse agent. When $\kappa = 0$, it is the policy of a fully rational agent, e.g. as in 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, $\hat{\pi}_t$, 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.

C.1.4. Proofs for the Merton portfolio problem.

Proof of Lemma C.1. Here we present a proof sketch, in part because those notions are well-known. We record the values with a time-discounting of D_t , with $D_t = e^{-\rho t}$ in the infinite horizon, but D_t could be different to capture finite-time horizon effects. For instance, with a finite horizon of T , and a terminal weight b on the last consumption, then $D_t = ae^{-\rho t}1_{t \leq T} + b\delta(t - T)$.

First, in the SDF approach, the problem is

$$\max \mathbb{E} \int_0^\infty D_t c_t^{1-\gamma} dt \text{ s.t. } \mathbb{E} \int_0^\infty M_t c_t dt = w_0$$

This leads to $Dc_t^{-\gamma} = k'M_t$ and $c_t = kD_t^\psi M_t^{-\psi}$ for constant k, k' . The constant is determined by the budget constraint, $w_0 = \mathbb{E} \int_0^\infty M_t c_t dt = k \mathbb{E} \int_0^\infty k D_t^\psi M_t^{1-\psi}$. This leads to a utility derivative $V_w = (\mu_0 w_0)^{-\gamma}$, with

$$\mu_0^{-1} = \mathbb{E} \left[\int_0^\infty D_t^\psi M_t^{1-\psi} dt \right] \quad (\text{C.14})$$

When M_t follows (C.3), routine calculations show that

$$\mu_0^{-1} = \mathbb{E} \left[\int_0^\infty D_t^\psi e^{-(1-\psi) \int_0^t R_u du} dt \right]$$

We next proceed to a Taylor expansion:

$$\begin{aligned} \mu_0^{-1} &= \mathbb{E} \left[\int_0^\infty D_t^\psi e^{-(1-\psi) \int_0^t (R_* + \hat{R}_u) du} dt \right] \\ &= \mathbb{E} \left[\int_0^\infty D_t^\psi e^{-(1-\psi) R_* t} \left(1 - (1-\psi) \int_0^t \hat{R}_u du \right) dt \right] \end{aligned}$$

With an infinite horizon, $D_t = e^{-\rho t}$ and

$$\begin{aligned} \mu_0^{-1} &= \mathbb{E} \left[\int_0^\infty e^{-\mu_* t} \left(1 - (1-\psi) \int_0^t \hat{R}_u du \right) dt \right] \\ &= \frac{1}{\mu_*} - (1-\psi) \mathbb{E} \left[\int_{t=0}^\infty e^{-\mu_* t} \int_0^t \hat{R}_u du dt \right] \\ &= \frac{1}{\mu_*} - (1-\psi) \mathbb{E} \left[\int_0^t \left(\int_{t=u}^\infty e^{-\mu_* t} \right) \hat{R}_u du \right] \\ &= \frac{1}{\mu_*} - (1-\psi) \mathbb{E} \left[\int_0^t \frac{1}{\mu_*} e^{-\mu_* u} \hat{R}_u du \right] \\ &= \frac{1}{\mu_*} - (1-\psi) \frac{1}{\mu_*^2} \mathbb{E} \left[\int_0^t \mu_* e^{-\mu_* u} \hat{R}_u du \right] \\ &= \frac{1}{\mu_*} - (1-\psi) \frac{1}{\mu_*^2} (R_0 - R_*) \text{ with } R_0 - R_* \\ &= \mathbb{E} \left[\int_0^t \mu_* e^{-\mu_* u} \hat{R}_u du \right] \\ &= \frac{1}{\mu_* + (1-\psi)(R_0 - R_*)} + O(R_0 - R_*)^2. \end{aligned}$$

Then,

$$\begin{aligned}
\mu_0 &= \mu_* + (1 - \psi) (R_0 - R_*) \\
&= \psi \rho + (1 - \psi) R_* + (1 - \psi) (R_0 - R_*) \\
&= \psi \rho + (1 - \psi) R_0
\end{aligned}$$

When the consumer has a finite horizon and only cares about date T consumption, then $D_t = \delta(t - T)$, and

$$\begin{aligned}
\mu_0^{-1} &= \mathbb{E} \left[\int_0^\infty D_t^\psi e^{-(1-\psi)R_*t} \left(1 - (1 - \psi) \int_0^t \hat{R}_u du \right) dt \right] \\
&= e^{-\mu_* T} - e^{-\mu_* T} (1 - \psi) \mathbb{E} \left[\int_0^T \hat{R}_u du \right]
\end{aligned}$$

so the MPC is 0 but we have

$$\mu_t^{-1} = e^{-\mu_* (T-t)} \left(1 - (1 - \psi) \mathbb{E} \left[\int_t^T \hat{R}_u du \right] \right) \quad (\text{C.15})$$

so again μ_t is related to the present value of future portfolio returns.

□

Proof of Lemma C.2. In semi-discrete notation the asset demand at time t comes from

$$\max_\theta \mathbb{E}_t [V(w(1 + r_t dt + \theta d\tilde{r}_t), X_t + dX_t)]$$

where, with $\pi_t = \pi_* + f' X_t$,

$$\begin{aligned}
\mathbb{E}_t [dV_t] &= \mathbb{E}_t [V(w(1 + r_t dt + \theta d\tilde{r}_t), X_t + dX_t) - V(w, X_t)] \\
&= V_w w (r_t + \theta' \pi_t) dt + V_{wX} w \langle \theta' d\tilde{r}_t, dX_t \rangle \\
&\quad + V_{ww} w^2 \theta' \Sigma_t \theta dt + \frac{1}{2} \text{Tr} (V_{XX} \Sigma^{X,X}) dt \\
&= V_w w \left[\theta' (\pi_t + H_t) - \frac{\gamma}{2} \theta' \Sigma_t \theta dt \right] dt + \frac{1}{2} \text{Tr} (V_{XX} \Sigma^{X,X}) dt
\end{aligned}$$

where

$$\theta' H_t = \frac{V_{wX}}{V_w} \langle \theta d\tilde{r}_t, dX_t \rangle$$

is the hedging demand premium term. This implies

$$\theta = \frac{1}{\gamma} \Sigma_t^{-1} (\pi_t + H_t)$$

To calculate H_t more fully, recall that $V_w = \mu(X_t)^{-\gamma} w^{-\gamma}$, so that $\ln V_w = -\gamma \mu(X_t) - \gamma \ln w$, and

$$\frac{V_{wX}}{V_w} = -\gamma \frac{\mu_X}{\mu_*} = -\gamma (1 - \psi) \frac{R_X}{\mu_*} = (1 - \gamma) \frac{R_X}{\mu_*} =: B'$$

with

$$\frac{R_X}{\mu_*} = b' (\mu_* I + \Phi_R)^{-1}$$

Note that $R_t = r + \frac{1}{2\gamma} \pi'_t \Sigma^{-1} \pi_t$ with $\pi_t = \pi_* + \hat{\pi}_t$, so, with $R_* = r + \frac{1}{2\gamma} \pi'_* \Sigma^{-1} \pi_*$

$$\begin{aligned} R_t &= R_* + \frac{1}{\gamma} \pi'_* \Sigma^{-1} \hat{\pi}_t = R_* + \theta'_* \hat{\pi}_t \\ R_t &= R_* + \theta'_* f' X_t \end{aligned} \tag{C.16}$$

hence

$$b = f\theta \tag{C.17}$$

Hence,

$$\theta' H_t = \frac{V_{wX}}{V_w} \langle \theta d\tilde{r}_t, dX_t \rangle = \sum_{i,j} \theta_i \langle d\tilde{r}_{it}, B_j dX_j \rangle = \theta_i \Sigma_{ij}^{r,X} B_j$$

so that

$$\begin{aligned} H_t &:= \Sigma^{r,X} B = (1 - \gamma) \Sigma^{r,X} \frac{\bar{R}'_X}{\mu_*} = (1 - \gamma) \text{cov} \left(d\tilde{r}, d \frac{R_t}{\mu_*} \right) \\ &= (1 - \gamma) \Sigma^{r,X} (\mu_* I + \Phi'_R)^{-1} b \end{aligned}$$

Proof of Lemma C.3. Suppose

$$d\tilde{r} = (r + \pi_* + fX_t) dt + \sigma dZ_t$$

and that agents have a constant MPC μ_* :

$$\begin{aligned} \frac{dw_t}{w_t} &= (r - \mu_*) dt + \theta' d\tilde{r} = (r - \mu_* + \theta\pi_* + \theta' f' X_t) dt + \theta\sigma dZ_t \\ &= (g_* + \theta' f' X_t) dt + \theta\sigma dZ_t \\ \frac{dw_t}{w_t} &= (g_* + b' X_t) dt + \theta\sigma dZ_t \end{aligned}$$

with $b = f\theta$ and

$$g_* := r + \theta\pi_* - \mu_*$$

We want to calculate (assuming the policy $c_t = \mu_* w_t$, which leads only to second order losses)

$$U = \mathbb{E} \left[\frac{1}{1 - \gamma} \int_0^\infty e^{-\rho s} c_s^{1-\gamma} ds \right] = \frac{\mu_*^{1-\gamma}}{1 - \gamma} \mathbb{E} \left[\int_0^\infty e^{-\rho s} w_s^{1-\gamma} ds \right]$$

Calling

$$m_s = e^{-\rho s} w_s^{1-\gamma}$$

We calculate

$$\begin{aligned}
\frac{dm_t}{m_t} &= -\rho + (1-\gamma) \left(g_* + b' X_t - \frac{\gamma}{2} \|\theta\sigma\|^2 \right) dt + (1-\gamma) \theta' \sigma dZ_t \\
&= (-a + (1-\gamma) b' X_t) dt + (1-\gamma) \theta' \sigma dZ_t \\
a &:= \rho - (1-\gamma) \left(g_* - \frac{\gamma}{2} \|\theta\sigma\|^2 \right) \\
&= \rho - (1-\gamma) \left(r + \theta\pi_* - \mu_* - \frac{\gamma}{2} \|\theta\sigma\|^2 \right) = \rho - (1-\gamma) (R_* - \mu_*) \\
&= \mu_*
\end{aligned}$$

We calculate linearly generating (LG moments. We assume $dX_t = -\Phi X_t dt + \sigma^X dZ_t + O(\|X_t\|^2)$):

$$\begin{aligned}
\mathbb{E} \left[\frac{dm_t}{m_t} \right] / dt &= -\mu_* + (1-\gamma) b' X_t \\
\mathbb{E} \left[\frac{d(m_t X_t)}{m_t} \right] / dt &= (-\mu_* + (1-\gamma) b' X_t) X_t - \Phi X_t + (1-\gamma) \langle \theta' d\tilde{r}_t, dX_t \rangle \\
&= (1-\gamma) \theta' \langle d\tilde{r}_t, dX_t \rangle + (-\mu_* - \Phi) X_t + O(\|X_t\|^2)
\end{aligned}$$

so the LG generator (Gabaix (2009)) is

$$\omega = \begin{pmatrix} \mu_* & -(1-\gamma) b' \\ -(1-\gamma) \Sigma^{X,r} \theta & \mu_* + \Phi \end{pmatrix}$$

Hence, the present value is $V = (1, 0) \omega^{-1} \cdot (1, X_t)$

We use the formula for the inversion of the block matrix:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1} BD^{-1} \\ * & * \end{pmatrix}$$

where $*$ are terms we will not use. We have

$$\begin{aligned}
(1, 0) \omega^{-1} &= \left(f, f(1-\gamma) b' (\mu_* + \Phi)^{-1} \right) \\
f &:= \left(\mu_* - (1-\gamma)^2 b' (\mu_* + \Phi)^{-1} \Sigma^{X,r} \theta \right)^{-1}
\end{aligned}$$

so

$$V = f \left(1 + (1-\gamma) b' (\mu_* + \Phi)^{-1} X_t \right) \tag{C.18}$$

The value function has the Taylor expansion: $V(w_t, X_t) = v(X_t) \mu_*^{1-\gamma} \frac{w_t^{1-\gamma}}{1-\gamma}$

$$\begin{aligned}
v(X_t) &= \frac{1 + (1-\gamma) b'(\mu_* + \Phi)^{-1} X_t}{\mu_{**}} \\
\mu_{**} &= \mu_* - (1-\gamma)^2 b'(\mu_* I + \Phi)^{-1} \Sigma^{X,r} \theta \\
&= \mu_* - \frac{(1-\gamma)^2}{\left(1 - \frac{1}{\psi}\right)} H_t' \theta \text{ using (C.11), } H_t = \left(1 - \frac{1}{\psi}\right) \Sigma^{r,X} (\mu_* I + \Phi')^{-1} b \\
&= \mu_* - (1-\gamma) H_t' \theta
\end{aligned}$$

Rewrite

$$\begin{aligned}
V &= v(X_t) \mu_*^{1-\gamma} = \frac{1+K}{\mu_* + L} \mu_*^{1-\gamma} \text{ with} \\
K &= (1-\gamma) b'(\mu_* + \Phi)^{-1} X_t \\
L &= -(1-\gamma) H_t' \theta \\
V &= (\mu_* + C)^{-\gamma} = \mu_*^{-\gamma} (1 - \gamma \mu_*^{-1} C) \\
&= \mu_*^{-\gamma} (1 + K - \mu_*^{-1} L)
\end{aligned}$$

Hence,

$$\begin{aligned}
\mu_t - \mu_* &= C = -\frac{\mu_*}{\gamma} K + \frac{1}{\gamma} L = (1-\psi) b' \mu_* (\mu_* + \Phi)^{-1} X_t + \frac{-1}{\gamma} (1-\gamma) H_t' \theta \\
&= (1-\psi) b' \mu_* (\mu_* + \Phi)^{-1} X_t + (1-\psi) H_t' \theta \\
\mu^I &:= \frac{1}{\gamma} L = \frac{-1}{\gamma} (1-\gamma) H_t' \theta = \left(1 - \frac{1}{\psi}\right) H_t' \theta
\end{aligned}$$

Intuition: the extra present value of returns is

$$\begin{aligned}
\frac{R_t - R_*}{\mu_*} &= \frac{C}{(1-\psi) \mu_*} = b'(\mu_* + \Phi)^{-1} X_t - \psi(1-\psi) \frac{1}{\mu_*} b'(\mu_* I + \Phi)^{-1} \theta \Sigma^{r,X} \\
&= b'(\mu_* + \Phi)^{-1} \left(X_t + (1-\gamma) \frac{1}{\mu_*} \theta \Sigma^{r,X} \right)
\end{aligned}$$

C.2. Linear-Quadratic models

Many economic problems can be conveniently expressed as linear-quadratic (LQ) models (Ljungqvist and Sargent (2012)). We show here how to systematically derive a BR version of those models.

We again write $z = (w, x)$, where w is the set of variables known under the default model, and x is the set of variables that are not considered in the default model. Utility is:

$$u(z, a) := \frac{1}{2} \begin{pmatrix} z \\ a \end{pmatrix}' \begin{pmatrix} U_{zz} & U_{za} \\ U_{az} & U_{aa} \end{pmatrix} \begin{pmatrix} z \\ a \end{pmatrix}$$

and the law of motion is:

$$z' = F^z(z, a) := \Gamma_z^z z + \Gamma_a^z a$$

where U and Γ are constant matrices. The rational value function is also LQ

$$V(z) = -\frac{1}{2} z' V_{zz} z = -\frac{1}{2} (w' V_{ww} w + 2w' V_{wx} x + x' V_{xx} x)$$

Under the default model V_{ww} is known, and

$$a^d(w) = A_w w$$

for A_w a constant. Our goal is to find V_{wx} , which affects the value function. To do so, we apply from A.1 :

LEMMA C.4. *In the linear-quadratic problem, the cross-partial derivative of the value function is*

$$V_{wx} = V_{xw} = \left[1 - \beta (D_w w') \cdot \Gamma_x^{x'} \right]^{-1} [U_{xw} + U_{xa} A_w + \beta \Gamma_x^w V_{ww} (D_w w')].$$

where $D_w w' = \Gamma_w^w + \Gamma_a^w A_w$. The impact on the action is $a = A_w w + A_x x$, where A_w is the default value, and

$$A_x = -\Psi_a^{-1} \Psi_x, \tag{C.19}$$

where

$$\begin{aligned} \Psi_a &= U_{aa} + \beta \Gamma_a^w V_{ww} \Gamma_a^w \\ \Psi_x &= U_{xa} + \beta V_{wx} \Gamma_a^w \end{aligned}$$

This illustrates that the value function can be written:

$$V(z) = -\frac{1}{2} z' V_{zz} z = -\frac{1}{2} w' V_{ww} w + w' V_{wx} x + O(\|x\|^2)$$

with matrix V_{wx} as expressed in closed form above.

Hence, the BR value function is simply:

$$V^s(z, m) = -\frac{1}{2} z' V_{zz} z = -\frac{1}{2} w' V_{ww} w + w' V_{wx} M(m) x + O(\|x\|^2)$$

for the diagonal attention matrix $M(m) = \text{diag}(m_{x_i})$.

PROPOSITION C.2. (Behavioral version of linear-quadratic problems) *In a linear-quadratic problem, the optimal attention is*

$$m_{x_i} = \mathcal{A}(A_{x_i} \Psi_a A_{x_i} \sigma_i^2 / \kappa) \tag{C.20}$$

and the optimal sparse action is

$$a = A_w w + A_x M x$$

where $M = \text{diag}(m_{x_i})$. Here we use the notations of Lemma C.4.

C.3. Precautionary saving

The consumer may save more when the future is uncertain, a phenomenon known as “precautionary savings.” This is easy to obtain in this BR model. Suppose that the income process is $\hat{y}_{t+1} = \rho_y \hat{y}_t + \varepsilon_{t+1}$, for ε a mean-0 variable. Then, the rational value function does not obtain in closed form, unless we assume very specific functional forms (CARA utility, Gaussian noise). What to do then?

Let us first derive the rational policy.⁷⁹

LEMMA C.5. (Rational policy with precautionary saving) *With stochastic income shocks, the rational value function is*

$$\begin{aligned} V(w_t, \hat{y}_t, \sigma_\varepsilon^2) &= \frac{R}{r} \mathbb{E} u \left(\frac{\bar{r}}{R} w + \bar{y} + \frac{\bar{r}}{R} \sum_{\tau > t} \frac{\hat{y}_\tau}{R^{\tau-t}} \right) + o(\sigma_\varepsilon^2) \\ &= \frac{R}{r} u \left(\frac{\bar{r}}{R} w + \bar{y} + \frac{\bar{r}}{R} \mathbb{E} \left[\sum_{\tau > t} \frac{\hat{y}_\tau}{R^{\tau-t}} \right] \right. \\ &\quad \left. - \frac{\Gamma}{2} \left(\frac{\bar{r}}{R} \right)^2 \text{var}_t \left(\sum_{\tau > t} \frac{\hat{y}_\tau}{R^{\tau-t}} \right) \right) + o(\sigma_\varepsilon^2) \end{aligned}$$

where $\Gamma = -\frac{u''(c^d)}{u'(c^d)}$ ($c^d = \frac{\bar{r}}{R} w_t + \bar{y}$) is the coefficient of absolute risk aversion. For instance, with an AR(1) process:

$$\hat{c}_t = \frac{\bar{r}}{R} \frac{\hat{y}_t}{R - \rho_y} - \frac{\Gamma}{2} \left(\frac{\bar{r}}{R} \right)^2 \frac{\text{var}(\varepsilon_{t+1}^y)}{(R - \rho_y)^2} \quad (\text{C.21})$$

Then, the agent may, or may not, take the noise into account.

Proof. First, observe that the two RHS are identical, up to $o(\sigma_\varepsilon^2)$ terms.

To see the second statement, let us take more successively more complex problems. First, for small noise, we have, as in Arrow-Pratt:

$$\mathbb{E}[u(X)] = u \left(\mathbb{E}[X] - \frac{\Gamma}{2} \text{var}(X) \right) + o(\text{var}(X))$$

with $\Gamma = -\frac{u''(0)}{u'(0)}$.

Second, suppose that $\hat{y}_{t+1} = \rho_y \hat{y}_t + \varepsilon_{t+1}^y$. We are looking for an expansion of the type $V(w, \hat{y}, \sigma_y^2) = V(w + A, \hat{y}, 0) = V^d \left(w + A + \frac{\hat{y}}{R - \rho_y} \right)$, for some A . We have

$$V^d(w + A, \hat{y}) = \max_c u(c) + \beta \mathbb{E} \left[V^d(R(w - c) + y + A, \mathbb{E}[\hat{y}_{t+1}] + \varepsilon_{t+1}^y) \right]$$

79. See Wang et al. (2016) and Achdou et al. (2022) for recent analytical progress on this issue.

so, taking the Taylor expansion of $\hat{y}_{t+1}$ around $\mathbb{E}_t \hat{y}_{t+1}$

$$V_w^d A = \beta V_w^d A + \frac{1}{2} \beta V_{yy}^d \text{var}(\varepsilon_{t+1}^y)$$

so, using $\beta = \frac{1}{R}$

$$\begin{aligned} A &= \frac{1}{2} \frac{\beta}{1 - \beta} \frac{V_{yy}^d}{V_w^d} \text{var}(\varepsilon_{t+1}^y) \\ &= -\frac{1}{2} \frac{\bar{r}}{R} \Gamma^{V^{d,0}} \frac{1}{(R - \rho_y)^2} \text{var}(\varepsilon_{t+1}^y) \end{aligned}$$

and as $\Gamma^{V^{d,0}} = \frac{\bar{r}}{R} \Gamma^u$,

$$A = -\left(\frac{\bar{r}}{R}\right)^2 \frac{\Gamma^u}{2} \frac{1}{(R - \rho_y)^2} \text{var}(\varepsilon_{t+1}^y).$$

Next, for a more general process with state vector z , we have, by the same reasoning,

$$A = \frac{1}{2} \frac{\bar{r}}{R} \frac{\mathbb{E}[\varepsilon_z V_{zz}^d \varepsilon_z]}{V_w^d}$$

Now, it is not trivial to get V_{zz} . Indeed, we have $V(w, z) = V^d(w + b \cdot z) + O(z^2)$, but that expression gives only part of V_{zz} . The “certainty equivalent” approach works well for income shocks, but not for uncertainty about interest rates, say. $\square$

A sparse agent will, in contrast, do

$$\hat{c}_t = m_y \frac{\bar{r}}{R} \frac{\hat{y}_t}{R - \rho_y} - m_{\sigma_y^2} \frac{\Gamma}{2} \left(\frac{\bar{r}}{R}\right)^2 \frac{\text{var}(\varepsilon_{t+1}^y)}{(R - \rho_y)^2}$$

with some inattention to risk $m_{\sigma_y^2}$. Hence, he will create a too small buffer of savings, compared to a rational agent.

PROPOSITION C.3. *A sparse agent saves too little against idiosyncratic shocks, compared to a rational agent.*

C.4. Investment Problems

Suppose that the agent needs to solve:

$$V(K_0) = \max_{I_t} \int e^{-\rho t} (\pi(K_t) - G(I_t)) dt \text{ s.t. } \dot{K}_t = -\delta K_t + I_t$$

where $\pi(K)$ is the profit rate, $G(I) = I + g(I)$ is the cost of investment, inclusive of adjustment cost I . Capital depreciates at a rate δ .

How will the agent proceed? We apply the generalized k/K procedure of section F.8, using $(X, a) := (K, I)$. As $u(K, I) = \pi(K) - g(I)$ we have

$$\bar{u}(k, K, I) = \pi(K) + \pi_K(K)(k - K) - g(I)$$

and $\bar{F}(k, K, I) = -\delta k + I$, so that

$$\dot{k}_t = -\delta k_t + I_t$$

Hence, the agent simply solves a model with linear profitability of capital:

$$V(k_0, K_0) = \max_{I_t} \int e^{-\rho t} (\pi_K(K_t) k_t - G(I_t)) dt \text{ s.t. } \dot{k}_t = -\delta k_t + I_t$$

Hence, optimal investment satisfies

$$G'(I_0) = V_k(k_0, K_0) = \mathbb{E}^s \int_0^\infty e^{-(\rho+\delta)t} \pi_K(K_t) dt$$

i.e. on the RHS with have the subjective expectation of marginal profits.

At the steady state, with $K_t = K_*$, the optimum is characterized by $G'(I_*) = \pi_K(K_*)/(\rho + \delta)$ with $I_* = \delta K_*$ – as in the traditional model. As in the general procedure, I assume that the agent perceives a linear mean-reversion of the state variable: $\dot{K}_t = -\Phi K_t$, for some perceived speed Φ .

Outside the steady state,

$$\begin{aligned} \hat{I}_0 &= \frac{\int_0^\infty e^{-(\rho+\delta)t} \pi_{KK}(K_*) \hat{K}_t dt}{G''(I_*)} \\ &= - \int_0^\infty e^{-(\rho+\delta)t} \xi e^{-\Phi t} \hat{K}_0 dt \\ &= - \frac{\xi}{\rho + \delta + \Phi} \hat{K}_0 \\ &= \frac{1}{G''(I_*)} \frac{\widehat{\pi_K}|_{t=0}}{\rho + \delta + \Phi} \end{aligned}$$

with $\xi := \frac{-\pi_{KK}(K_*)}{G''(I_*)} \geq 0$ and $\widehat{\pi_K} = \pi_{KK} \hat{K}_t$. Here,

$$\hat{I}_0 = \frac{1}{G''(I_*)} \frac{\widehat{\pi_K}|_{t=0}}{\rho + \delta + \Phi}$$

means that the agent's investment reacts to current (marginal) profitability $\widehat{\pi_K}|_{t=0}$, with a dampening indexed by Φ , which needs not be the “rational” amount of dampening. For instance, if Φ is low, the agent will overreact to current profitability. This shows, I hope, that the procedure is reasonable psychologically.

The investment policy $\hat{I}_t = \xi \frac{1}{\rho + \delta + \Phi} \hat{K}_t$ implies that the true law of motion of capital is

$$\dot{K}_t = -\delta \hat{K}_t + \hat{I}_t = - \left(\delta + \frac{\xi}{\rho + \delta + \Phi} \right) K_t$$

so that the true speed of mean-reversion is

$$\Phi^r = \delta + \frac{\xi}{r + \delta + \Phi}.$$

Under rational expectations, $\Phi = \Phi^r$, so that a fixed point needs to be solved for. In the more general model here, the agent perceives a speed of mean-reversion of profitability Φ , and reacts accordingly. De facto, he sets $G'' \cdot \hat{I}_0 = \mathbb{E}^s \int e^{-\rho t} \widehat{\pi_K} dt$, with $\widehat{\pi_K} = \pi_{KK} \hat{K}_t$, i.e. sets investment according to the perceived changes in future profitability of capital.

C.5. The Becker-Murphy model of Rational Addiction

The Becker and Murphy (1988) model of rational addiction is a peak of the use of rationality in economics. We will give a behavioral version of it. We shall see that the qualitative evidence in favor of the model (the fact that future increase in prices lower consumption today) are also consistent with this BR version - it shows that agent are at least *partially* rational (as in the present model), not that they are *fully* rational (as assumed by Becker-Murphy). This distinction is important: if people are BR enough, they'd be better off under a high tax, or a ban, of the addictive substance – while the optimal tax is 0 in the Becker-Murphy model. This analysis is in the spirit of Gruber and Köszegi (2001), who study a hyperbolic discounting addict, rather than a boundedly rational one in the sense of this paper.

We call c the consumption and x the level of addiction. Utility function is

$$u(c, x) = -\frac{1}{2}(c - x - A)^2 - Bx$$

Addiction x_t evolves as

$$x_{t+1} = \rho x_t + h c_t.$$

The BR agent has in mind the model

$$x_{t+1} = \rho^s x_t + h^s c_t.$$

We posit that in the default model the agent does not perceive any addiction dynamics: he perceives addiction as being constant.

$$\rho^d = 1, \quad h^d = 0.$$

When the agent has partial attention m to inattention dynamics, we have

$$x_{t+1} = (1 - m)x_t + m(\rho x_t + h c_t)$$

so

$$\rho^s = (1 - m) + m\rho, \quad h^s = mh$$

Let us now study the BR dynamics.

Warm-up: 2 period model. As before, it is helpful to study a 2-period model, with $t = 1, 2$.

Behavior at the last period, $t = 2$. The agent should and does consume his optimal consumption

$$c^d(x) = \arg \max_c u(c, x) = x + A$$

We define the resulting utility as $\bar{u}(x)$

$$\bar{u}(x) := \max_c u(c, x) = u(c^d(x), x) = -Bx.$$

To, the time-1 value function is

$$V^1(x) = \bar{u}(x). \quad (\text{C.22})$$

Behavior at period 1, $t = 1$. Given perceived dynamics, the problem is

$$\begin{aligned} & \text{smax}_{c; m} v(c, x, m) \\ v(c, x, m) &:= u(c, x) + \beta V(\rho^s(m)x + h^s(m)c) \end{aligned}$$

which gives:

$$\begin{aligned} 0 &= u_c + \beta h^s V'(\rho^s x + h^s c) \\ &= -c + x + A - \beta h^s B \\ c &= x + A - \beta h^s B \end{aligned} \quad (\text{C.23})$$

An interesting variant is to impose $c \geq 0$. Then, first period consumption is > 0 iff $A - h^s B > 0$. So, if $h^s B < A \leq hB$, then *the rational agent consumes 0, while the very behavioral agent consumes a positive amount and becomes addicted*.

The optimal attention is $m = \mathcal{A}(v_{cc}c_m^2/\kappa) = \mathcal{A}(-u_{cc}B^2h^2/\kappa)$.

Infinite horizon model. The value function satisfies

$$V(x) = \text{smax}_{c; m} u(c, x) + \beta V(\rho^s(m)x + h^s(m)c)$$

The FOC is

$$u_c(c, x) + \beta V'(\rho^s(m)x + h^s(m)c)h^s(m) = 0$$

i.e. the agent takes into account only *part* of the addiction costs, as $h^s(m) \leq h$. As a result, the agent is more addicted in the steady state. The greater the myopia, the greater the optimal tax.

PROPOSITION C.4. *In the Becker-Murphy model with boundedly rational agents, the consumption c given the stock of addiction x is*

$$c(x) = x + A + \beta b(m) m^h h$$

using $m = (m^h, m^V)$; the value function is

$$V(x, m) = a(m) + b(m)x$$

where $b(m) = -\frac{B}{1-\beta(1+m^V(\rho+h-1))}$ and $a(m)$ is in the proof. When using the plain (as opposed to iterated) sparse max, $m^V = 0$ and attention to addition is $m^h = \mathcal{A}\left(\frac{1}{\kappa} \left(\frac{\beta B h}{1-\beta}\right)^2\right)$.

Proof of Proposition C.4. We're looking for a solution of the form: $V(x) = a + bx$, for a, b to be determined. The FOC is: $u_c + \beta V_x h^s = 0$, i.e. $-(c - x - A) + \beta b h^s = 0$ and

$$\begin{aligned} c &= x + A + \beta b^s h^s \\ u(c(x), x) &= -\frac{1}{2} (\beta b^s h^s)^2 - Bx \end{aligned}$$

The self-consistency condition is

$$V(x) = u(c(x), x) + \beta V(\rho x + hc)$$

i.e.

$$a + bx = \frac{-1}{2} (\beta b^s h^s)^2 - Bx + \beta [a + b(\rho x + h(x + A + \beta b^s h^s))]$$

This gives

$$\begin{aligned} b &= \frac{-B}{1 - \beta \bar{\rho}} \\ a &= \beta \frac{hA + \beta b^s h h^s - \frac{1}{2} \beta (b^s h^s)^2}{1 - \beta} \end{aligned}$$

When the agent perceives $\rho' = 1 - m^V + m^V \rho$ and $h' = m^V h$ when forming the value function, we have the same expressions,

$$\begin{aligned} b^s(m) &= \frac{-B}{1 - \beta(\rho' + h')} = \frac{-B}{1 - \beta(1 + m^V(\rho + h - 1))} \\ a(m) &= \beta \frac{hA + \beta b^s(m) h h^s - \frac{1}{2} \beta (b^s(m) h^s)^2}{1 - \beta} \end{aligned}$$

To determine optimal attention m , observe that in the 1-step smax, at the beginning, $m^V = 0$, so the perceived value function is

$$V(x, m^V = 0) = u(c(x), x) + \beta V(x, m^V = 0)$$

so

$$V(x, m^V = 0) = \frac{u(c(x), x)}{1 - \beta} = \frac{-\frac{1}{2}(\beta b^s h^s)^2 - Bx}{1 - \beta}$$

This implies: $b^s = -\frac{B}{1-\beta}$, and

$$\begin{aligned} c &= x + A + \beta b^s (m^V = 0) h m \\ &= x + A - \frac{\beta B}{1 - \beta} h m \end{aligned}$$

so that the impact of thinking more about h , while keeping the future value function constant is

$$\frac{\partial c}{\partial m} = \beta b^s (m^V = 0) h = -\frac{\beta B h}{1 - \beta}$$

Hence, optimal attention is:

$$m = \mathcal{A} \left(\frac{1}{\kappa} \left(\frac{\partial c}{\partial m} \right)^2 u_{cc} \right) = \mathcal{A} \left(\frac{1}{\kappa} \left(\frac{\beta B h}{1 - \beta} \right)^2 \right)$$

C.6. Ricardian Equivalence: Reaction to taxes over time

For simplicity, I 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 .⁸⁰ 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, 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 consumption deviation from the default value: $\hat{c}_t = r\hat{w}_t - H$.

Before the enactment of taxes ($t < T$), will the consumer think of the tax H ? That tax lowers the present value of his income by $He^{-r(T-t)}$, so the consumer's response is

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

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.

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

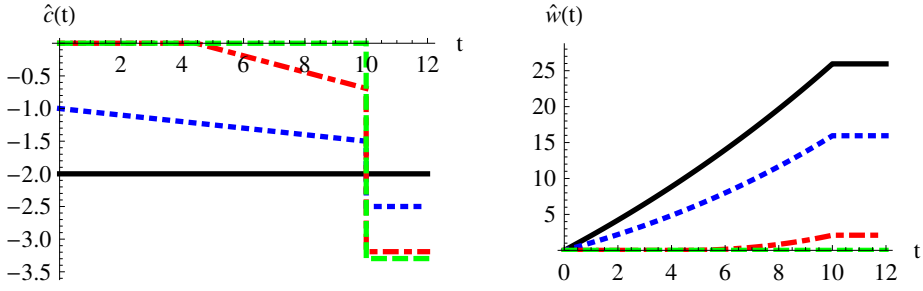


FIGURE C.1. Reaction of consumption and wealth to an increase of future taxes, for different level of κ . *Notes.* At time 0, it is announced that taxes will be paid start at time $T = 10$. This Figure plots the change in consumption and wealth. The solid line is the prediction of the rational model (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.

PROPOSITION C.5. (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 $s = \max(0, \min(T, \frac{1}{r} \ln \frac{\kappa}{He^{-rT}}))$. His consumption path is*

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

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

Let us take an example illustrated in Figure C.1, 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), $\kappa = 0.1$ (green, dashed). The traditional Ricardian consumer ($\kappa = 0$) immediately decreases his consumption by 2%, which leads to wealth accumulation until time T . In contrast, the 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 of 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 cutting 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).

Proof of Proposition C.5. Taxes lower the present value of his income by $He^{-r(T-t)}$, so the consumer's response is:

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

so wealth accumulation is: $\frac{d}{dt}\hat{w}_t = r\hat{w}_t - \hat{c}_t = \tau \left(He^{-r(T-t)}, \kappa \right)$. 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.

$$s = \max \left(0, \min \left(T, \frac{1}{r} \ln \frac{\kappa}{He^{-rT}} \right) \right) \quad (\text{C.24})$$

First, consider the case $s < T$.

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

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

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

So at $t = T$

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

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

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

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

We conclude that consumption is

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

and wealth is

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

C.7. 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 not matter: the impact of interest rates, future taxes, future income shocks etc. are the same whether a sparse agent uses the consumption frame or saving frame. However, the frame does matter for one variable: current income. Indeed, take the permanent-income setup.⁸¹

Which frame does the agent use? One might posit that the agent takes the frame that yields the higher expected utility. To analyze this, we note the following result.

PROPOSITION C.6. (Welfare under the consumption vs savings frame) *The consumption frame yields greater utility than the savings frame if and only if $\varphi_y > r$, i.e. if income shocks mean-revert faster than the interest rate.*

When $\varphi_y > r$ (which is probably the relevant case), the “consumption” frame is indeed better for the agent. The reason is that consumption should be smooth, while savings could be bumpy as they absorb transitory income shocks. When the agent chooses consumption in an inattentive manner, it makes consumption automatically rather smooth. However, if the agent chooses savings inattentively, he makes savings smooth, but consumption needs to absorb the shocks, hence is quite volatile. Therefore, generally, to keep consumption smooth, choosing consumption inattentively is better than choosing savings inattentively. However, when income shocks are a random walk ($\varphi_y = 0$), 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.

81. Recall that $\hat{c}_t^r = \frac{r}{r+\varphi} \hat{y}_t$, so

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

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

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

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

Proof of Proposition C.6. We use the content⁸² and notations of Proposition F.1. We set $x_t = \hat{y}_t$. We have $F^w(w, x, c) = rw + x_t - c_t$ and $F^x(w, x) = -\varphi x$.

Under the *consumption frame*, $a_t = c_t$, and $F_{aa}^w = 0$, so by Proposition F.1, noting $[V_{xx}^\delta]^C$ the value of $V_{xx}^\delta(w, 0)$ under the consumption frame:

$$[V_{xx}^\delta]^C = \frac{u''(c)}{r + 2\varphi_y} (c_y^s - c_y^r)^2 \quad (\text{C.25})$$

and as $c_y^s = mc_y^r$ with $c_y^r = \frac{r}{r+\varphi}$,

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

and the expected losses are (with $\sigma_y^2 = \mathbb{E}[\hat{y}_t^2]$)

$$\begin{aligned} L^C &= \frac{-1}{2} [V_{xx}^\delta]^C \sigma_y^2 = \frac{-1}{2} \frac{u''(c)}{r + 2\varphi} \sigma_y^2 (1 - m)^2 \left(\frac{r}{r + \varphi} \right)^2 \\ &= A(1 - m)^2 r^2 \end{aligned}$$

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

$$[V_{xx}^\delta]^S = \frac{u''(c)}{r + 2\varphi} (S_y^s - S_y^r)^2$$

and as $S_y^s = mS_y^r$, with $S_y^r = 1 - c_y^r = \frac{\varphi}{r+\varphi}$,

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

and expected losses are

$$L^S = \frac{-1}{2} [V_{xx}^\delta]^S \sigma_y^2 = A(1 - m)^2 \varphi^2$$

The consumption frame yields greater utility than the savings frame if and only if $\varphi_y > r$.

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

$$L^x = \frac{-u''(c)}{r + 2\varphi} \sigma_x^2 (c_x^s - c_x^{rat})^2 = \frac{-u''(c)}{r + 2\varphi} \sigma_x^2 c_x^2 (1 - m_x)^2$$

We parametrize the losses by the “equivalent permanent tax” λ^x such that

$$L^x = \mathbb{E} \int_0^\infty e^{-\rho t} [u(c_t) - u(c_t(1 - \lambda))] dt$$

82. We could also draw on the results in Cochrane (1989), with a variety of adjustments. Proposition F.1 extend Cochrane’s results (derived for consumption) to general dynamic problems.

Hence, using a Taylor expansions, $\lambda^x = \frac{L}{u'(c)c/r}$. This gives

$$\lambda^x = \frac{1}{2} \frac{-\frac{u''(c)}{u'(c)c/r}}{r + 2\varphi} \sigma_x^2 c_x^2 (1 - m_x)^2$$

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

$$\lambda^x = \frac{1}{2} \frac{r\gamma}{r + 2\varphi} \left[\frac{c_x \sigma_x}{c} (1 - m_x) \right]^2 \quad (\text{C.26})$$

PROPOSITION C.7. *The losses from paying only attention m_x to variable x , expressed in terms of an “equivalent proportional losses in consumption”, λ^x are*

$$\lambda^x = \frac{1}{2} \frac{r\gamma}{r + 2\varphi_x} \left[\frac{c_x \sigma_x}{c} (1 - m_x) \right]^2 \quad (\text{C.27})$$

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

The calibration gives

$$\lambda^r = (1 - m_r)^2 \times 0.03\%, \quad \lambda^y = (1 - m_y)^2 \times 3.0\% \quad (\text{C.28})$$

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 : $\hat{y}_t = \rho^{t-1} \varepsilon$ for $t = 0, 1, 2$, with ε a mean-0 shock. Normatively, that should induce the change $\hat{c} = (\hat{c}_t)_{t=0,1,2} = (1, 1, 1) \frac{1+\rho+\rho^2}{3} \varepsilon$ (indeed, the total value of income has increased by $(1 + \rho + \rho^2) \varepsilon$). Let us now consider a BR agent with $m = 0$. However, under the consumption frame, $\hat{c}^C = (0, \frac{1}{2}, \frac{1}{2} + \rho + \rho^2) \varepsilon$ (as there is no reaction of c_0 , so that time-1 wealth increases by $\hat{w}_1 = \varepsilon$, of which half is consumed at time 1, so $\hat{c}_1^C = \frac{\varepsilon}{2}$). Under the savings frame, we get $\hat{c}^S = (1, \rho, \rho^2) \varepsilon$ (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 .⁸³ Indeed, when $\rho = 0$, $\hat{c}^C = (0, \frac{1}{2}, \frac{1}{2}) \varepsilon$ and $\hat{c}^S = (1, 0, 0) \varepsilon$, so there is more smoothing under the consumption frame. Other the other hand, with $\rho = 1$, $\hat{c}^C = (0, \frac{1}{2}, \frac{5}{2}) \varepsilon$ and $\hat{c}^S = (1, 1, 1) \varepsilon$, and there is more smoothing under the savings frame.

C.8. Intertemporal elasticity of substitution: controversies about its value

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

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

of 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

$$\ln c_{t+1} - \ln c_t = \psi^e \frac{r_t}{R} + a + \varepsilon_{t+1} \quad (\text{C.29})$$

where ψ^e is the measured IES, a is a constant, and ε_{t+1} is mean-zero noise

I apply the infinite-horizon framework of Section 4.2. 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), then the estimated IES will be: $\psi^e = m_r \psi$. This is a point prediction that goes beyond Chetty (2012)’s prediction of an interval bound. The formal statement is cleanest to write in the limit of no savings ($w_t - \bar{y} = 0$). It is as follows.

PROPOSITION C.8. *An econometrician fitting an Euler equation as in (C.29) will estimate a downwardly-biased IES (intertemporal elasticity of substitution) if the agent is sparse:*

$$\psi^e = m_r \psi \quad (\text{C.30})$$

where ψ^e is the estimated IES, ψ the true IES and m_r is the attention to the interest rate, endogenized in Proposition 8.

Proof: We use the results of Section 4.2. Given an interest rate deviation $\hat{r}_t$, the behavioral consumer increases his consumption by $\hat{c}_t = m_r B_r \hat{r}_t$, which decreases next period wealth by $\hat{w}_{t+1} = -R m_r B_r \hat{r}_t$, and so that next period consumption deviation is

$$\hat{c}_{t+1} = m_r B_r \hat{r}_{t+1} + \frac{r}{R} \hat{w}_{t+1} = m_r B_r (\hat{r}_{t+1} - r \hat{r}_t)$$

which implies, as $\mathbb{E}[\hat{r}_{t+1}] = \rho_t \hat{r}_t$,

$$\mathbb{E}[\hat{c}_{t+1} - \hat{c}_t] = (m_r B_r (\rho_r - r - 1)) \hat{r}_t = -m_r b_r \hat{r}_t = -m_r \frac{\psi}{R} c^d \hat{r}_t$$

Hence, the regression coefficient in (C.29) is $\psi^e = m_r \psi$. $\square$

The above calibration yields Figure C.2, which plots the measured IES ψ^e 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.⁸⁴

84. 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 et al. (2010) for a model where agents’ use of simplified models leads to departures from the standard aggregate model.

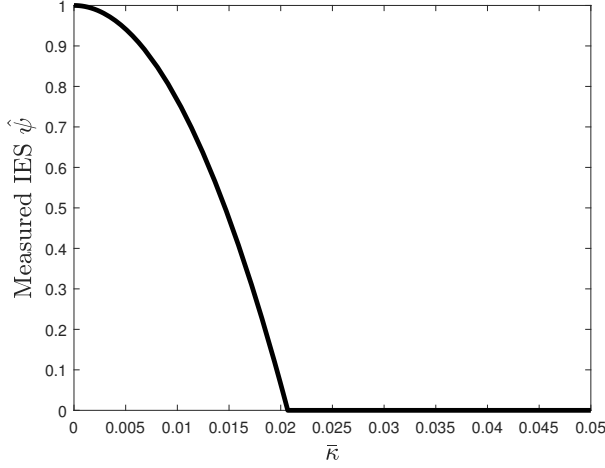


FIGURE C.2. Measured intertemporal elasticity of substitution (IES), $\hat{\psi}$, if the consumer is sparse with cost κ , while the econometrician assumes he is fully rational. The true IES is $\psi = 1$.

C.9. Source-dependent Marginal Propensity to Consume

The agent has initial wealth w and 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 $\max_c V = u(c) + \mathbb{E}[v(y + R(w - c))]$, 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 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 this model, we obtain (assuming exponential utility with absolute risk aversion γ for simplicity)

PROPOSITION C.9. *Time-1 consumption is $c = \frac{1}{1+R}(Rw + \delta/\gamma - \gamma\sigma_\varepsilon^2/2 + y_* + \sum_i m_i y_i)$, $m_i = \tau(1, \frac{\kappa^m \sigma_{c2}}{\sigma_{y_i}})$. The marginal propensity to consume (MPC) at time 1 out of income source i is*

$$MPC_i^s = MPC_i^r \cdot m_i, \quad (\text{C.31})$$

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.

Different income sources have different marginal propensities to consume – reminiscent of Thaler (1985)’s mental accounts. Equation (C.31) makes another prediction: 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 (C.31), lends itself more directly to empirical evaluation.

C.10. Cognitive Discounting of Future News

Suppose that the agent is told that he may receive \$7 in 3 periods, but that those \$7 may disappear, with a survival probability of ρ_y per period. Then, the \$7 should count at $\rho_y^3 7$. However, a sparse agent could replace ρ_y by $\rho_y(m_{y'}) = m_{y'}\rho_y + (1 - m_{y'})\rho_y^d$. If $\rho_y(m_{y'}) < \rho_y$, he's thinking "Let me not bother with those potential future things", or "I'll believe it when I see it", or something akin to "a bird in the hand is worth two in the bush".

The framework easily accommodates that behavior of "cognitive discounting". Call f_t^y the vector of future flows, whose s -th component will arrive in s periods.⁸⁵ In the rational model, with no decay,

$$f_{t+1}^y = Lf_t + \varepsilon_{t+1}^{f^y}$$

where L is the left-shift operator $L(f_1, f_2, f_3, \dots) = (f_2, f_3, \dots)$. An innovation $\varepsilon_{t+1}^{f^y}$ codifies announcement. For instance, in the initial example $\varepsilon_t^{f^y} = (0, 0, 7, 0, 0, \dots)$. When there is cognitive discounting, operator L is replaced by $\rho_{y'}(m)L$.

To come back to our consumption problem, the problem is, under the subjective model (with $Z = (w, r_t, f_t^y, f_t^r)$):

$$\begin{aligned} F^w(Z, m) &= (R + m_r \hat{r}_t)(w_t - c_t) + \bar{y} + m_y \hat{y}_t \\ f_{t+1}^x &= m_{fx} Lf_t^x + \varepsilon_{t+1}^x, \quad \hat{x}_t = k_1^x \cdot f_t^x \text{ for } x = \hat{r}, \hat{y} \end{aligned}$$

Here, the value of m_{fx} "dampens" the appreciation of future movements in variable x . Intuitively, because the future is harder to predict, its simulations are dampened.

85. More generally, $E_t[y_{t+s}] = (f_t^y)_s = k_s^y \cdot f_t^y$, with $k_s^y = (0, \dots, 0, 1, 0, 0, \dots)$ the vector selecting the s -th component.

PROPOSITION C.10. *In the cognitive discounting specification, the behavioral policy is:*

$$\hat{c}_t^s = \mathbb{E}_t \left[\sum_{\tau \geq t} \frac{1}{R^{\tau-t+1}} \left(m_r m_{r'}^{\tau-t} b_r(w_t) \hat{r}_\tau + m_y m_{y'}^{\tau-t} b_y \hat{y}_\tau \right) \right], \quad (\text{C.32})$$

All those expressions hold up to second order terms.

Formulation (C.32) encapsulates two different forms of inattention. First, the agent may not think about interest rate at all if $m_r = 0$. Second, he may discount future news, if $m_{r'} < 1$. Indeed, he discounts future news arriving in T periods by a factor $m_{r'}^T$. In addition, this discounting is source-specific: if news about future interest rates are less important than news about future income (something we will compute soon), they are (cognitively) discounted more.⁸⁶

C.11. Extension of the basic 3-period example

Using the simplification function. The value (39) is a bit complicated. This is where the simplification operator S (defined in Definition F.2) intervenes. Applying it (with the same notations as in the motivating example before and after Definition F.2), we obtain $V^{1,S} := S(V^1)$, i.e.

$$V^{1,S}(w, x) = 2u \left(\frac{w_1 + x}{2} \right) \quad (\text{C.33})$$

The value is the same as V^1 , up to $O(x^2)$ terms: $V^1(w_1, x) = V^{1,S}(w, x) + O(x^2)$. The attention-augmented value function at time 1 is

$$V^1(w, x, m^V) = m^V V^1(w, x) + (1 - m^V) V^{1,S}(w, x)$$

At time 0, the agent does $\text{smax}_{c_0; m} v^0(c_0, x, m_0)$, with $m_0 = (m_0^x, m^V)$ and

$$v^0(c_0, w_0, x, m_0) := u(c_0) + V^1(w_0 - c_0, m_0^x x, m^V) \quad (\text{C.34})$$

The FOC is $v_{c_0}^0 = 0$ with

$$v_{c_0}^0 = u'(c_0) - V_w^1(w_0 - c_0, m_0^x x, m^V).$$

We have $V_{w, m^V}^1 = 0$ at the default $m_0^d = (0, 0)$, so $\frac{\partial c_0}{\partial m^V}|_{m^d=0} = 0$ and the optimal attention is $m^V = 0$: the agent uses the proxy value function, not the exactly rational one (we will see soon that attention m^V can be non-zero using the 2-step smax , but it is still likely to be 0 if κ is not too small).

86. Psychologically, why the decay? This may be because often “promises are not kept” , or “something intervenes” (so that the agent anchors on a decay $\rho_y^d < 1$), or simply because things far in the future are generally less easy to perceive.

We note that if $m^V > 0$, the FOC is more complex. The FOC is

$$u'(c_0) = (1 - m^V) u' \left(\frac{w_1 + m_0 x}{2} \right) + m^V \frac{1}{2} \left[u' \left(\frac{w_1 + m_1 x}{2} \right) + u' \left(\frac{w_1 + (2 - m_1) x}{2} \right) \right]$$

Still, to the first order, the decision is the same (as per Proposition 2). Making the problem simpler at every period, via the $m^V = 0$ device, makes the problem more tractable for both the agent and the economist examining him. We next study this, using the 2-step sparse max.

Using the 2-step sparse max. We have so far used the plain sparse max. This led to $m^V = 0$, the exclusive reliance on the simplified value function. We now calculate what happens when using the twice-iterated sparse max of Definition F.1.

To endogenize m^V , we use the twice-iterated smax: $\text{smax}_{c;m}^2 v^0(c_0, w_0, x, m)$ with $m = (m_0^x, m^V)$. At the first round, $v_{c_0, m^V}^0 = 0$, so $m_0^V(1) = 0$, and, as before, $m_0^x(1) = \mathcal{A} \left(\frac{1}{6\kappa} u'' \left(\frac{w_0}{3} \right) \sigma_x^2 \right)$.

At the second round, now $m^d = (m_0^x(1), 0)$. The easy part is the attention to x , which is slightly different than at step 1:

$$m_0^x(1) = \mathcal{A} \left(\frac{1}{6\kappa} u'' \left(\frac{w_0 + m_0^x(0) x}{3} \right) \sigma_x^2 \right)$$

The more novel part is to calculate m_V . We have, with $w_1 = w_0 - c_0$ and calling $x^s := m_0^x x$,

$$\begin{aligned} v_{c, m^V}^0(c_0, w_0, x, m_0^x, m^V) &= \partial_c [V^1(w_0 - c_0, x^s) - V^1(w_0 - c_0, x^s, m^V = 0)] \\ &= -\frac{1}{2} u' \left(\frac{w_1 + m_1 x^s}{2} \right) - \frac{1}{2} u' \left(\frac{w_1 + (2 - m_1) x^s}{2} \right) + u' \left(\frac{w_1 + x^s}{2} \right) \end{aligned}$$

Doing a Taylor expansion of the consumptions $\frac{w_1 + m_1 x^s}{2}$ and $\frac{w_1 + (2 - m_1) x^s}{2}$ around their mean

$$c^d = \frac{w_1 + x^s}{2} = \frac{w_1 + m_0^x x}{2}$$

we obtain

$$\begin{aligned} v_{c, m^V}^0 &= -\frac{1}{2} u' \left(c^d + (m_1 - 1) \frac{x^s}{2} \right) - \frac{1}{2} u' \left(c^d - (m_1 - 1) \frac{x^s}{2} \right) + u'(c^d) \\ &= -\frac{1}{2} u'''(c^d) (m_1 - 1)^2 \left(\frac{x^s}{2} \right)^2 \times 2 + o(x^2) \\ &= -\frac{1}{4} u'''(c^d) (m_1 - 1)^2 (m_0^x x)^2 + o(x^2) \end{aligned}$$

Likewise, $v_{cc|m=m_0^d(1)}^0 = \frac{3}{2} u''(c^d)$. So, the impact of m_V is

$$\frac{\partial c_0}{\partial m_V} = -\frac{v_{c, m^V}^0}{v_{cc}^0} = -\frac{1}{6} \frac{u'''(c^d)}{u''(c^d)} (m_1 - 1)^2 (m_0^x x)^2 + o(x^2)$$

Hence, for a small x , the attention m^V to the difference between the difference between the true and proxy value functions (i.e., $V^1(w_1, x, m^V)$ for $m^V = 1$ vs $m^V = 0$) is:

$$\begin{aligned}
m_0^V &= A \left(\frac{1}{\kappa} \mathbb{E} \left[\left(\frac{\partial c_0}{\partial m^V} \right)^2 v_{cc}^0 \right] \right) \\
&= A \left(\frac{1}{\kappa} \mathbb{E} \left[\left(\frac{1}{6} \frac{u'''(c^d)}{u''(c^d)} (m_1 - 1)^2 (m_0^x)^2 x^2 \right)^2 \right] \frac{3}{2} u''(c^d) \right) \\
&= A \left(\frac{1}{24\kappa} \left(\frac{u'''(c^d)}{u''(c^d)} \right)^2 (m_1 - 1)^4 (m_0^x)^4 \mathbb{E}[x^4] u''(c^d) \right) \quad (\text{C.35})
\end{aligned}$$

It is instructive to take the limit of small κ using a sparsity-inducing cost function ($g'(0) > 0$). To have $m_0^x > 0$, we need $\frac{\sigma_x^2}{\kappa}$ large enough, so $\sigma_x \succeq \kappa^{1/2}$. To have $m_0^V > 0$, we need $\frac{\sigma_x^4}{\kappa}$ large enough, i.e. $\sigma_x \succeq \kappa^{1/4}$, which is a much higher hurdle ($\frac{\kappa^{1/4}}{\kappa^{1/2}} \rightarrow \infty$) for small κ . We formalize this.

PROPOSITION C.11. (Attention to a variable, vs attention to the fine properties of how the value function depends on that variable) *Suppose a succession of problems (indexed by κ going to 0) such that there are positive constants B, B', ε such that for κ small enough: $B\kappa^{1/2-\varepsilon} \leq \sigma_x(\kappa) \leq B'\kappa^{1/4+\varepsilon}$. Then, the agent will have $m_0^x > 0$ and $m_0^V = 0$ when κ is small enough. This is, the agent pays attention to the disturbance x , but not to the subtle difference between the true and proxy value functions (i.e., $V^1(w_1, x, m^V)$ for $m^V = 1$ vs $m^V = 0$).*

In plain terms: because thinking about the nuances m^V in $V(x, m^V)$, one needs to think about x at all. Hence, in many situations, we have $m^V = 0$ and $m^x > 0$. Indeed, we cannot have, with just one state, variable $m^x = 0$ and $m^V > 0$.

In particular, for our 3-period problem for κ small enough but not too small, $m^V = 0$ and $m_0^x > 0$ - the agent uses the simplified value function, but still pays attention to x , like in the basic smax case. This is one reason it is useful to use the basic smax: it gets to the essence of the more complex patterns that can later be refined using the iterated smax.

Appendix D: Complements to the life-cycle model

Here I record variants on the life-cycle models of Section 2. What happens with various attention functions, in discrete and continuous time, etc. This is useful to get a “feel” for the model in a concrete, substantial setting.

D.1. Derivation in continuous time

It is instructive to do the proof in continuous time, using the notations of Section 2. The budget constraint is $\dot{w}_t := \frac{dw_t}{dt} = y_t - c_t$, where a dot denotes a time-derivative.

If the agent is rational, his value function is

$$V^r(w_t, x, t) = (T - t) u\left(\bar{y} + \frac{w_t - x}{T - t}\right)$$

Indeed, the optimal policy is to smooth consumption over the time $[t, T]$, exhausting final resources, which are $w_t + (T - t)\bar{y} - x$, so that $c_t = \frac{w_t + (T - t)\bar{y} - x}{T - t}$, and the value function is $V^r = (T - t)u(c_t)$. I take $V^p = V^r$ for the proxy value function.

Hence, at time t , the Bellman equation is

$$\max_{c; m} v(c_t, m, x, t) \text{ with } v(c_t, m, x, t) := u(c_t)dt + V^r(w_t - c_t dt, mx, t).$$

The FOC is:

$$\begin{aligned} v_c &= u'(c_t)dt - V_w^r(w_t - c_t dt, mx, t)dt \\ &= [u'(c_t) - V_w^r(w_t, mx, t)]dt + O(dt^2) \end{aligned}$$

$$\begin{aligned} v_{cc} &= u''(c_t)dt + V_w^r(w_t - c_t dt, mx, t)(dt)^2 \\ &= u''(c_t)dt + O(dt^2) \end{aligned}$$

Hence, the optimal policy is given by $v_c = 0$, i.e. $u'(c_t) = V_w^r(w_t, mx, t) = u'\left(\bar{y} + \frac{w_t - mx}{T - t}\right)$, so

$$c_t = \bar{y} + \frac{w_t - mx}{T - t} = c_t^d - mb_t,$$

where

$$b_t := \frac{x}{T - t}$$

is the normative sensitivity response to x .

Next, for the allocation of attention, we form

$$\max_m \frac{-1}{2} v_{cc|m=0} \left(\frac{\partial c_t}{\partial m} \right)^2 (m_t - 1)^2 - \kappa_t g(m) dt$$

i.e. the attention is

$$m_t^* = \mathcal{A} \left(\frac{|u''(c_t^d)| b_t^2}{\kappa} \right)$$

Using the definition of the truncation function $\tau(b, k) := b\mathcal{A}\left(\frac{b^2}{k^2}\right)$ (equation 12) the response is

$$c_t - c_t^d = m_t^* b_t = \tau \left(b_t, \sqrt{\frac{\kappa_t}{|u''(c_t^d)|}} \right) \quad (\text{D.1})$$

D.2. Calculation of the consumption path

The basic result (Proposition 1) gives the consumption policy, which is enough to simulate the consumption path, say with a computer. Here I explore specifications to obtain an analytic representation. As always, continuous time is much simpler to derive and cleaner than discrete time. Still, because it is more elementary, I start with discrete time.

D.2.1. Consumption Path in Discrete Time. We will see that the path is as follows. Consider

$$B(s) := \frac{x}{T-s} - \frac{\bar{\kappa}^2}{x} (T-s-1) \quad (\text{D.2})$$

which is related to marginal net benefit of thinking.

PROPOSITION D.1. (Basic life-cycle: consumption path in discrete time) *Call s the first time at which the agent thinks about retirement, $s = \inf \{s \in [0, L] : B(s) \geq 0\}$. Hence, the agent can think of retirement at time $s = 0$ (this is the case if $B(0) \geq 0$), at time a later time before retirement (this is the case if $B(0) < 0 < B(L)$, and s is the solution of $B(s) = 0$ if that is an integer). Or he may never think about it until actual retirement, $s = L$.*

Before he thinks of retirement ($t < s$), consumption is $c_t = \bar{y} + \frac{w_0}{T}$ and $w_t = (1 - \frac{t}{T}) w_0$. After the agent thinks of retirement ($t \in [s, L]$):

$$c_t = \frac{w_0}{T} + \bar{y} - \frac{2\kappa^2}{x} (t-s) - B(s)$$

and wealth is $w_t = (1 - \frac{t}{T}) w_0 + \frac{\kappa^2}{x} (t-s)(t-s-1) + B(s)(t-s)$ for $t \in [s, L]$.

Consumption after retirement (for $t \in [L, T)$) is constant, at $c_t = \frac{w_L}{T-K} + \bar{y} + \hat{y}$.

We next derive this. We follow Section 2, which gives:

$$c_t = \frac{w_t - m_t x}{T-t} + y$$

We suppose the scaling: $\kappa = \bar{\kappa}^2 u''(c_t^d)$, given $v'' = u'' \left(1 + \frac{1}{T-t-1}\right)$,

$$\begin{aligned} m_t &= \mathcal{A} \left(\frac{v_{cc} c_m^2}{\kappa} \right) = \mathcal{A} \left(\frac{v'' c_m^2}{u'' \bar{\kappa}^2} \right) = \mathcal{A} \left(\left(1 + \frac{1}{T-t-1}\right) \left(\frac{x}{T-t}\right)^2 \right) \\ &= \mathcal{A} \left(\frac{1}{(T-t-1)(T-t)} \frac{x^2}{\bar{\kappa}^2} \right) \end{aligned}$$

Using $\mathcal{A}(y) = \max\left(1 - \frac{1}{|y|}, 0\right)$,

$$\begin{aligned} m_t \frac{x}{T-t} &= \frac{x}{T-t} \left(1 - \frac{\bar{\kappa}^2 (T-t-1)(T-t)}{x^2}\right) = \frac{x}{T-t} - \frac{\bar{\kappa}^2 (T-t-1)}{x} \\ &= B(s) \end{aligned}$$

where $B(s)$ was defined in (D.2).

To concentrate on the key difficulty, I take the case $w_0 = 0$.

If the solution is interior ($s \in (0, T)$), the agent thinks about retirement for s such that

$$B(s) = 0 \quad (\text{D.3})$$

Hence, when attention is positive, we have:

$$c_t = y + \frac{w_t - x}{T-t} + \frac{\kappa^2 (T-t-1)}{x} \quad (\text{D.4})$$

We need to calculate the case where the agent saves before retirement. We look for a solution of the type, for $t \in [s, L]$:

$$c_t = y + A(t-s) + b$$

for some constants A, b . At $t = s$, (D.4) implies $b = -B(s)$. For time $t \in [s, L]$, we have:

$$w_t = - \sum_{t'=s}^{t-1} (c_{t'} - y) = -\frac{A}{2} (t-s)(t-s-1) + B(s)(t-s)$$

We want to verify (D.4), which we will express: $g(t) = 0$ with

$$\begin{aligned} g(t) &:= (T-t) \left(c_t - y - \frac{w_t - x}{T-t} - \frac{\kappa^2 (T-t-1)}{x} \right) \\ &= (T-t)(c_t - y) - w_t + x - (T-t)(T-t-1) \frac{\kappa^2}{x} \\ &= (T-t)(A(t-s) - B(s)) + \frac{A}{2} (t-s)(t-s-1) \\ &\quad + B(s)(t-s) + x - (T-t)(T-t-1) \frac{\kappa^2}{x} \end{aligned}$$

This is a polynomial in t . The coefficient of t^2 must be 0, so $-\frac{A}{2} - \frac{\kappa^2}{x} = 0$, and

$$A = \frac{-2\kappa^2}{x}$$

To make sure that $g(t) = 0$, we check first $g(s) = 0$. But

$$g(s) = -(T-s)B(s) + x - \frac{\bar{\kappa}^2 (T-s)(T-s-1)}{x} = 0$$

because of (D.2). Finally,

$$g(T) = -\frac{\kappa^2}{x}(T-s)(T-s-1) - B(s)(T-s) + x = 0$$

again. Given g is a polynomial of degree at most 1, with $g(s) = g(T) = 0$, g is identically 0.

D.2.2. Consumption Path in Continuous Time.

Using the scaling $\kappa_t = \bar{\kappa}^2 |u''(c_t^d)|$. I set: $\kappa_t = \bar{\kappa}^2 |u''(c_t^d)|$. The consumption policy before retirement is

$$c_t = \frac{w_t}{T-t} + \tau \left(\frac{-x}{T-t}, \kappa \right) + y$$

The agent will not think about the change in income when $\frac{x}{T-t} < \kappa$. Call $s \in [0, L)$ the first moment when he thinks about it. If the solution is interior, $\frac{x}{T-s} = \kappa$, i.e. $s = T - \frac{x}{\kappa}$. In general, we need to windorize at 0 and L :

$$s := \max \left(0, \min \left(L, T - \frac{x}{\kappa} \right) \right) \quad (\text{D.5})$$

Note that this is for $\hat{y} \leq 0$. In general have $s := \max \left(0, \min \left(L, T - \frac{|x|}{\kappa} \right) \right)$.

For $t \leq L$, define the deviations from the policy that doesn't pay attention to retirement x :

$$\begin{aligned} \hat{c}_t &= c_t - \frac{w_0}{T} - y \\ \hat{w}_t &= \int_0^t -\hat{c}_t dt = w_t - \frac{T-t}{T} w_0 \end{aligned}$$

We plug $\hat{c}_t$ and $\hat{w}_t$ into the optimal consumption policy to get

$$\hat{c}_t = \frac{\hat{w}_t}{T-t} + \tau \left(\frac{-x}{T-t}, \kappa \right)$$

First case, using truncation function $\tau_1(b, \kappa) = b \max \left(1 - \frac{\kappa^2}{b^2}, 0 \right)$. I first assume that $s < L$. For $t \in [s, L)$ the agent's response is

$$\begin{aligned} \hat{c}_t &= \frac{\hat{w}_t}{T-t} + \tau \left(\frac{-x}{T-t}, \kappa \right) = \frac{\hat{w}_t}{T-t} - \frac{x}{T-t} \left(1 - \frac{\kappa^2}{\left(\frac{x}{T-t} \right)^2} \right) \\ &= \frac{\hat{w}_t - x}{T-t} - \frac{\kappa^2}{x} (T-t) \end{aligned}$$

Take derivative with respect to t , this yields, using $\frac{d\hat{w}_t}{dt} = -\hat{c}_t$,

$$\begin{aligned}\frac{d\hat{c}_t}{dt} &= \frac{-\hat{c}_t}{T-t} + \frac{\hat{w}_t - x}{(T-t)^2} + \frac{\kappa^2}{x} \\ &= \frac{-\hat{c}_t}{T-t} + \frac{\hat{c}_t + \frac{\kappa^2}{x}(T-t)}{(T-t)} + \frac{\kappa^2}{x} \\ &= \frac{2\kappa^2}{x}\end{aligned}$$

with boundary condition that $\hat{w}_s = 0$. Solving for $\hat{c}_t$, for $t \in [s, L]$

$$\hat{c}_t = \frac{2\kappa^2}{x} (t - s)$$

The wealth at time t is

$$\hat{w}_t = \int_s^t -\hat{c}_{t'} dt' = -\frac{\kappa^2}{x} (t - s)^2$$

Second case. Now use truncation function:

$$\tau_{L_1}(b, \kappa) = \text{sign}(b) \max(|b| - |\kappa|, 0).$$

After time s , the agent's response is

$$\hat{c}_t = \frac{\hat{w}_t}{T-t} + \tau\left(\frac{-x}{T-t}, \kappa\right) = \frac{\int_s^t -\hat{c}_t dt}{T-t} - \left(\frac{x}{T-t} - \kappa\right)$$

Taking derivative with respect to t yields

$$\begin{aligned}\frac{d\hat{c}_t}{dt} &= \frac{-\hat{c}_t}{T-t} + \frac{\hat{w}_t}{(T-t)^2} - \frac{x}{(T-t)^2} \\ &= \frac{-\hat{c}_t}{T-t} + \frac{\hat{c}_t + \frac{x}{T-t} - \kappa}{(T-t)} - \frac{x}{(T-t)^2} \\ &= \frac{-\kappa}{T-t}\end{aligned}$$

with boundary condition that $\hat{c}_s = 0$. We can solve $\hat{c}_t$ as,

$$\hat{c}_t = \kappa \ln\left(\frac{T-t}{T-s}\right)$$

The wealth at time $t \leq L$ will be

$$\hat{w}_t = \int_s^t -\hat{c}_t dt = \kappa(T-s) \left(\frac{T-t}{T-s} \ln \frac{T-t}{T-s} - \frac{T-t}{T-s} + 1 \right)$$

Case 1: With $\tau_1(b, \kappa) = b \max\left(1 - \frac{\kappa^2}{b^2}, 0\right)$, his consumption path is

$$c_t = \begin{cases} \frac{w_0}{T} + y & \text{for } t < s \\ \frac{w_0}{T} + y + 2\frac{\kappa^2}{x} \frac{t-s}{T-s} & \text{for } s \leq t < L \\ \frac{w_L}{T-L} + y + \hat{y} & \text{for } t \geq L \end{cases}$$

with $w_L = \hat{w}_L + \frac{B}{T}w_0 = \kappa \frac{(L-s)^2}{T-s} + \frac{B}{T}w_0$.

Case 2: With $\tau_{L_1}(b, \kappa) = \text{sign}(b) \max(|b| - |\kappa|, 0)$, his consumption path is

$$c_t = \begin{cases} \frac{w_0}{T} + y & \text{for } t < s \\ \kappa \ln\left(\frac{T-t}{T-s}\right) + \frac{w_0}{T} + y & \text{for } s \leq t < L \\ \frac{w_L}{T-L} + y + \hat{y} & \text{for } t \geq L \end{cases}$$

with $w_L = \hat{w}_L + \frac{B}{T}w_0 = \kappa(T-s) \left(\frac{B}{T-s} \ln \frac{B}{T-s} - \frac{B}{T-s} + 1 \right) + \frac{B}{T}w_0$.

Using a different scaling, $\kappa_t = \bar{\kappa}^2 \left| (c_t^d)^2 u''(c_t^d) \right|$. Here I explore a different scaling $\kappa_t = \bar{\kappa}^2 \left| (c_t^d)^2 u''(c_t^d) \right|$, with $c_t^d = \bar{y} + \frac{w_t}{T-t}$. This renders $\bar{\kappa}$ dimensionless and potentially portable from one situation to the next. Then (D.1) gives:

$$c_t - c_t^d = \tau \left(b_t, \sqrt{\frac{\kappa_t}{|u''(c_t^d)|}} \right) = \tau(b_t, \bar{\kappa} c_t^d)$$

The agent is all set – he just follows that policy. Now let's turn to the economist's role, to trace out the implications of that policy.

I use the truncation function $\tau_{L_1}(b, \kappa) = \text{sign}(b) \max(|b| - |\kappa|, 0)$, because it yields simpler calculations. I assuming $\hat{y} < 0$, to focus on the retirement case. We have, when $b_t := \frac{-x}{T-t} < -\bar{\kappa} c_t^d$, $\tau(b_t, \bar{\kappa} c_t^d) = b_t + \bar{\kappa} c_t^d$, so

$$\begin{aligned} \hat{c}_t &:= c_t - \bar{y} = (c_t^d - \bar{y}) + (c_t - c_t^d) \\ &= \frac{\hat{w}_t}{T-t} + b_t + \bar{\kappa} c_t^d = \frac{\hat{w}_t}{T-t} - \frac{x}{T-t} + \bar{\kappa} \left(\bar{y} + \frac{\hat{w}_t}{T-t} \right) \end{aligned}$$

i.e.

$$\hat{c}_t = \frac{(1 + \bar{\kappa}) \hat{w}_t - x}{T-t} + \bar{\kappa} \bar{y} \quad (\text{D.6})$$

Let us examine the first time s at which the agent thinks about retirement. It if it in $(0, L)$, we have $\hat{w}_s = 0$, and $-\frac{x}{T-t} + \bar{\kappa} \bar{y} = 0$, so $s = T - \frac{x}{\bar{\kappa} \bar{y}}$. In general, we windsorize by 0 and L , and obtain:

$$s := \max \left(0, \min \left(L, T - \frac{x}{\bar{\kappa} \bar{y}} \right) \right) \quad (\text{D.7})$$

To find c_s , we again apply (D.6), observing that $\hat{w}_s = 0$:

$$\begin{aligned} c_s &= \bar{y} - \frac{x}{T} + \bar{\kappa} \bar{y} \text{ if } s = 0 \\ &= \bar{y} \text{ if } s \in (0, L) \end{aligned} \quad (\text{D.8})$$

From then on, we assume $s < L$.

After the agent has starting thinking about retirement, we have (D.6), so taking the time derivative, using $\hat{w}_t = -\hat{c}_t$, we have:

$$\begin{aligned}\hat{c}_t &= \frac{(1 + \bar{\kappa}) \hat{w}_t - x}{(T - t)^2} + \frac{(1 + \bar{\kappa})}{T - t} \hat{w}_t = \frac{\hat{c}_t - \bar{\kappa} \bar{y}}{T - t} - \frac{(1 + \bar{\kappa}) \hat{c}_t}{T - t} \\ &= -\bar{\kappa} \frac{\hat{c}_t + \bar{y}}{T - t}\end{aligned}$$

As $c_t = \hat{c}_t + \bar{y}$, we have $\dot{c}_t = -\kappa \frac{X_t}{T-t}$, so $\frac{d \ln c_t}{dt} = \frac{-\kappa}{T-t}$, $\ln c_t = \kappa \ln(T - t) + a$, and

$$c_t = A (T - t)^\kappa$$

for a constant A . Given the value at c_s

$$c_t = \left(\frac{T - t}{T - s} \right)^\kappa c_s. \quad (\text{D.9})$$

The wealth at time $t \leq L$ is

$$\hat{w}_t = \int_s^t -(c_t - \bar{y}) dt = (t - s) \bar{y} - \frac{1}{1 + \kappa} \left((T - s) - \frac{(T - t)^{1 + \kappa}}{(T - s)^\kappa} \right) c_s. \quad (\text{D.10})$$

D.3. Lifecycle: Policy of the hyperbolic log agent

The log case is particularly clean (Barro (1999)). The agent at time t has decision utility: $\ln c_t + \beta \sum_{\tau=t+1}^{T-1} \ln c_\tau$. Given full wealth (including discounted future income y_t) call Ω_t , his policy will be some $c_t = \mu_t \Omega_t$, for μ_t independent of Ω_t , so we shall have

$$\sum_{\tau=t}^{T-1} \ln c_\tau = (T - t) \ln \Omega_t + K_t$$

for some constant K_t . This implies a consumption policy

$$c_t = \arg \max u(c_t) + \beta (T - t - 1) \ln(\Omega_t - c_t) + K_{t+1}$$

The first order condition is

$$\frac{1}{c_t} = \frac{\beta (T - t - 1)}{\Omega_t - c_t}$$

i.e.

$$c_t = \frac{\Omega_t}{1 + \beta (T - t - 1)}$$

We see that there is no “cliff” at retirement. The agent makes no distinction between pre and post retirement income.

Appendix E: Proofs Omitted in the Paper

Proof of Lemma 7. Exact result. The problem is

$$\max_{(c_t)_{t \geq 0}} \sum_{t \geq 0} \beta_t \frac{c_t^{1-\gamma}}{1-\gamma} \text{ s.t. } \sum_{t \geq 0} q_t c_t \leq \Omega_0$$

where $q_t = \frac{1}{(1+r_0)\dots(1+r_{t-1})}$ is the time-0 Arrow-Debreu price of a dollar received at t , and β_t is the discount factor (which is not necessarily of the form β^t here), and

$$\Omega_0 := w_0 + \sum_{t > 0} q_t y_t$$

is the full wealth. Forming the Lagrangian,

$$L = \sum_{t \geq 0} \beta_t \frac{c_t^{1-\gamma}}{1-\gamma} + \lambda \left(\Omega_0 - \sum_{t \geq 0} q_t c_t \right)$$

we have $\beta_t c_t^{-\gamma} = \lambda q_t$, i.e. (with $\psi = \frac{1}{\gamma}$), $c_t = c_0 \left(\frac{\beta_t}{q_t} \right)^\psi$ for some c_0 . The budget constraint gives:

$$\Omega_0 = \sum_{t \geq 0} q_t c_t = c_0 \sum_{t \geq 0} \beta_t^\psi q_t^{1-\psi}$$

i.e.

$$c_0 = \mu \Omega_0, \quad \mu := \frac{1}{\sum_{t \geq 0} \beta_t^\psi q_t^{1-\psi}} \quad (\text{E.1})$$

Given $V'(\Omega_0) = u'(c_0) = u'(\mu \Omega_0)$, we have (as the function is also homogeneous of degree $1-\gamma$): $V(\Omega_0) = \frac{1}{\mu} u(\mu \Omega_0)$.

Suppose now that $\beta_t = \beta^t$ and $\beta R = 1$. Then the interest rate is constant, $q_t = R^{-t}$, and $\mu = \frac{1}{\sum_{t \geq 0} R^{-t}} = \frac{1}{1-\frac{1}{R}} = \frac{\bar{r}}{R}$ and $\Omega_0 = w_0 + \frac{\bar{y}}{r}$. So, $c_0 = \mu \Omega_0 = \frac{r w_0 + \bar{y}}{R}$.

Taylor expansion, first in the deterministic case.

The impact of a change dy_τ is very easy: $dc_0 = \mu d\Omega_0 = \frac{\bar{r}}{R} \frac{dy_\tau}{R^{\tau+1}}$.

The impact of an interest rate is more delicate. Consider a change change dr_τ , for just one date τ . It creates a bond price change $dq_t = \frac{-1}{R^{t+1}} dr_\tau 1_{t > \tau}$, so that

$$\sum_{t \geq 0} dq_t = \sum_{t \geq 0} \frac{-1}{R^{t+1}} dr_\tau 1_{t > \tau} = \sum_{t \geq \tau+1} \frac{-1}{R^{t+1}} dr_\tau = \frac{-1}{r R^{\tau+1}} dr_\tau$$

This gives

$$\begin{aligned} \frac{d\mu}{\mu} &= -\mu(1-\psi) \sum_{t \geq 0} \beta_t^\psi q_t^{-\psi} dq_t = -\frac{\bar{r}}{R} (1-\psi) \sum_{t \geq 0} dq_t \\ &= (1-\psi) \frac{\bar{r}}{R} \frac{1}{r R^{\tau+1}} dr_\tau = \mu(1-\psi) \frac{dr_\tau}{R^{\tau+2}} \end{aligned}$$

Also, $d\Omega_0 = \bar{y} \sum_{t \geq 1} dq_t = \frac{-\bar{y}}{r} \frac{dr_\tau}{R^{\tau+1}}$. Recalling that $c_0 = \mu\Omega_0$:

$$\begin{aligned} dc_0 &= \mu\Omega_0 \frac{d\mu}{\mu} + \mu d\Omega_0 = c_0 (1 - \psi) \frac{dr_\tau}{R^{\tau+2}} + \frac{\bar{r}}{R} \frac{-\bar{y}}{r} \frac{dr_\tau}{R^{\tau+1}} \\ &= \left(-\psi c_0 + \frac{rw_0 + \bar{y}}{R} - \bar{y} \right) \frac{dr_\tau}{R^{\tau+2}} = \left(-\psi c_0 + \frac{r(w_0 - \bar{y})}{R} \right) \frac{dr_\tau}{R^{\tau+2}} \end{aligned}$$

This gives announced value.

Value function. Write $c_t = \mu_t(w_t - \bar{y}) + \nu_t$, for values μ_t, ν_t independent of w_t . Because one dollar now can be consumed today, we have $V'(w_t) = u'(c_t)$, and indeed $V(w_t) = \frac{1}{\mu_t} u(\mu_t(w_t - \bar{y}) + \nu_t)$. Using (47), in particular $c_t^d = \frac{r(w_t - \bar{y})}{R} + \bar{y}$ and $b_r(w_t) := \frac{\bar{r}(w_t - \bar{y}) - \psi c_t^d}{R}$, we obtain:

$$\begin{aligned} \mu_t w_t + \nu_t &= c_t = c_t^d + \hat{c}_t \\ &= c_t^d + \mathbb{E}_t \left[\sum_{\tau \geq t} \frac{1}{R^{\tau-t+1}} (b_r(w_t) \hat{r}_\tau + b_y \hat{y}_\tau) \right] \\ &= \frac{\bar{r}(w_t - \bar{y})}{R} + \bar{y} \\ &\quad + \mathbb{E}_t \left[\sum_{\tau \geq t} \frac{1}{R^{\tau-t+1}} \left(\frac{\bar{r}(w_t - \bar{y}) - \psi \left(\frac{r(w_t - \bar{y})}{R} + \bar{y} \right)}{R} \hat{r}_\tau + b_y \hat{y}_\tau \right) \right] \end{aligned}$$

which gives the announced values.

Stochastic case. As we are reasoning up to $O(\|x\|^2)$, we can take the certainty equivalent, e.g. use $\mathbb{E}[f(x)] = f(\mathbb{E}[x]) + O(\text{var}(x))$ for a C^2 function f . Technically, we assume that f is C^2 with $|f''(x)|$ bounded over the support of the distribution of X .⁸⁷ This, way, we can move from the deterministic version of (47) to its expectation, capturing absorbing the uncertainty terms in the $O(\|x\|^2)$.

Proof of Lemma 8. In the AR(1) case,

$$\begin{aligned} \mathbb{E}_t \left[\sum_{\tau \geq t} \frac{1}{R^{\tau-t+1}} \hat{r}_\tau \right] &= \sum_{\tau \geq t} \frac{1}{R^{\tau-t+1}} \rho_r^{\tau-t} \hat{r}_t = \frac{1}{1 - \frac{\rho_r}{R}} \frac{\hat{r}_t}{R} = \frac{1}{R - \rho_r} \hat{r}_t \\ \mu_t &= \frac{\bar{r}}{R} + (1 - \psi) \frac{\bar{r}}{R} \mathbb{E}_t \left[\sum_{\tau \geq t} \frac{\hat{r}_\tau}{R^{\tau-t+2}} \right] = \frac{\bar{r}}{R} + (1 - \psi) \frac{\bar{r}}{R^2} \frac{1}{R - \rho_r} \hat{r}_t \end{aligned}$$

87. The proof is standard. Normalizing $\mathbb{E}X = 0$, and calling x here just a real number, we use $f(x) = f(0) + f'(0)x + R(x)$ with $R(x) = \int_0^x f''(y)(x-y)dy$. So $f(X) = f(0) + f'(0)X + R(X)$, and

$$\mathbb{E}f(X) = f(\mathbb{E}X) + \mathbb{E}R(X)$$

Using $|R(x)| \leq Cx^2$ for $C := \frac{1}{2} \sup_{x \in I} |f''(x)|$, we have $|\mathbb{E}f(X) - f(\mathbb{E}X)| \leq C\mathbb{E}[X^2] = C\text{var}(X)$. So, $\mathbb{E}f(X) = f(\mathbb{E}X) + O(\text{var}(X))$.

$$\nu_t = \bar{y} + \mathbb{E}_t \left[\sum_{\tau \geq t} \frac{-\psi \bar{y}}{R} \hat{r}_\tau + b_y \hat{y}_\tau \right] = \bar{y} + \frac{-\psi \bar{y}}{R} \frac{1}{R - \rho_r} \hat{r}_t + b_y \frac{1}{R - \rho_y} \hat{y}_t$$

Proof of Lemma 3. This is a variant on the standard proof. We have

$$\mathcal{T}(V, V^p)(Z) - \mathcal{T}(\tilde{V}, V^p)(Z) = \beta \mathbb{E}[V(Z') - \tilde{V}(Z')] \quad (\text{E.2})$$

with $Z' := F(z, a(Z, V^p), \mu)$.

For (i), using $|V(Z') - \tilde{V}(Z')| \leq \|V - \tilde{V}\|_\infty$,

$$\begin{aligned} |\mathcal{T}(V, V^p)(Z) - \mathcal{T}(\tilde{V}, V^p)(Z)| &\leq \beta \mathbb{E}[|V(Z') - \tilde{V}(Z')|] \\ &\leq \beta \mathbb{E}[\|V - \tilde{V}\|_\infty] = \beta \|V - \tilde{V}\|_\infty \end{aligned}$$

and taking the sup on the left-hand side,

$$\|\mathcal{T}(V, V^p) - \mathcal{T}(\tilde{V}, V^p)\|_\infty \leq \beta \|V - \tilde{V}\|_\infty.$$

For (ii), if $V(Z') - \tilde{V}(Z') \leq 0$ for all Z' , then (E.2) implies that $\mathcal{T}(V, V^p)(Z) - \mathcal{T}(\tilde{V}, V^p)(Z) \leq 0$. The operator is monotone.

Proof of Lemma 4. This is the usual fixed point argument. Define $V_0 := V^p$, and for $n \geq 0$, $V_{n+1} = \mathcal{T}(V_n, V^p)$. By Lemma 3,

$$\|V_{n+1} - V_n\|_\infty = \|\mathcal{T}(V_n, V^p) - \mathcal{T}(V_{n-1}, V^p)\|_\infty \leq \beta \|V_n - V_{n-1}\|_\infty$$

hence V_n is a Cauchy sequence and converges in a complete metric space.

Proof of Proposition 2. Proposition 4 implies that $V(w, x) = V^r(w, x) + O(\|x\|^2)$.⁸⁸ Next, Lemma 5 implies that the optimal policy satisfies $a(w, x) = a^r(w, x) + O(\|x\|^2)$. Next, decompose $a^r(w, x) = a^d(w) + \sum_i b_i(w) x_i + O(\|x\|^2)$. Then, using the policy V without sparse max, we have

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

Finally, using Lemma 2, we have:

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

Proof of Proposition 4. It is in Section G.1.

Proof of Lemma 5. The reasoning is a direct transposition of the arguments in the proof of Proposition 4.

⁸⁸ Proposition 2 is stated in the text before some results this proof uses (this is useful to make the flow of the paper more natural), but there is no logical circularity.

Proof of Lemma 6. Given a value function $V(w, x)$, Lemma A.A.2 shows that, up to $O(\|x\|^2)$ terms, $a(w, x)$ just depend on $V(w, 0)$, $V_w(w, 0)$ (but not on $V_{xx}(w, 0)$). The Lemma assumes that the two functions $V^p(w, x) = V^{p'}(w, x)$ have the same values of $V(w, 0)$, $V_x(w, 0)$. Hence, their actions $a(w, x)$ are the same up to $O(\|x\|^2)$.

Proof of Proposition 5. We will prove by induction on $q \geq 0$ that the following property holds $H_q : V^{(q)}(z) = V^r(z) + O(\|x\|^2)$.

This is true by assumption for $q = 0$. Suppose H_q holds, we will see that H_{q+1} holds. By Lemma 6,

$$a(z, V^{(q)}) = a(z, V^{(0)}) + O(\|x\|^2) \quad (\text{E.3})$$

Because $a^{(0)}(z)$ is C^1 , we also have $a(z, V^{(0)}) = a^r(z) + O(\|x\|)$, so $a(z, V^{(q)}) = a^r(z) + O(\|x\|)$. Lemma 5 in turns implies that $V^{(q+1)}(z) = V^r(z) + O(\|x\|^2)$.

Proof of Lemma 8. One proof is that it is just a straightforward corollary of (53). Here I show another proof, via a Taylor expansion of the default value function. For notational simplicity, in this proof I call r and r' the deviation of the interest rate from the average value, $\bar{r}$.

The Bellman equation is

$$V(w, r) = \max_c u(c) + \beta V((R+r)(w-c) + y', r') \quad (\text{E.4a})$$

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 $R = 1 + \bar{r}$,

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

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

$$\begin{aligned} V_r(w, r) &= \beta V'_{w'}(w', r') \frac{\partial w'}{\partial r} + \beta V'_{r'}(w', r') \frac{\partial r'}{\partial r} \\ V_r(w, r) &= \beta V'_{w'}(w', r') (w-c) + \beta V'_{r'}(w', r') \rho \end{aligned} \quad (\text{E.5})$$

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 , $D_w V = \partial_w V + \frac{da}{dw} \partial_a V$, 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 $c(w) = \bar{r}w/R + \bar{y}$, so $D_w c = \bar{r} : R$, and $D_w w' = d(R(w - c))/dw = R - R\bar{r}/R = 1$.

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

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 E.5. We obtain

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

so, using

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

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

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

so

$$\begin{aligned} \frac{dc}{dr} &= \frac{V_{wr}}{u''(c)} = \frac{-1}{u''(c)} \frac{V_w}{R} \frac{1 - \gamma \bar{r} (\frac{w}{c} - 1)}{R - \rho_r \beta} \\ &= \frac{-1}{\gamma u'(c) c} \frac{V_w}{R} \frac{1 - \gamma \bar{r} (\frac{w}{c} - 1)}{R - \rho_r} \\ \frac{dc}{c} &= \frac{1}{R} \frac{\bar{r} (\frac{w}{c} - 1) - 1/\gamma}{R - \rho_r} dr = b^r dr \\ b_r &= \frac{\bar{r} (w_t - c^d) - \frac{c^d}{\gamma}}{R(R - \rho_r)} \\ &= \frac{\bar{r} (w_t - \frac{rw_t + y}{R}) - \frac{c^d}{\gamma}}{R(R - \rho_r)} = \frac{\bar{r} (w_t - y) - \frac{c^d}{\gamma}}{R(R - \rho_r)} \end{aligned}$$

We note that the result

$$b_y = \frac{\bar{r}}{R(R - \rho_y)}, \quad b_r = \frac{\bar{r} (w_t - c^d) - \frac{c^d}{\gamma}}{R(R - \rho_r)}$$

becomes, in continuous time:

$$b_y = \frac{\bar{r}}{\bar{r} + \varphi_y}, \quad b_r = \frac{\bar{r} w_t - \psi c_t^d}{\bar{r} + \varphi_r} \quad (\text{E.6})$$

Proof of Proposition 11. When $\varphi > 0$, we saw that

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

Let $\psi := \frac{r + \varphi}{\xi} \neq 0$. Then

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

which is equivalent to

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

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

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

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

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

Appendix F: Extensions of the Basic Model

F.1. Evaluating the benefits of thinking at the true model

In the basic sparse max, benefits of thinking are evaluated at the default model. Here is a simple extension where they're evaluated at the true model. This is useful to avoid “starvation” in some extreme examples.

Call

$$\omega_i(m) := \mathbb{E} \left[-a'_{m_i}(m^d, x) u_{aa}(a(m^d, x), m, x) a_{m_i}(m^d, x) \right]$$

the prospective benefits of thinking, evaluated at model m . The basic sparse max sets attention to

$$m_i^* = \mathcal{A} \left(\frac{1}{\kappa} \omega_i(m^d) \right)$$

But one could enrich it e.g. as

$$m_i^* = \mathcal{A} \left(\frac{1}{\kappa} \max \left(\omega_i(m^d), \frac{\omega_i(\mu)}{K} \right) \right)$$

with $K > 1$. The max features two term: in the first one ($\omega_i(m^d)$), the benefits are evaluated at the default model; in the second term ($\omega_i(\mu)$) benefits are evaluated at the true model. To capture the fact that this is a more complex procedure, a penalty of $K > 1$ is applied, for some K , e.g. $K = 10$.

A benefit is that then the model “detects the danger of starvation”. In the 3-period model, we have

$$m_1^* = \mathcal{A} \left(\frac{\sigma_x^2}{2\kappa} \max \left\{ \left| u'' \left(\frac{w_1}{2} \right) \right|, \frac{1}{K} \left| u'' \left(\frac{w_1}{2} + x \right) \right| \right\} \right)$$

so that if $\frac{w_1}{2} + x$ is too close to a starvation level, then the second part is “active”, and attention becomes higher (if $u''' > 0$). For instance, if $\left| u'' \left(\frac{w_1}{2} + x \right) \right| = \infty$, then $m_1^* = 1$, and the consumer becomes fully attentive.

Likewise, we'll have

$$m_0^* = \mathcal{A} \left(\frac{\sigma_x^2}{6\kappa} \max \left\{ \left| u'' \left(\frac{w_0}{3} \right) \right|, \frac{1}{K} \left| u'' \left(\frac{\frac{2w_0}{3} + x}{2} \right) \right| \right\} \right)$$

as the value function is evaluated as a derivative of $V^{1,p}(w_1, x) = 2u \left(\frac{w_1 + x}{2} \right)$.

F.2. Finite-difference in the sensitivity to m

When we calculate $a_{m_i} = \frac{\partial a}{\partial m_i}$ in Definition 1, the following variant $\bar{\partial}_{m_i} a$ is sometimes useful. We first need to define the finite-difference operator:

$$\Delta_{m_i} g(m) := g(m)_{m_i=1, m_{-i}=0} - g(0) = g(0, \dots, 0, 1, 0, \dots, 0) - g(0)$$

where the 1 is at the i -th coordinate of m . This is the “finite difference” analogue to $\partial_{m_i} g(m) = \frac{\partial g}{\partial m_i} \big|_{m=0}$. Next, we define:

$$\begin{aligned}\bar{\partial}_{m_i} a(m, x) &:= \Delta_{m_i} (a_x \cdot x) \\ &= \Delta_{m_i} \left(\sum_k \left(\frac{\partial}{\partial x_k} a(m, x) \big|_{x=0} \right) x_k \right) \\ &= (\Delta_{m_i} \partial_x) (a) \cdot x\end{aligned}$$

Note that if $a(x, m) = \sum b_i m_i x_i$, then $\bar{\partial}_{m_i} a(m, x) = b_i m_i = \partial_{m_i} a(m, x)$. However, the definition using $\bar{\partial}_{m_i}$ generalizes better. For instance, if m is one-dimensional ($m = m_1$) and

$$a(m_1, x) = \sum_{i=1}^3 m_1^i b_i x_i$$

then $\bar{\partial}_{m_1} a(m, x) = \sum_{i=1}^3 b_i x_i$, whereas $\partial_{m_1} a(m, x) = \frac{\partial a(x)}{\partial m_1} \big|_{m=0} = b_1 x_1$. The higher-power terms m^2, m^3 are “invisible” when using ∂_{m_1} , but “visible” when using $\bar{\partial}_{m_1}$.

F.3. Taking into account the costs of thinking in the value function

One could take into account the costs of thinking in the value function. This will complicate the issues a bit and change the optimal action only by second order terms. Therefore, it’s best not to do that in the first model.

Should thinking costs be taken into account?

There are some reasons to do it. If attention is a resource, then its cost should be taken into account.

There are also reasons *not* to do it. First, we’re modeling a BR agent and imagining that the BR agent optimizing today will take into account future thinking costs may assume too much rationality. Technically when optimizing, the agent may take a default value of 0 for κ^V . Second, it could be that the costs in his decision utility (the ones used when deciding) are not the actual costs of thinking. This is the case if the agent misoptimizes on inattention, i.e. does as if the cost was κ – but perhaps the true cost is $\kappa^V = 10\kappa$ or $\kappa^V = 0.1\kappa$.

In the basic statement of the model, I opted for the simplest version of the framework. Here is an expanded version that does take them into account.

The selection of the action is still (27). Calling $m^*(Z, V^p)$ the attention return by the smax in (27), and the value function iteration in (29) becomes

$$\begin{aligned}\mathcal{T}(V, V^p)(Z) &:= u(a(Z, V^p), Z, \mu) - \kappa^V G(m^*(Z, V^p)) \\ &\quad + \beta \mathbb{E}[V(F(z, a(Z, V^p), \mu))]\end{aligned}\tag{F.1}$$

where $G(m) = \sum_i g(m_i)$, and κ^V is the perceived cost as included in the value function, and the state vector Z is expanded to include κ^V as a state variable.

The basic sparse max corresponds to taking $\kappa^V = 0$ at the Bellman iteration stage.

How do results change? First, the basic smax agent does not change his action at all. What changes is for the q -iterated agent case with $q \geq 1$. Second, in accordance of Section 3.5, this change makes only a second order difference in the action.

Formally, this can be interpreted by enriching the action space to $\bar{a} = (a, m)$ and having attention be part of the action vector and an expanded utility function:

$$\bar{u}(a, m, Z, \mu) = u(a, Z, \mu) - \kappa^V G(m)$$

Then, the perceived decision utility may or may not capture the correct value of κ . This is close to the perspective taken in Farhi and Gabaix (2020).

F.4. Iterated Static Sparse Max

In some cases, it is useful to have a generalization of the basic sparse max.

DEFINITION F.1. (*Iterated sparse max for static problems*) The K -times iterated sparse max, $\text{smax}_{a; m|m^d}^K u(a, x)$, is defined by the following procedure. Define $m^d(k)_{k=1}$ to be the initial default attention, m^d .

Start at round $k = 1$. At each round $k \leq K$, apply the regular smax, using the default $m^d(k)$: $\text{smax}_{a; m|m^d(k)} u(a, x, m)$, and call $m^*(k)$ and $a^*(k)$ the resulting attention. Define then $m^d(k+1) = m^*(k)$.

Stop at the end of round $k = K$, and return $m^*(K)$ and $a^*(K)$, the optimal attention and action at the last iteration.

Illustration. Suppose that $u(a, x) = -\frac{1}{2}(a - x_1(1 + x_2))^2$ so that the rational policy is $a^r(x_1, x_2) = x_1(1 + x_2)$. If the agent doesn't think of x_1 (replacing it with $x_1 = 0$), then he should not think about x_2 .

We next apply the iterated smax outlined in Definition F.1, iterating twice ($K = 2$). Initial default attention is $m^d(1) = (0, 0)$. We start at step $k = 1$. We observe that so $a_{x_1}^r = 1 + x_2$, $a_{x_2}^r = x_1$, which gives

$$m_1^*(1) = \mathcal{A}\left(\frac{\sigma_1^2}{\kappa}\right), \quad m_2^*(1) = 0$$

So, at the beginning of the second step, the default is $m^d(2) = m^*(1)$. Again applying the plain smax but with that default $m^d(2)$, we have

$$m_1^*(2) = \mathcal{A}\left(\frac{\sigma_1^2}{\kappa}\right), \quad m_2^*(2) = \mathcal{A}\left(\frac{m_1^*(1)^2 \sigma_2^2}{\kappa}\right)$$

Hence, the action is $a = a^r (m^* (2) \odot x) = m_1^* (2) x_1 (1 + m_2^* (2) x_2)$. We also see that, as $\kappa \rightarrow 0$, the action converges to the rational action.⁸⁹

F.5. Proportional Thinking

Here is a simple microfoundation for the scale-free κ of Section A.2.2, equation (F.2). Thinking about m_i implies some “mental costs” $g(m_i)$. These costs translate into some trembling in the action, so that, with $a^{**} = \arg \max_a v(a, x, m^*)$, the actual action is:

$$a = a^* + \tilde{\eta}$$

where $\tilde{\eta}$ is a mean 0 noise with standard deviation:

$$Stdev(\eta) = \sqrt{2\bar{\kappa}a^\natural} \left(\sum_i g(m_i) \right)^{1/2}$$

The size of the noise proportional to the typical scale $a^\natural$ of the action (this proportional is encountered in much of psychophysics, e.g. in the Weber-Fechner law)⁹⁰, and increasing in mental activity m . We call $\sqrt{2\bar{\kappa}}$ the factor of proportionality.

Hence, the utility losses from this noise $L = -\mathbb{E}[v(a^* + \eta, x) - v(a^*, x)]$ are, to the leading order:

$$L = -\frac{1}{2} \mathbb{E}[\eta^2] v_{aa} = -v_{aa} \bar{\kappa}^2 (a^\natural)^2 \sum_i g(m_i)$$

Hence, as in Gabaix (2014), (Lemma 2), the utility losses from imperfect attention are, to the leading order:

$$\frac{1}{2} \sum_{i,j=1\dots n} (1 - m_i) \Lambda_{ij} (1 - m_j) + \kappa \sum_i g(m_i)$$

with $\Lambda_{ij} := -\mathbb{E}[a_{m_i} V_{aa} a_{m_j}]$ and

$$\kappa = (\bar{\kappa} a^\natural)^2 |v_{aa}(a^d(x, m^d))|. \quad (\text{F.2})$$

89. This iterated smax suffices for the problems considered in this paper. For other purposes, one could imagine a variant where the default is at say $m^d = (\varepsilon, \dots, \varepsilon)$, for some $\varepsilon > 0$, so as to better “probe” the importance of all variables.

90. The microfoundation of that is probably that noisy computations are made in the brain in a scale-free space, and then at the end multiplied by $a^\natural$ to get an action to scale. This generates the proportionality to $a^\natural$. This type of thinking, however, it still speculative at this stage Glimcher (2011).

F.6. Simplification of Functions

F.6.1. Taylor expansion inside a value function. We develop here a bit of simple machinery to reflect how the agent can “simplify” a function (in practice a value function), by forcing them to have a given functional form.

A motivating example. Suppose that the agent consumes $c_1 = \frac{w}{2} + y_1$ and $c_2 = \frac{w}{2} + y_2$, where $y = (y_1, y_2)$ can be viewed as small. His rational value function, assuming no discounting, is

$$v(y) = u\left(\frac{w}{2} + y_1\right) + u\left(\frac{w}{2} + y_2\right)$$

The agent may wish to use a simplified representation of this function. We observe that $v(y) = v^S(y) + O(\|y\|^2)$ with

$$v^S(y) := 2u\left(\frac{w + y_1 + y_2}{2}\right)$$

We shall take this function V^S as a “simplified” representation of v . We can then form a more general function: $v(y, m^V) := (1 - m^V) v^S(y) + m^V v(y)$. If $m^V = 1$, then the agent uses the rational value function. If $m^V = 0$, then the agent uses the proxy value function v^S , which is in some sense simpler.

The following definition generalizes that thought and codifies the creation of a “simplified” value function.

DEFINITION F.2. (Simplifying function) Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function such that $f_{x_i}(x)|_{x=0} \neq 0$ for all i , and $\varphi : \{1, \dots, p\} \rightarrow \{1, \dots, n\}$. Call $\mathcal{E}^f := \{v \in C^1(\mathbb{R}^p, \mathbb{R}) \text{ such that } v(0) = f(0)\}$. We define the simplification function $S_{f,\varphi} : \mathcal{E}^f \rightarrow \mathcal{E}^f$ by

$$(S_{f,\varphi}(v))(y) := f(b \cdot y) \tag{F.3}$$

where b is the uniquely determined matrix $b \in \mathbb{R}^{n \times p}$ such $b_{ij} = 0$ unless $i = \varphi(j)$ and

$$v(y) = f(b \cdot y) + o(\|y\|) \tag{F.4}$$

Furthermore, $b_{ij} = \frac{v_{y_j}(y)|_{y=0}}{f_{x_i}(x)|_{x=0}}$ if $i = \varphi(j)$, $b_{ij}(x) = 0$ otherwise.

We prove the b is indeed unique.

Proof. We want $v(y) = f(b \cdot y) + o(\|y\|)$. This is equivalent to

$$v_{y_j}(y)|_{y=0} = \sum_i f_i b_{ij} = f_{\varphi(j)} b_{\varphi(j)j}.$$

Inspecting the Taylor expansions gives the result. $\square$

This defines an attention-augmented function

$$v(y, m^V) := (1 - m^V) v^S(y) + m^V v(y)$$

where parameter m^V captures the attention to the true value v .

Basically $f(b \cdot y)$ is like a non-linear Taylor expansion of $v(y)$. For instance, in our introductory example, $f(x) = 2u\left(\frac{w+x}{2}\right)$, $y = (y_1, y_2)$, $n = 1$, $p = 2$, $\varphi(j) = 1$, and $b = (1, 1)$.

Here are two other variants of the same idea. Suppose that we have a stochastic variable and a variant of the Black-Scholes model, with stochastic volatility. Then, we may approximate the value function in by tweaking the implied volatility: $V(x_t, S, K, r, t) = V^{\text{BS}}(\bar{\sigma} + ax_t + o(x_t), S, K, r, t)$, where V^{BS} is the regular Black-Scholes formula, so that the simplified function is

$$V^S(x_t, S, K, r, t) = V^{\text{BS}}(\bar{\sigma} + ax_t, S, K, r, t)$$

Suppose that the agent estimates a distribution, $h(y)$, where y are parameters of the distribution. The agent may wish to replace this distribution by a distribution with a simpler functional form, say a Gaussian: then f is a Gaussian distribution approximating the distribution h , perhaps by matching h 's mean and variance.

F.6.2. Just paying attention to first order terms. Suppose that the problem is:

$$\max_a u(a, x)$$

which gives $a = a^d + b \cdot x + O(\|x\|^2)$, with $b = -u_{aa}^{-1}u_{ax}$. Define $u^d(a) := u(a, 0)$, and $a^d = \arg \max_a u(a, 0)$. Suppose we have an agent that actually does:

$$\tilde{a} = a^d + b \cdot x$$

i.e. exactly discards the second order terms. How to we represent that agent?

First, we could define a ‘‘Taylor sparse max’’, that given a problem $u(a, x)$, returns the linearized optimum $\tilde{a}$, or a sparse version of it, $\tilde{a} = a^d + \sum_i b_i m_i x_i$

Second, we can say that the agent uses a proxy utility function. We observe that for (a, x) close to $(a^d, 0)$, we have:

$$\begin{aligned} u(a, x) &= u^i(a, x) + O(x^2) \\ u^1(a, x) &= u^d(a - bx) + u_x(a^d, 0)x \\ u^2(a, x) &= u^d(a^d, x) + \frac{1}{2}u_{aa}^d(a - a^d)^2 + (a - a^d)u_{ax}x \\ u^3(a, x) &= u^d(a - bx) + u(a^d, x) - u^d(a^d) \end{aligned}$$

The above representations u^i all ‘‘work’’, i.e. deliver the linear expansion.

F.6.3. Linearizing a relation. Suppose that there’s a non-linearity, say it’s $\hat{r}_t = f^r(x)$, e.g. $f^r(x) = \frac{1}{2}x + bx^3$, then the agent may use a linearized policy, i.e.,

$$f(x, m) = f(0) + m_1 f'(0)x + m_2 (f(x) - f(0) - f'(0)x)$$

This is, the function is approximated by its constant, first order term (with weight m_1), and higher order terms (with weights m_2). This way, one has a “simpler” representation by linearization.

F.7. Notes on the design of the model

Here I record some notes about modeling choices of the model. This section should only interest people thinking about the foundations of the approach (hence, potential ways to change it), not its direct practical use.

F.7.1. Breaking the explosion of Thinking about thinking about thinking... Why not model iterated expectations, such as the agent’s perception at time 0 of his perception at time 2 of his perception at time 5? The short answer is that this leads to a combinatorial explosion of the complexity of the model. This motivates the particular formulation of sparse dynamic programming, which eschews such a combinatorial explosion.

I record this phenomenon in this subsection, using the simple 3-period model of Section 4.1, extended here to $T + 1$ periods. There we obtain about 2^T state variables for a T -period model.

Utility is $\sum_{t=0}^T u(c_t)$ and $R = 1$. The agent receives w_0 at time 0, and x at time t . So, $w_t = w_{t-1} - c_{t-1} + x1_{t=T}$ for $t \geq 1$. The rational problem is

$$\max \sum_{t=0}^T u(c_t) \text{ s.t. } \sum_{t=0}^T c_t = w_0 + x$$

and the optimal consumption policy at time t is $c_t = \frac{w_t + x}{T - t + 1}$

$$V^T(w_T, x) = u(w_T + x)$$

Next, at time $T - 1$, the rational value function is (RE stands for rational expectations):

$$\begin{aligned} V^{T-1, RE}(w_{T-1}, x, M_{T-1}^{RE}) &= u\left(\frac{w_{T-1} + m_{T-1}^* x}{2}\right) \\ &\quad + u\left(\frac{w_{T-1} + (2 - m_{T-1}^*) x}{2}\right) \end{aligned}$$

for the value m_{T-1} chosen at time $T - 1$. However, the perceived value function could be

$$V^{T-1, RE}(w_{T-1}, x, M_{T-1})$$

for some other perceived M_{T-1} .

At time $T - 2$, the problem is

$$V^{T-2}(w_{T-2}, x, M^{RE}) = u(c_{T-2}) + V^{T-1}(w_{T-2} - c, x, M_{T-1}^{RE}),$$

with

$$c_{T-2} = \arg \max_c u(c) + V^{T-1}(w_{T-2} - c, m_{T-2}^x, E_{T-2}[M_{T-1}])$$

so that at each stage, the agent gets either “ M^{RE} ” or “ M^S ”. So, the relevant perception vector is $M_{T-2} = (m_{T-2}, M_{T-1}, E_{T-2}^s[M_{T-1}]) \in \mathbb{R}^3$ – the datum of the perception of x , M_{T-1} and the actual M_{T-1} .

More generally, at time t , the value function is

$$V_t(w_t, x, M_t) = u(c_t^*) + V_{t+1}(w_t - c_t^*, x, M_{t+1})$$

with

$$c_t^* = \arg \max_{c_t} u(c_t) + V_{t+1}(w_t - c_t, m_t^x, E_t^s[M_{t+1}])$$

So, the perception vector at time t is

$$M_t = (m_t^x, M_{t+1}, E_t^s[M_{t+1}])$$

i.e. it’s formed of the attention m_t^x to x , tomorrow’s perception vector M_{t+1} , and today’s perception about tomorrow’s vector, $E_t^s[M_{t+1}]$. Ideally, the agent should keep track of all those. Call $D_t = \dim M_t$. We have $D_T = 0$, $D_{T-1} = 1$, $D_t = 1 + 2D_{t+1}$. This yields that the dimension of the attention vector at time t is

$$D_t = 2^{T-t} - 1.$$

which is overwhelming.

Simplified value function. This is why the main sparse max cuts through the difficulty to allow for just two value functions: $V^{t,RE}(w_t, x)$, the rational expectation value function, and $V^{t,S}(w_t, x)$, the “simplified value function”. Then, we form

$$V^t(w_t, x, m_t^V) = m_t^V V^{t,RE}(w_t, x) + (1 - m_t^V) V^{t,S}(w_t, x)$$

LEMMA F.1. *The value function $V^t(w, x, M)$ is independent of M , up to $O(x^2)$ terms.*

Hence, we define the germ $V^t(w, x, M^S)$ to be a “simplified” version of $V^t(w, x, M)$. In many cases, it’s the $V^t(w, x, M^S) = V^t(w, x, M^r)$, assuming that the agent will be rational. Typical case, $u_c = V_w(w, x, M)$, with $c = w(A + Bx) + C + Dx + o(x)$. (This applies with time-varying interest rate, income and equity premium). Then, we set $V(w, x, M^S) := u(w(A + Bx) + C + Dx) / (A + Bx)$. If chosen representation is $c = we^{A+Bx} + o(x)$, then we set: $V(w, x, M^S) := u(we^{A+Bx}) / e^{A+Bx}$.

More abstractly, if it’s $c = \varphi(A(w) + B(w)x) + o(x)$, for some function φ , then we set $V_w(w, x, M^S) = u_c(\varphi(A(w) + B(w)x))$.

F.8. Generalization: The k, K Procedure

Here I discuss how to do an expansion when V_{xx} is required but not known by the agent. This greatly generalizes the Cass-Koopmans of Section 5.

Suppose that we want to solve

$$V(X_0) = \max_{(a_t)_{t \geq 0}} \sum_{t \geq 0} \beta^t u(X_t, a_t) \text{ s.t. } X_{t+1} = F(X_t, a_t) \quad (\text{F.5})$$

but do not know $V_{XX}(X_0)$, which is required to find a_{X_0} (recall that the first order condition is $u_a + \beta V_X F_a$, so calculating a_{X_0} involves V_{XX}). What to do?

I posit the following description of the agent's world view and behavior. He considers x_t "his" variable, and X_t the value created by the environment, which is perceived to be exogenous to him: this is the same way that in much of macro, k_t is his wealth, and K_t is the aggregate capital stock. He has a mental model of the law of motion of X_t on the equilibrium path, e.g. as in $X_{t+1} - X_t = B(X_t - X_*)$ for some matrix B , and X_* is the steady state value of X_t . He solves in his mind the "micro" problem:

$$V(x_0, X_0) = \max_{(a_t)_{t \geq 0}} \sum_t \beta^t \bar{u}(x_t, X_t, a_t) \text{ s.t. } x_t = \bar{F}^x(x_t, X_t, a_t) \quad (\text{F.6})$$

with X_t exogenous to his actions and I define the modified utility and production function, in a manner that separates the micro variable x and the macro variable X :

$$\begin{aligned} \bar{u}(x, X, a) &:= u(X, a) + u_X(X, a)(x - X), \\ \bar{F}^x(x, X, a) &:= F(X, a) + F_X(X, a)(x - X) \end{aligned}$$

This way, they capture the marginal contribution of x .^{91,92}

Next, I assume that the agent knows V, V_x, V_{xx} evaluated at $(x, X) = (X_*, X_*)$ — i.e., the agent has solved the "microeconomic" problem of optimizing when the macroeconomic environment is at the steady state, i.e. $X_t = X_*$ at all dates, but his microeconomic variable x_t is off equilibrium.⁹³

91. One could also imagine using A , the other agent's action, and setting $\bar{F}(x, X, a, A) := F(X, A) + F_X(X, A)(x - X) + F_a(X, A)(a - A)$, or similar variants. What's important is that values and first order derivatives are preserved around $(x, a) = (X, A)$: more precisely, function $\bar{F}(x, X, a, A)$ must satisfy: $\bar{F}(x + \varepsilon, x, a + \eta, a) = F(x + \varepsilon, a + \eta) + O(\varepsilon^2 + \eta^2)$, and similarly $\bar{u}(x, X, a, A)$ must satisfy: $\bar{u}(x + \varepsilon, x, a + \eta, a) = u(x + \varepsilon, a + \eta) + O(\varepsilon^2 + \eta^2)$.

92. For instance, in the Cass-Koopmans problem, $(x, X, a) := (k, K, c)$, $F(K, c) = f(K) - c$, and

$$\bar{F}^k(k, K, c) = f(K) - c + f'(K)(k - K) = f'(K)k + [f(K) - Kf'(K)] - c$$

so that $f'(K)k$ is the return on capital, and $[f(K) - Kf'(K)]$ is labor income.

93. This is actually easy to derive in a number of canonical problems: For instance, in RCK this is saying that the agent knows the "micro" problem of the life-cycle with constant interest rates. This is also true in a canonical real investment problem derived below.

This gives the value $a_x(x, X_*)$. This way, we can calculate $V_{x,X}$ using the procedure in Section A.1 (using the change in notations $(w, x) := (x, X)$).

Hence the action is $\hat{a} = a_x \hat{x}_t + a_X \hat{X}_t$ and on the equilibrium path $\hat{x}_t = \hat{X}_t$. To calculate a_X we proceed as in section A.1.⁹⁴ Section C.4 applies this to an optimal real investment problem.

F.9. Tools to expand a simple model: 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 following gives a continuous-time version of Section A.1.

The laws of motion are

$$\begin{aligned}\dot{w}_t &= F^w(w, x, a) \\ \dot{x}_t &= F^x(w, x)\end{aligned}$$

and the Bellman equation is

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

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

Call, for some function f , $D_w f = \partial_w f + a_w \partial_a f$ the “total impact” of a change in w . Then differentiate the Bellman equation with respect to x ,

$$\rho V_x = u_x + V_w F_x^w + V_x F_x^x + V_{xx} F^x \quad (\text{F.7})$$

Now, we differentiate with respect to w and evaluate at $x = 0$:

$$\rho V_{wx} = D_w(u_x + V_w F_x^w) + V_{wx} F_x^x + V_x F_{wx}^x$$

so

$$V_x = (\rho - F_x^x)^{-1} [u_x + V_w F_x^w] \quad (\text{F.8})$$

$$V_{wx} = (\rho - F_x^x)^{-1} [D_w(u_x + V_w F_x^w) + V_x F_{wx}^x] \quad (\text{F.9})$$

94. The FOC is $\Psi(x, X, a) = 0$ with

$$\Psi(x, X, a) := \bar{u}_a(x, X, a) + \beta V_x(x, X) \bar{F}_a^x(x, X, a)$$

This gives the marginal impact of X on the action $a_X = -\Psi_a^{-1} \Psi_X$, with

$$\begin{aligned}\Psi_a &= \bar{u}_{aa} + \beta V_x \bar{F}_{aa}^x = u_{aa} + F_{aa} \\ \Psi_X &= \bar{u}_{aX} + \beta V_{xX} \bar{F}_a^x + \beta V_x \bar{F}_{aX}^x = \bar{u}_{aX} + \rho V_{xX} \bar{F}_a^x.\end{aligned}$$

where in the last two equations, the last part of the right-hand side is evaluated at $x = X$.

As a satisfies $\Psi = 0$ with

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

where we have used here $F_a^x = 0$. Hence, the impact of x on the optimal action is

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

$$\Psi_a = u_{aa} + V_w F_{aa}^w$$

$$\Psi_x = u_{ax} + V_w F_{ax}^w + V_{wx} F_a^w$$

The terms $u_{ax} + V_w F_{ax}^w$ are the “instantaneous impact” of a change x . The term $V_{wx} F_a^w$ depends on the present value of future x , given that V_{wx} depends on its present value (as shown by the term $(\rho - F_x^x)^{-1}$).

F.9.1. Losses from suboptimal policies. Calculation of V_{xx} . We now turn to the more difficult case of V_{xx} . Using $D_x f = \partial_x f + a_x \partial_a f$ the “total impact” of a change in x , we have:

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

Next, differentiating at $x = 0$,

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

hence

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

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

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

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

We use the following elementary Lemma:

LEMMA F.2. *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,

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

so, using the previous Lemma,

$$V_{xx}^s - V_{xx}^r = (\rho - 2F_x^x)^{-1} (a_x^s - a_x^r) (u_{aa} + V_w F_{aa}^w) (a_x^s - a_x^r) \quad (\text{F.10})$$

We gather the results.

PROPOSITION F.1. (What are the losses from a suboptimal policy?) *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*

$$V^\delta = 0, V_w^\delta = 0, V_{ww}^\delta = 0, V_x^\delta = 0, V_{wx}^\delta = 0 \quad (\text{F.11})$$

and

$$V_{xx}^\delta = (\rho - 2F_x^x)^{-1} (a_x^s - a_x^r) (u_{aa} + V_w F_{aa}^w) (a_x^s - a_x^r) \quad (\text{F.12})$$

Equation (F.12) 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 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 $\Lambda_t = -\frac{1}{2} x_t a_x^\delta v_{aa} a_x^\delta x_t$. The full utility loss is the present discounted value of that, i.e.

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

It is enough to study the “static” utility losses to derive the dynamic utility losses. This proposition F.1 is a dynamic application of the Proposition 26 in Gabaix (2014) (online appendix) regarding losses from a suboptimal policy. For convenience, we restate this proposition here. With static problem $\max u(a, x)$ s.t. $b(a, x) \geq 0$ and a Lagrangian $L(a, x) = u(a, x) + \lambda b(a, x)$, the losses from a suboptimal policy $a^\delta = a - a^r$ (where a^r is the optimal policy) are to the leading order: $\frac{1}{2} a^{\delta'} L_{aa} a^\delta$.

Here the Lagrangian is $L = \int e^{-\rho t} [u(a_t, z_t) + \lambda_t (-\dot{z}_t + F^z(a_t, z_t))] dt$, where $z_t = (w_t, x_t)$ is the state vector. Hence, the loss Λ is expressed by (to the

leading order)

$$2\Lambda = a' L_{aa} a = \int a_t^\delta L_{a_t a_t} a_t^\delta = \int e^{-\rho t} a_t^\delta [u_{a_t a_t} + \lambda_t F_{a_t a_t}] a_t^\delta dt$$

Suppose that we can linearize, $a_t^\delta = A x_t$, we have

$$2\Lambda = \int e^{-\rho t} x_t' A' [u_{a_t a_t} + \lambda_t F_{a_t a_t}] A x_t dt$$

Consider the ergodic limit, where x_t has a distribution independent of t . Recall that

$$\mathbb{E} x_t' B x_t = \mathbb{E} \sum_{i,j} x_i B_{ij} x_j = \sum_{i,j} B_{ij} \mathbb{E} [x_i x_j] = \text{Trace} (B \mathbb{E} [x x'])$$

Hence,

$$2\Lambda = \frac{1}{\rho} \text{Trace} (B \mathbb{E} [x x'])$$

$$B = A' [u_{a_t a_t} + \lambda_t F_{a_t a_t}] A = A' L_{a_t a_t} A$$

F.10. Some ancillary results

Call $G(Z, m) = Z(m)$ a transformation function for the state vector Z . E.g. in the basic life-cycle example, $G(w, \hat{y}, \hat{r}, m) = (w, m_y \hat{y}, m_r \hat{r})$. [Note: below, the notation bar isn't ideal, as bar refers to means; perhaps tilde would be better] When can we express the perceived model as a rational model, with different utility and transition functions? The following Lemma gives the answer.

LEMMA F.3. *Let $G(Z, m)$ be a function and define $\bar{Z}_t = G(Z_t, m)$. Suppose that we can write*

$$u(a, Z, m) = \bar{u}(a, G(Z, m))$$

$$G(F(a, Z, m), m) = \bar{F}(a, G(Z, m))$$

for two functions $\bar{u}, \bar{F}$. Then the model evaluated at m is the same as a rational model with state variables $\bar{Z}_t$, utility $\bar{u}$, transition function $\bar{F}$. We also have $V(Z) = \bar{V}(G(Z, m))$

Proof. We have

$$\bar{Z}_{t+1} = G(Z_{t+1}, m) = G(F(a, Z_t, m), m) = \bar{F}(a, G(Z_t, m)) = \bar{F}(a, \bar{Z}_t)$$

The value function $\bar{V}$ satisfies, with Z s.t. $\bar{Z} = G(Z, m)$

$$\bar{V}(\bar{Z}) = \max_a \bar{u}(a, \bar{Z}) + \beta \bar{V}(\bar{F}(a, \bar{Z}))$$

$$= \max_a u(a, Z, m) + \beta \bar{V}(G(F(a, Z, m), m))$$

Define $V(Z) := \bar{V}(G(Z, m))$. Then,

$$V(Z) = \bar{V}(\bar{Z}) = \max_a u(a, Z, m) + \beta V(F(a, Z, m), m)$$

So indeed V satisfies the Bellman equation. $\square$

Also, we have $F^w(Z, m) = ((R + \bar{r})(w - \bar{y}), \rho)$

$$\hat{y}_{t+1} = m_F \rho_y \hat{y}_t + \varepsilon_{t+1}^y$$

gives

$$m_y \hat{y}_{t+1} = m_{y'} \rho_y (m_y \hat{y}_t + m_y \varepsilon_{t+1}^y)$$

i.e.

$$\bar{F}^{\bar{y}}(a, \bar{Z}) = m_{y'} \rho_y (\bar{y}_t + m_y \varepsilon_{t+1}^y)$$

I conclude with a remark which will be useful later, drawing again on Gabaix (2014). As κ has the units of utils, one can make it more endogenous with the primitive, unitless parameter $\bar{\kappa}$, by setting:

$$\kappa = \bar{\kappa}^2 \text{var}(u(a^d(x), x, m^d))^{1/2} \quad (\text{F.13})$$

Appendix G: Other Results

G.1. Second Order Losses From Inattention

G.1.1. Statement. The intuition is that as x is small, and the action is close to the optimum, we get only second order losses $O(\|x\|^2)$ from misoptimization. The idea is simple, but it turns out that it requires some extra care, in particular to ensure that differentiability, and to define formally the objects of interest. I present that here. This subsection does not contain surprising results, so should be skipped at the first reading.

Recall that w is a set of variables thought about in the default model ($m = 0$), and x is a set of variable not thought about in the default model. We also set $z = (w, x)$. Formally, we assume that $u^t(a, w, x, m)$ and $F^t(a, w, x, m)$ are independent of m if $x = 0$. We also assume that $F^{x,t}(a, w, x, m) = 0$ if $x = 0$, i.e. so that small x 's at t map in small x 's at $t + 1$.

We suppose that the attention function $\mathcal{A}$ is C^∞ (this can easily be ensured).

Recapitulations of the notations. With 1 period problems, the action is

$$a(z, m) := \arg \max_a u(a, z, m) \quad (\text{G.1})$$

With exogenous attention m , the value function is the utility evaluated at the true model of the world:

$$V(z, m, \mu) := u(a(z, m), z, \mu) \quad (\text{G.2})$$

where μ denotes the true state of the world (typically $\mu = (1, \dots, 1)$). With endogenous attention, we have

$$m_i^*(w, \sigma_x^2, \mu) := \arg \max_{m_i \in [0, \mu_i]} E \left[\frac{1}{2} V_{m_i m_i}(w, x, m) \Big|_{m=0} (m_i - \mu_i)^2 + G(m_i) \right] \quad (\text{G.3})$$

and the value function with endogenous attention is

$$V(z, \mu) := V(z, m^*(w, \sigma_x^2, \mu), \mu) \quad (\text{G.4})$$

We sometimes drop the explicit dependence on μ .

With T period problems, the notions are the same, but recursive. We use $m = (m_t)_{t=0 \dots T}$. The action at time t is:

$$a^t(z, m) := \arg \max_a u^t(a, z, m_t) + \beta V^{t+1}(F^t(a, z, m_t), m_{t+1}) \quad (\text{G.5})$$

which defines the value function, evaluated at the true model of the world. With exogenous attention:

$$V^t(z, m, \mu) := u^t(a^t(z, m), z, \mu_t) + \beta V^{t+1}(F^t(a, z, m_t), \mu_{t+1}) \quad (\text{G.6})$$

When attention is endogenous, we have

$$m_i^{t,*}(w, \sigma_x^2, \mu) := \arg \max_{m_{t,i} \in [0, \mu_{t,i}]} E \left[\frac{1}{2} V_{m_i m_i}^t(w, x, m) \Big|_{m_t=0} (m_{t,i} - \mu_{t,i})^2 + G(m_{t,i}) \right]$$

and

$$V^t(z, \mu) := V^t(z, m^{t,*}(w, \sigma_x^2, \mu), \mu). \quad (\text{G.7})$$

We call $V^r(w, x) := V^t(w, x, (\mu_t, \mu_t)_{t=0 \dots T})$ the rational value function.

We prove two propositions: the first one is elementary to state, the other one a bit more cumbersome.

PROPOSITION G.1. (Second order losses from inattention) *Suppose that u^t and F^t are C^∞ . Recall the decomposition $z = (w, x)$. Then, with exogenous m , $a^t(z, (m_\tau, \mu_\tau)_{\tau=t \dots T})$ and $V^t(z, (m_\tau, \mu_\tau)_{\tau=t \dots T})$ are C^∞ and,*

$$\begin{aligned} & V^t(w, x, (m_\tau, \mu_\tau)_{\tau=t \dots T}) - V^{t,r}(w, x) \\ &= O(\|x\|^2) O\left(\sum_{\tau=t}^T \|m_\tau - \mu_\tau\|^2\right) \end{aligned} \quad (\text{G.8})$$

With endogenous m , $a^t(z)$ and $V^t(z)$ are C^∞ , and

$$V^t(w, x) - V^{t,r}(w, x) = O(\|x\|^2) \quad (\text{G.9})$$

PROPOSITION G.2. (Second order losses from inattention, with finite differentiability) *With exogenous m , assume that there is an $\ell \geq 3$ such that u^t is $C^{t+\ell}$, F^t (the transition function from t to $t+1$) is $C^{t+\ell}$. Then $a^t(z, (m_\tau, \mu_\tau)_{\tau=t \dots T})$ and $V^t(z, (m_\tau, \mu_\tau)_{\tau=t \dots T})$ are $C^{t+\ell-1}$ for all $t = 0, \dots, T$, and (G.8) holds.*

With endogenous m , assume that there is an $\ell \geq 5$ such that u^t ($t \leq T$), F^t (the transition from t and $t+1$) are $C^{3t+\ell}$ for some ℓ . Then $a^t(z)$ and $V^t(z)$ are $C^{3t+\ell-3}$ and (G.9) holds.

G.1.2. *Proof of Propositions G.1 and G.2.* The proof starts with the most elementary cases (1 period), then amplifies it to 2 and more periods.

1 period, with exogenous m . We suppose that $u(a, w, x, m)$ is C^k , with $k \geq 4$, and that the rational problem is

$$\max_a u(a, w, x, m)$$

We shall see that $V(w, x, m)$ is C^{k-1} and $V(w, x)$ is C^{k-3} .

First, let us suppress the dependence on w , and consider $\max_a u(a, x, m)$.

By the implicit function theorem, $a(x, m) := \arg \max_a u(a, x, m)$ is C^{k-1} (it is the solution of $u_a(a, x, m) = 0$). This makes $V(x, m, \mu) := u(a(x, m), x, \mu)$ be C^{k-1} .

As $u(a, 0, m)$ is independent of m , $a(0, m)$ is independent of m , so we can write: $a(x, m) = a(0, 0) + b(m)x + O(x^2)$. As $a_x(x, m)$ is C^{k-2} with $k-2 \geq 1$ there is a constant L such for that all m, μ in the compact $[0, 1]^n$, we have $\|b(m) - b(\mu)\| \leq L \|m - \mu\|$. We consider the loss from imperfect perception:

$$\begin{aligned} R(x, m, \mu) &:= V(x, m, \mu) - V(x, \mu, \mu) \\ &= u(a(x, m), x, \mu) - u(a(x, \mu), x, \mu) \\ &= (a(x, m) - a(x, \mu))' u_{aa}(a(x, m) - a(x, \mu)) \\ &\quad + o(a(x, m) - a(x, \mu))^2 \\ &= (b(m) - b(\mu))' x u_{aa} x (b(m) - b(\mu))' + o(\|x\|^2) \end{aligned}$$

So we have

$$V(x, m, \mu) - V(x, \mu, \mu) = R(x, m, \mu) = O(\|x\|^2) O(\|m - \mu\|^2) \quad (\text{G.10})$$

Reinserting the dependence in w , the same reasoning show that with

$$R(w, x, m, \mu) := V(w, x, m, \mu) - V(w, x, \mu, \mu) \quad (\text{G.11})$$

we have

$$R(w, x, m, \mu) = O(\|x\|^2) O(\|m - \mu\|^2). \quad (\text{G.12})$$

1 period, with endogenous m . We supposed that $u(a, w, x, m)$ was C^k , which led $V(w, x, m)$ to be C^{k-1} . Now, we endogenize m . We have

$$m_i^*(w, \sigma_x^2) = \arg \max_m E \left[-\frac{1}{2} V_{m_i m_i}(w, x, m) \Big|_{m=0} (m_i - \mu_i)^2 - G(m_i) \right]$$

So, $m^*(w)$ is a C^{k-3} function of w . So $V(w, x) := V(w, x, m^*(w))$ is a C^{k-3} function.

Hence, starting from a C^k function $u(a, w, x, m)$, we obtain a C^{k-3} value function $V(w, x)$. We “lost” 3 orders of differentiability.⁹⁵

2 period problems. There are 2 periods, 0 and 1. We call

$$\begin{aligned} V^1(w, x, m, \mu) &:= u^1(a(w, x, m), w, x, \mu) \\ V^1(w, x, \mu) &:= u^1(a(w, x, m^*(w, \sigma_x^2)), w, x, \mu) \end{aligned}$$

the value function at the beginning of period 1, with respectively exogenous and endogenous attention.

We assume that u^1 is C^k .

2 periods, with exogenous m_1 . The problem is now

$$\max_a v^0(a, w, x, m_0, m_1, \mu_1)$$

with

$$v^0(a, x, m_0, m_1, \mu_1) := u^0(a, x, m_0) + V^1(F(a, w, x, m_0), m_1, \mu_1)$$

where the last function $V^1(w, x, m_1)$, which is C^{k-1} , and $F = (F^w, F^x)$ gives the transition functions for both w and x . We assume that F, V^1, u^0 are C^{k-1} . So, function v^0 is C^{k-1} .

The reasoning in the 1 period case applies, and $a^0(w, x, m_0, m_1, \mu_1)$ is C^{k-2} , so

$$V^0(x, m_0, \mu_0, m_1, \mu_1) := v^0(a^0(x, m_0, m_1, \mu_1), x, \mu_0, m_1, \mu_1)$$

is C^{k-2} . If $k \geq 3$, we have

$$V^0(x, m_0, \mu_0, m_1, \mu_1) - V^0(x, \mu_0, \mu_0, m_1, \mu_1) = O(\|x\|^2) O(\|m_0 - \mu_0\|^2)$$

by the time-0 result.

Also, we have

$$V^0(x, \mu_0, \mu_0, m_1, \mu_1) - V^0(x, \mu_0, \mu_0, \mu_1, \mu_1) = O(\|x\|^2) O(\|m_1 - \mu_1\|^2)$$

so summing the two differences:

$$\begin{aligned} &V^0(x, m_0, \mu_0, m_1, \mu_1) - V^0(x, \mu_0, \mu_0, \mu_1, \mu_1) \\ &= O(\|x\|^2) O(\|m_0 - \mu_0\|^2 + \|m_1 - \mu_1\|^2) \end{aligned} \tag{G.13}$$

95. If there was no w , then we'd just have m^* a value independent of x , and function $V(w, x)$ would be C^{k-1} .

2 periods, with endogenous m_0 . The problem is now

$$\max_a v^0(a, w, x, m_0)$$

with

$$v^0(a, w, x, m_0) := u^0(a, w, x, m_0) + V^1(F(a, w, x, m_0))$$

where the last function is $V^1(w, x)$, which is C^{k-3} , and $F = (F^w, F^x)$ gives the transition functions for both w and x . We assume that F, V^1, u^0 are C^{k-3} . So, function $v^0(a, x, m_0)$ is C^{k-3} .

By the reasoning before with 1 period, $a(w, x, m_0)$ is C^{k-4} and

$$V^0(a, w, x, m_0) = v^0(a(w, x, m_0), w, x, m_0)$$

is C^{k-4} . Next, to endogenize m_0 , again by the reasoning done with 1 period, $m_0(w, \sigma_x^2)$ is C^{k-6} , so that

$$V^0(a, w, x) = V^0(a, w, x, m_0^*(w, \sigma_x^2)) \quad (\text{G.14})$$

is C^{k-6} .

We next move to more than 2 periods. The reasoning is very similar to the 2 period case.

$T + 1$ periods, exogenous m . We assumed that u^t is $C^{t+\ell}$, F^t (from t to $t + 1$) is $C^{t+\ell}$ for some $\ell \geq 3$. That implies that a^T is $C^{T+\ell-1}$, V^T is $C^{T+\ell-1}$, and by backward induction on $t = T \dots 0$, that a^t and V^t are $C^{t+\ell-1}$ for all $t = 0, \dots, T$. So, if $\ell \geq 3$, then V^0 is C^2 , and

$$\begin{aligned} & V^0(w, x, (m_t, \mu_t)_{t=0 \dots T}) - V^0(w, x, (\mu_t, m_t)_{t=0 \dots T}) \\ &= O(\|x\|^2) O\left(\sum_{t=0}^T \|m_t - \mu_t\|^2\right) \end{aligned} \quad (\text{G.15})$$

holds as well.

$T + 1$ periods, endogenous m . We assumed that u^t ($t \leq T$), F^t (the transition from t and $t + 1$) is $C^{3t+\ell}$ for some ℓ . Then, by backwards induction on $t = T \dots 0$, V^t is $C^{3t+\ell-3}$. Indeed, by the reasoning done in the 2 period case, given that u^t is C^k , we have that V^t is C^{k-3} . As F^{t-1}, u^{t-1} are assumed to be C^{k-3} , we have that V^{t-1} is C^{k-6} .

This ends the proof of Propositions G.1 and G.2. $\square$

G.1.3. Extensions.

Noise. In the problems above, there is no noise. Adding noise is straightforward, but adds yet another layer of notations. Formally, we assume bounded noise:

$$\|\varepsilon_{t+1}^x\| \leq K \sigma_x \text{ almost surely}$$

for some K , and where $\sigma_x = E \left[\|x\|^2 \right]^{1/2}$. Then, the statements in Propositions G.1 and G.2 are replaced by

$$\begin{aligned} V^t(w, x, (m_\tau, \mu_\tau)_{\tau=t \dots T}) - V^{t,r}(w, x) \\ = O \left(\|x\|^2 + \sigma_x^2 \right) O \left(\sum_{\tau=t}^T \|m_\tau - \mu_\tau\|^2 \right) \end{aligned} \quad (\text{G.16})$$

and

$$V^t(w, x) - V^{t,r}(w, x) = O \left(\|x\|^2 + \sigma_x^2 \right) \quad (\text{G.17})$$

This is, both the actual value and the variance of x matter.

References for Online Appendix

- Abel, Andrew B., Janice C. Eberly, and Stavros Panageas (2013). “Optimal inattention to the stock market with information costs and transactions costs.” *Econometrica*, 81(4), 1455–1481.
- Acemoglu, Daron (2008). *Introduction to modern economic growth*. Princeton university press.
- Acemoglu, Daron, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2012). “The network origins of aggregate fluctuations.” *Econometrica*, 80(5), 1977–2016.
- Acemoglu, Daron and Martin Kaae Jensen (2023). “Equilibrium Analysis in Behavioural One-Sector Growth Models.” *Review of Economic Studies*.
- Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll (2022). “Income and wealth distribution in macroeconomics: A continuous-time approach.” *The Review of Economic Studies*, 89(1), 45–86.
- Aguiar, Mark and Erik Hurst (2013). “Deconstructing Life Cycle Expenditure.” *Journal of Political Economy*, 121(3), 437–492.
- Akerlof, George A. and Janet L. Yellen (1985). “Can small deviations from rationality make significant differences to economic equilibria?” *The American Economic Review*, 75(4), 708–720.
- Allcott, Hunt and Dmitry Taubinsky (2015). “Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market.” *American Economic Review*, 105(8), 2501–2538.
- Angeletos, George-Marios and Zhen Huo (2021). “Myopia and anchoring.” *American Economic Review*, 111(4), 1166–1200.
- Angeletos, George-Marios and Chen Lian (2018). “Forward guidance without common knowledge.” *American Economic Review*, 108(9), 2477–2512.
- Auclert, Adrien (2019). “Monetary policy and the redistribution channel.” *American Economic Review*, 109(6), 2333–2367.

- Bansal, Ravi and Amir Yaron (2004). "Risks for the long run: A potential resolution of asset pricing puzzles." *The Journal of Finance*, 59(4), 1481–1509.
- Barberis, Nicholas (2000). "Investing for the long run when returns are predictable." *The Journal of Finance*, 55(1), 225–264.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998). "A model of investor sentiment." *Journal of financial economics*, 49(3), 307–343.
- Barro, Robert J (1974). "Are Government Bonds Net Wealth?" *Journal of Political Economy*, 82(6), 1095–1117.
- Barro, Robert J (1999). "Ramsey meets Laibson in the neoclassical growth model." *The Quarterly Journal of Economics*, 114(4), 1125–1152.
- Barro, Robert J (2009). "Rare disasters, asset prices, and welfare costs." *American Economic Review*, 99(1), 243–264.
- Basak, Suleyman and Georgy Chabakauri (2010). "Dynamic mean-variance asset allocation." *The Review of Financial Studies*, 23(8), 2970–3016.
- Becker, Gary S and Kevin M Murphy (1988). "A theory of rational addiction." *Journal of political Economy*, 96(4), 675–700.
- Bernheim, B Douglas, Jonathan Skinner, and Steven Weinberg (2001). "What accounts for the variation in retirement wealth among US households?" *American Economic Review*, 91(4), 832–857.
- Beshears, John, James J Choi, Joshua Hurwitz, David Laibson, and Brigitte C Madrian (2015). "Liquidity in retirement savings systems: An international comparison." *American Economic Review*, 105(5), 420–425.
- Bigio, Saki and Jennifer La'o (2020). "Distortions in production networks." *The Quarterly Journal of Economics*, 135(4), 2187–2253.
- Bilbiie, Florin O (2021). "Monetary Policy and Heterogeneity: An Analytical Framework." *CEPR Discussion Paper No. 12601*.
- Blanchard, Olivier and Stanley Fisher (1989). *Lectures on Macroeconomics*. The MIT Press, Cambridge, Mass.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013). "Salience and Consumer Choice." *Journal of Political Economy*, 121(5), 803–843.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2018). "Diagnostic expectations and credit cycles." *The Journal of Finance*, 73(1), 199–227.
- Brown, Jennifer, Tanjim Hossain, and John Morgan (2010). "Shrouded attributes and information suppression: Evidence from the field." *The Quarterly Journal of Economics*, 125(2), 859–876.
- Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein (2021). "A model of relative thinking." *The Review of Economic Studies*, 88(1), 162–191.
- Caballero, Ricardo J. (1995). "Near-Rationality, Heterogeneity, and Aggregate Consumption." *Journal of Money, Credit and Banking*, 27(1), 29–48.
- Caballero, Ricardo J and Alp Simsek (2013). "Fire sales in a model of complexity." *The Journal of Finance*, 68(6), 2549–2587.

- Campbell, John Y. and N. Gregory Mankiw (1989). "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence." *NBER Macroeconomics Annual*, 4, 185–216.
- Campbell, John Y and Luis M Viceira (2002). *Strategic asset allocation: portfolio choice for long-term investors*. Clarendon Lectures in Economic.
- Candes, Emmanuel J. and Terence Tao (2006). "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, 52(12), 5406–5425.
- Carroll, Christopher D. and Andrew A. Samwick (1997). "The nature of precautionary wealth." *Journal of Monetary Economics*, 40(1), 41–71.
- Carvalho, Vasco M. and Basile Grassi (2019). "Large Firm Dynamics and the Business Cycle." *American Economic Review*, 109(4), 1375–1425.
- Chetty, Raj (2012). "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica*, 80(3), 969–1018.
- Chetty, Raj, Adam Looney, and Kory Kroft (2009). "Salience and taxation: Theory and evidence." *American economic review*, 99(4), 1145–1177.
- Cochrane, John H. (1989). "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives." *The American Economic Review*, 79(3), 319–337.
- Del Negro, Marco, Marc P. Giannoni, and Christina Patterson (2023). "The Forward Guidance Puzzle." *Journal of Political Economy Macroeconomics*, 1(1), 43–79.
- Evans, George W and Seppo Honkapohja (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.
- Evans, George W. and Seppo Honkapohja (2013). *Rethinking Expectations : The Way Forward for Macroeconomics*, chap. 2. Learning as a Rational Foundation for Macroeconomics and Finance, pp. 68–111. Princeton University Press.
- Farhi, Emmanuel and Xavier Gabaix (2020). "Optimal Taxation with Behavioral Agents." *American Economic Review*, 110(1), 298–336.
- Farhi, Emmanuel and Iván Werning (2019). "Monetary Policy, Bounded Rationality, and Incomplete Markets." *American Economic Review*, 109(11), 3887–3928.
- Fuster, Andreas, Benjamin Hebert, and David Laibson (2012). "Natural expectations, macroeconomic dynamics, and asset pricing." *NBER Macroeconomics Annual*, 26(1), 1–48.
- Fuster, Andreas, David Laibson, and Brock Mendel (2010). "Natural expectations and macroeconomic fluctuations." *The Journal of Economic Perspectives*, 24(4), 67–84.
- Gabaix, Xavier (2009). "Linearity-Generating Processes: A Modelling Tool Yielding Closed Forms for Asset Prices." *Working paper*.
- Gabaix, Xavier (2011). "The granular origins of aggregate fluctuations." *Econometrica*, 79(3), 733–772.

- 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 (2014). “A sparsity-based model of bounded rationality.” *Quarterly Journal of Economics*, 129(4), 1661–1710.
- Gabaix, Xavier (2019). “Behavioral Inattention.” *Handbook of Behavioral Economics*, 2, 261–344.
- Gabaix, Xavier (2020). “A behavioral New Keynesian model.” *American Economic Review*, 110(8), 2271–2327.
- Gabaix, Xavier and Ralph SJ Koijen (2022). “In search of the origins of financial fluctuations: The inelastic markets hypothesis.” *NBER Working Paper No. 28967*.
- 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(2), 505–540.
- Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006). “Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model.” *American Economic Review*, 96(4), 1043–1068.
- Ganong, Peter and Pascal Noel (2019). “Consumer spending during unemployment: Positive and normative implications.” *American economic review*, 109(7), 2383–2424.
- Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus (2021). “Five facts about beliefs and portfolios.” *American Economic Review*, 111(5), 1481–1522.
- Glimcher, Paul W. (2011). *Foundations of Neuroeconomic Analysis*. Oxford University Press.
- Gourinchas, Pierre-Olivier and Jonathan A Parker (2002). “Consumption over the life cycle.” *Econometrica*, 70(1), 47–89.
- Greenwood, Robin and Samuel G Hanson (2015). “Waves in ship prices and investment.” *The Quarterly Journal of Economics*, 130(1), 55–109.
- Gruber, Jonathan and Botond Köszegi (2001). “Is Addiction Rational? Theory and Evidence.” *Quarterly Journal of Economics*, 116(4), 1261–1303.
- Hall, Robert E (1988). “Intertemporal substitution in consumption.” *Journal of political economy*, 96(2), 339–357.
- Hansen, Lars Peter and Thomas J Sargent (2008). *Robustness*. Princeton university press.
- Harris, Christopher and David Laibson (2001). “Dynamic choices of hyperbolic consumers.” *Econometrica*, 69(4), 935–957.
- Harris, Christopher and David Laibson (2013). “Instantaneous gratification.” *The Quarterly Journal of Economics*, 128(1), 205–248.
- Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh (2009). “Driven to distraction: Extraneous events and underreaction to earnings news.” *The*

- Journal of Finance*, 64(5), 2289–2325.
- Hong, Harrison, Jeremy C Stein, and Jialin Yu (2007). “Simple forecasts and paradigm shifts.” *The Journal of Finance*, 62(3), 1207–1242.
- Hsieh, Chang-Tai (2003). “Do consumers react to anticipated income changes? Evidence from the Alaska permanent fund.” *American Economic Review*, 93(1), 397–405.
- Judd, Kenneth L (1998). *Numerical methods in economics*. MIT press.
- Jung, Junehyuk, Jeong Ho Kim, Filip Matějka, and Christopher A Sims (2019). “Discrete actions in information-constrained decision problems.” *The Review of Economic Studies*, 86(6), 2643–2667.
- Kahneman, Daniel (2011). *Thinking, fast and slow*. Macmillan.
- Kaplan, Greg, Benjamin Moll, and Giovanni L Violante (2018). “Monetary policy according to HANK.” *American Economic Review*, 108(3), 697–743.
- Khaw, Mel Win, Ziang Li, and Michael Woodford (2021). “Cognitive imprecision and small-stakes risk aversion.” *The Review of Economic Studies*, 88(4), 1979–2013.
- Kőszegi, Botond and Adam Szeidl (2013). “A Model of Focusing in Economic Choice.” *Quarterly Journal of Economics*, 128(1), 53–104.
- Krusell, Per and Anthony A Smith, Jr (1998). “Income and wealth heterogeneity in the macroeconomy.” *Journal of political Economy*, 106(5), 867–896.
- Kueng, Lorenz (2018). “Excess sensitivity of high-income consumers.” *The Quarterly Journal of Economics*, 133(4), 1693–1751.
- Laibson, David (1997). “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics*, 112(2), 443–478.
- Lian, Chen (2021). “A theory of narrow thinking.” *Review of Economic Studies*, 88(5), 2344–2374.
- Ljungqvist, Lars and Thomas J. Sargent (2012). *Recursive Macroeconomic Theory*. The MIT Press, URL <http://www.jstor.org/stable/j.ctt5vjq05>.
- Lucas Jr, Robert E (1976). “Econometric policy evaluation: A critique.” In *Carnegie-Rochester conference series on public policy*, vol. 1, pp. 19–46, North-Holland.
- Lusardi, Annamaria and Olivia S Mitchell (2011). “Financial literacy and retirement planning in the United States.” *Journal of pension economics & finance*, 10(4), 509–525.
- Lusardi, Annamaria, Daniel Schneider, Peter Tufano, Adair Morse, and Karn M. Pence (2011). “Financially Fragile Households: Evidence and Implications.” *Brookings Papers on Economic Activity*, pp. 83–150.
- Maćkowiak, Bartosz and Mirko Wiederholt (2015). “Business Cycle Dynamics Under Rational Inattention.” *Review of Economic Studies*, 82(4), 1502–1532.
- Malamud, Semyon and Grigory Vilkov (2018). “Non-myopic betas.” *Journal of Financial Economics*, 129(2), 357–381.
- Mankiw, N. Gregory and Ricardo Reis (2002). “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips curve.”

- Quarterly Journal of Economics*, 117(4), 1295–1328.
- McKay, Alisdair, Emi Nakamura, and Jón Steinsson (2016). “The Power of Forward Guidance Revisited.” *American Economic Review*, 106(10), 3133–3158.
- Meggiorini, Greta (2023). “Behavioral New Keynesian Models: An empirical assessment.” *Journal of Macroeconomics*, 77, 103538.
- Merton, Robert C (1971). “Optimum consumption and portfolio rules in a continuous-time model.” *Journal of Economic Theory*, 3(4), 373–413.
- Modigliani, Franco and Richard H. Brumberg (1954). “Utility analysis and the consumption function: an interpretation of cross-section data.” In *Post-Keynesian Economics*, edited by Kenneth K. Kurihara, pp. 388–436. Rutgers University Press, New Brunswick, NJ.
- O’Donoghue, Ted and Matthew Rabin (2001). “Choice and procrastination.” *The Quarterly Journal of Economics*, 116(1), 121–160.
- Prescott, Edward C. and Rajnish Mehra (1980). “Recursive Competitive Equilibrium: The Case of Homogeneous Households.” *Econometrica*, 48(6), 1365–1379.
- Rabin, Matthew (2013). “An approach to incorporating psychology into economics.” *American Economic Review*, 103(3), 617–622.
- Rees-Jones, Alex and Dmitry Taubinsky (2020). “Measuring ‘schmeduling’.” *The Review of Economic Studies*, 87(5), 2399–2438.
- Reis, Ricardo (2006). “Inattentive consumers.” *Journal of Monetary Economics*, 53(8), 1761–1800.
- Romer, David (2018). *Advanced Macroeconomics*. New York: McGraw Hill Companies.
- Samuelson, William and Richard Zeckhauser (1988). “Status quo bias in decision making.” *Journal of risk and uncertainty*, 1, 7–59.
- Sargent, Thomas J et al. (1993). *Bounded rationality in macroeconomics*. Oxford University Press.
- Schwartzstein, Joshua (2014). “Selective attention and learning.” *Journal of the European Economic Association*, 12(6), 1423–1452.
- Sims, Christopher A (2003). “Implications of rational inattention.” *Journal of Monetary Economics*, 50(3), 665–690.
- Smets, Frank and Rafael Wouters (2007). “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review*, 97(3), 586–606.
- Thaler, Richard (1985). “Mental Accounting and Consumer Choice.” *Marketing Science*, 4(3), 199–214.
- Tibshirani, Robert (1996). “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.
- Tversky, Amos and Daniel Kahneman (1981). “The framing of decisions and the psychology of choice.” *science*, 211(4481), 453–458.

- Van Nieuwerburgh, Stijn and Laura Veldkamp (2010). “Information acquisition and under-diversification.” *The Review of Economic Studies*, 77(2), 779–805.
- Veldkamp, Laura L. (2011). *Information Choice in Macroeconomics and Finance*. Princeton University Press.
- Wachter, Jessica A and Motohiro Yogo (2010). “Why do household portfolio shares rise in wealth?” *The Review of Financial Studies*, 23(11), 3929–3965.
- Wang, Chong, Neng Wang, and Jinqiang Yang (2016). “Optimal consumption and savings with stochastic income and recursive utility.” *Journal of Economic Theory*, 165, 292–331.
- Woodford, Michael (2013). “Macroeconomic Analysis without the Rational Expectations Hypothesis.” *Annual Reviews of Economics*, 5, 303–346.
- Woodford, Michael (2019). “Monetary policy analysis when planning horizons are finite.” *NBER Macroeconomics Annual*, 33(1), 1–50.
- Woodford, Michael (2020). “Modeling imprecision in perception, valuation, and choice.” *Annual Review of Economics*, 12, 579–601.