A Sparsity-Based Model of Bounded Rationality, Applied to Basic Consumer and Equilibrium Theory*

Xavier Gabaix
NYU Stern, CEPR and NBER
September 26, 2012

Abstract
This paper defines and analyzes a “sparse max” operator, which generalizes the traditional max operator used everywhere in economics. The agent builds (as economists do) a simplified model of the world which is sparse, considering only the variables of first-order importance. His stylized model and his resulting choices both derive from constrained optimization. Still, the sparse max operator remains tractable to compute. Moreover, the induced outcomes reflect basic psychological forces governing limited attention.

The sparse max provides a behavioral extension to several parts of economics. Here we study the basic theory of consumer demand (choosing a consumption bundle subject to a budget constraint) and competitive equilibrium. We obtain a behavioral version of classic pillars of economics, such as Marshallian and Hicksian demand, Slutsky matrix and Edgeworth boxes, and competitive equilibrium. The Slutsky matrix is not symmetric anymore – non-salient prices are associated with anomalously small demand elasticities. In the Edgeworth box, the offer curve is “extra-dimensional.” In the behavioral model it is a two-dimensional surface rather than a one-dimensional curve. This leads to a robust one-dimensional continuum of competitive equilibria that corresponds to levels of economic activity or exchange. An equilibrium is efficient if and only if agents have the same misperceptions. We can see which parts of basic microeconomic theory are robust, and which are not, to the assumption of perfect maximization.

---

*xgabaix@stern.nyu.edu. I thank David Laibson for a great many enlightening conversations about bounded rationality over the years. For very good research assistance I am grateful to Jonathan Libgober and Farzad Saïdi, and for helpful comments to Andrew Abel, Daniel Benjamin, Douglas Bernheim, Pierre-André Chiappori, Vincent Crawford, Stefano Dellavigna, Alex Edmans, David Hirshleifer, Botond Koszegi, Sendhil Mullainathan, Matthew Rabin, Antonio Rangel, Larry Samuelson, Yuliy Sannikov, Thomas Sargent, Andrei Shleifer, and participants at various seminars and conferences. I am grateful to INET for financial support.
1 Introduction

This paper proposes a tractable model of some dimensions of bounded rationality (BR). It develops a “sparse maximum” operator, which is a behavioral generalization of the traditional “maximum” operator. In the sparse max, the agent pays less or no attention to some features of the problem. The sparse max has a psychological foundation and is quite versatile. It easily handles problems of maximization with constraints.

This research has two goals: one is to provide an easy-to-use, plausible alternative to full maximization. Another is to probe which parts of economic theory are robust, and which are not, to that assumption of perfect maximization. Given the application-oriented focus of this paper (where theory is meant to shed light on other, central, theories), I submit a proof of concept by re-examining two pillars of economic theory: basic consumer theory (problem 1), and basic equilibrium theory. We obtain a simple way to give a fresh perspective on those topics: a behavioral version of Marshallian and Hicksian demand, Edgeworth boxes etc., and competitive equilibrium sets. We can see which parts of basic microeconomics are fragile to the assumptions of perfect maximization, and which are robust. We also obtain an enrichment of perfect-rationality basic microeconomics, that is arguably more plausible, and often equally convenient (or more, as BR agents tend to use simpler problems).

The principles behind the sparse max are the following. First, the decision maker (DM) in the model builds a simplified model of the world, somewhat like economists do. He builds a representation of the world that is simple enough, and thinks about the world through his partial model. Second, this representation is “sparse,” i.e., uses few parameters that are non-zero or different from the usual state of affairs.1 These choices are controlled by an optimization of his representation of the world, that depends on the problem at hand. I draw from fairly recent literature on statistics and image processing to use a notion of “sparsity” that still leads to well-behaved, convex maximization problems (see Tibshirani (1996), Candès and Tao (2006), Donoho (2006), Mallat (2009)). The idea is to think of “sparsity” (having lots of zeroes in a vector) instead of “simplicity” (which is an amorphous notion), and measure the lack of “sparsity” by the sum of absolute values. This apparently simple step leads to a rich set of results in statistics and signal processing – largely because of the tractable (convex) notion of “simplicity” it leads to (through linear penalties, rather than fixed costs).2

---

1 The meaning of “sparse” is that of a sparse vector or matrix. For instance, a vector in $\theta \in \mathbb{R}^{100,000}$ with only a few non-zero elements is sparse.

2 Econometricians have already successfully used sparsity (e.g. Belloni and Chernozhukov 2010).
paper follows this lead to use sparsity notions in economic modelling, and to the best of my knowledge is the first to do so.

“Sparsity” is also a psychologically realistic feature of life. For any decision, in principle, thousands of considerations are relevant to the DM; his income, but also GDP growth in his country, the interest rate, recent progress in the construction of plastics, interest rates in Hungary, the state of the Amazonian forest, etc. Since it would be too burdensome to take all of these variables into account, he is going to discard most of them. The traditional modelling for this is to postulate a fixed cost for each variable. However, that often leads to intractable problems (fixed costs, with their non-convexity, are notoriously ill-behaved). In contrast, the notion of sparsity I use (again, following the statistics literature mentioned above, using linear rather than fixed costs) leads to problems that are easy to solve.

The model rests on very robust psychological notions, which are discussed below. It incorporates limited attention, of course. To supply the “missing elements” due to limited attention, people rely on defaults. When taking into account some information, they anchor on the default and do a limited adjustment towards the truth. This “anchoring and adjustment” is at the heart of the model. At the same time, attention is allocated purposefully, towards features that are likely to be important. Sparsity is a way to capture this allocation of attention.

If the agent is confused about prices, how is the budget constraint still satisfied? I propose a way to incorporate maximization under constraint. For this purpose I tried to strike a good balance between psychological plausibility and tractability. The formulation of the sparse max with constraints has some nice properties, e.g. of duality, which are quite useful.

After the sparse max has been defined, I probe its usefulness by re-examining two building blocks of basic microeconomics: consumer theory and competitive equilibrium theory. By basic consumer theory, I mean the optimal choice of a consumption bundle subject to a budget constraint:

\[
\max_{c_1, \ldots, c_n} u(c_1, \ldots, c_n) \text{ subject to } p_1 c_1 + \ldots + p_n c_n \leq w
\]  

where \(u\) is a utility function, \(c_i\) is the quantity consumed of good \(i\), \(p_i\) the price, and \(w\) the available budget. There does not yet appear to be any systematic treatments of this building block with a non-standard model other than sparsity.

---

3 Ignoring variables altogether and assuming that they do not differ from their usual values are the same thing in the model. For instance, in most decisions we do not pay attention to the quantity of oxygen that is available to us because there is plenty of it. In the model, ignoring the oxygen factor is modeled as assuming that the quantity of oxygen available is the normal quantity. Indeed, the two are arguably the same thing.

4 In companion projects, I examine how to inject sparsity into game theory, and dynamic programming.

5 Dufwenberg et al. (2011) analyze general equilibrium with other-regarding preferences, especially when
One might think that there is little to add to that oldest of topics. However, it turns out that (sparsity-based) limited rationality leads to enrichments that may be both realistic and intellectually intriguing. I assume that agents do not fully pay attention to all prices. The price they perceive is a weighted average of the true price and the default price. The sparse max determines how much attention they pay to each good, and how they adjust their budget constraint. The comparative statics are sensible. People pay more attention to goods that have more volatile prices, and to which they are fundamentally more price elastic. The agent exhibits a form of nominal illusion. If all prices and budget increase by 10%, say, the consumer does not react in the traditional model; however, in the sparsity model, the DM under-perceives the increase of non-salient prices; hence, his demand changes. This apparently simple departure has many consequences.

The model allows us to work out the sparse counterpart of the key building blocks of consumer theory: Marshallian and Hicksian demand, Slutsky matrix, Edgeworth boxes, Roy’s identity, Shepard’s lemma, etc. We can see which results are robust under BR, and which are not (see section 1). One result is that the Slutsky matrix is not symmetric any more. Non-salient prices will lead to small terms in the matrix. At the same time, the model offers a parsimonious deviation from the rational model. I argue below that indeed, the extant evidence seems to favor the effects theorized here. In addition, deviations from Slutsky symmetry could useful for potential future empirical work: they allow to recover quantitatively the extant of limited attention (up to a multiplicative factor).

I next revisit equilibrium theory and the venerable Edgeworth box. Recall that the “offer curve” of an agent is the set of consumption bundles he chooses as prices change (those price changes affecting also dollar value of his endowment). In the traditional Edgeworth box analysis (with two goods), this offer curve is, well, a curve: a one-dimensional object. However, in the sparsity model, it becomes a two-dimensional object (see Figure 2). We obtain an “extra-dimensional” offer curve. This is due to the fact that, when the prices of the two goods change, in the traditional model only their ratio matters. However, in the present model, both prices matter, not just their ratio, and we have a two-dimensional curve. This effect, though quite basic, appears to be new.

Next, we study equilibrium theory with those sparse agents. When two-dimensional offer curves intersect, we do not expect a unique equilibrium. The key finding is that robustly, imply as-if rational behavior. In problem (1), their preferences are rational (though they depend on other people’s actions).

Chetty, Looney and Kroft (2009) show how this matters empirically. Their model cannot handle most the issues dealt in this paper, particularly in section 4-5. The “fixed cost” model they use does not lend itself to marginal analysis.

This notion is very different from the idea of a “thick demand curve”, which basically means that the consumer is indifferent between dominated bundles: a sparse consumer has only thin demand curves.
in the sparse model, there is one-dimensional continuum of competitive equilibria, rather than a (locally) unique equilibrium. The intuitive, rough reason is that, high prices, even when coming with a high budget, have a real effect on their demand (by the nominal illusion mentioned above), so the equilibrium is changed. A more refined reason is the following: in the traditional model, there is a one-dimensional continuum set of prices supporting the equilibrium (basically some price vector and then scalar multiples of it). This is still qualitatively true in the sparse model. However, these different prices correspond, in the sparse model, to different real allocations. Hence, the sparsity model generates a nontrivial one-dimensional set of equilibria. These equilibria can be interpreted as different levels of “economic activity” (or “exchange” in an exchange economy). We obtain effects of the Keynesian macroeconomic style (e.g., people supply more labor when the perceived wage is higher), in a basic microeconomic context.

I gather what appears to be robust and not robust in the basic microeconomic theory of consumer behavior and competitive equilibrium – when the specific deviation is a sparsity-seeking agent.\(^8\)

**What is robust in basic microeconomics?**

**Propositions that are not robust**

*Tradition:* The Slutsky matrix is symmetric. *Sparse model:* The Slutsky matrix is asymmetric, as elasticities to non-salient prices become small.

*Tradition:* The Marshallian demand \(c^a(p, w)\) is homogeneous of degree 0, i.e. there is no money illusion. *Sparse model:* Lack of attention leads to some nominal illusion.

*Tradition:* The offer curve is one-dimensional in the Edgeworth box. *Sparse model:* It is typically two-dimensional.

*Tradition:* The equilibrium set is made of isolated points. *Sparse model:* There is a one-dimensional continuum of equilibria.

*Tradition:* A competitive equilibrium is Pareto-efficient. *Sparse model:* An equilibrium is efficient if and only if agents have the same misperceptions.

*Tradition:* The Slutsky matrix is the second derivative of the expenditure function. *Sparse model:* They are linked, but in a richer way modulated by price salience.

*Tradition:* The Slutsky matrix is negative semi-definite. *Sparse model:* That property generally fails.

**Small Robustness: Propositions that hold at the default price, but not away from it, to the first order**

Marshallian and Hickisian demands, Shepard’s lemma and Roy’s identity: the values of

\(^8\)I use the sparsity benchmark not as “the truth,” of course, but as a plausible benchmark for a less than fully rational agent. The paper contains those statements, and the underlying technical conditions.
the underlying objects are the same in the traditional and sparse model at the default price, but differ (to the first order in $p - p^d$) away from the default price.

**Greater robustness: Objects are very close around the default price, up to second order terms**

Expenditure function $e(p, v)$, indirect utility function $V(p, w)$: their values are the same, under the traditional and sparse models, up to second order terms in the price deviation from the default ($p - p^d$).\(^9\)

*Traditional* economics gets the signs right — or, more prudently put, the signs predicted by the rational model remain correct under a sparsity variant. Those predictions are of the type “if the price of good 1 does down, demand for it goes up”, or, more generally “if there’s a good incentive to do X, people will indeed tend to do X”\(^10\) Those sign predictions make intuitive sense, and, not coincidentally, they hold in the sparse model: they hold even when the agent has a limited, qualitative understanding of his situation. Indeed, when economists think about the world, or in much applied microeconomic work, it is often the sign predictions that are used and trusted, rather than the detailed quantitative predictions.

After this behavioral version of basic micro, the paper concludes with a discussion of other approaches to inattention / behavioral economics. The main theme is that this approach is more tractable than other approaches that have been proposed, hence it allows us to analyze new, more complex things; in particular, consumer theory and basic equilibrium theory. Take the “noise and signal extraction” approach (surveyed in Velkamp 2011). Even though it has proven very useful in a wide range of economic situations, those situations are pretty much all Gaussian / quadratic. Outside the Gaussian / quadratic setting, this approach is immensely complex. It seems that we are very far from being able to analyze problem (1) with the traditional “noise and signal extraction” because they are not linear-quadratic. Likewise, models with automata, or many fixed costs, or discrete categories, are so complex to use that they also cannot, at least for now, address this problem (they also lead to NP-complete problems). The sparse approach, in its simplicity, can analyze them.

The plan of the paper is as follows. Section 2 states the basic sparse max, analyzes it, and discusses its psychological underpinnings. Section 3 extends the sparse max so that it can handle constraints, such as budget constraints. Section 4 develops the “pure theory of consumer’s behavior”. Section 5 extends the analysis to Edgeworth boxes and competitive

---

\(^9\)This is just a consequence of the envelope’s theorem.

\(^10\)This is true for “direct” effects, though not necessarily once indirect effects are taken into account. For instance, this is true for compensated demand (see the part on the Slutsky matrix), and in partial equilibrium. This is not necessarily true for uncompensated demand (where income effects arise) or in general equilibrium – though in many situations those “second round” effects are small.
equilibrium. Those two sections are the richest in fairly concrete results. Section 6 discusses links with existing themes in behavioral and information economics. Section 7 presents concluding remarks. Many proofs are delegated to the appendix or the online appendix. The online appendix contains extensions and other applications, e.g. basic producer theory and optimal pricing with sparse agents.

2 The Basic Sparse Max Operator

2.1 The Sparse Max Operator without Constraints

I state the sparse max operator, which is a sparse version of the usual max operator. The agent faces a maximization problem which is, in its traditional version, \( \max_a u(a, \mu, x) \) subject to \( B(a, \mu, x) \geq 0 \), where \( u \) is a utility function, and \( B \) is a constraint (or a vector of constraints). I want to define the “sparse max” operator:

\[
\text{smax}_a u(a, \mu, x) \text{ subject to } B(a, \mu, x) \geq 0,
\]

(2)
a sparse version of the “max” operator. We first shall examine cases without constraints, i.e. study \( \max_a u(a, \mu, x) \). To fix ideas, take the following quadratic example:

\[
u(a, \mu, x) = -(a - \sum_{i=1}^{n} \mu_i x_i)^2.
\]

(3)

Then, the traditional optimal action is \( a(\mu) = \sum_{i=1}^{n} \mu_i x_i \). For instance, to choose consumption \( a \) (normalized from some baseline), the decision maker should consider not only his wealth, \( x_1 \), and the deviation of GDP from its trend, \( x_2 \), but also the interest rate, \( x_{10} \), demographic trends in China, \( x_{100} \), recent discoveries in the supply of copper, \( x_{200} \), etc. There are \( n > 10,000 \) (say) factors that should in principle be taken into account. However, most of them have a small impact on his decision, i.e., their impact \( \mu_i \) is small in absolute value.

Notations are as follows: \( a \) is the action; it is potentially multi-dimensional. \( \mu \) is an ideal parameter (typically a vector). However, the agent might pick a sparser parameter \( m \) (we will specify how) and maximize \( u(a, m, x) \) rather than \( u(a, \mu, x) \). Then, he will pick the action: \( a(m) = \sum_{i=1}^{n} m_i x_i \), for some vector \( m \) that endogenously has lots of zeros, i.e., \( m \) is “sparse.” For instance, if the agent only pays attention to his wage and the state of the economy, \( m_1 \) and \( m_2 \) will be non-zero, and the other \( m_i \)’s will be zero.

I now present a procedure that the agent might follow. I first describe it, then analyze its
properties, then justify it (justification is easier after the properties are clear). The inputs are:

- \( m^d \), the “default parameter” (typically \( m^d = 0 \)). This corresponds psychologically to what the agent considers if he has “no time” to think about it. Conceptually, it’s the flash of intuition corresponding to the System 1 in Kahneman’s system 1. In the quadratic example, it will be \( m^d = (0,\ldots,0) \), i.e. the agent thinks about nothing.

- A parameter \( \kappa \geq 0 \), the penalty for lack of sparsity. If \( \kappa = 0 \), the agent will be the traditional, rational agent model. If \( \kappa > 0 \), the agent has a real taste for sparsity.

- A probability measure \( P \) over the set of \( x \): the DM behaves as if the \( x \) were drawn from \( P \). We assume that \( \mathbb{E}_P [x_i] = 0 \). Two cases are worth distinguishing:

  1. **Attention chosen before seeing the variables (Ex ante attention).** The DM chooses the weight \( m_i \) before seeing the \( x_i \). As a proxy for the likely magnitude of \( x_i \), he uses \( \sigma_{x_i} \), the standard deviation of \( x_i \) under \( P \),

  2. **Attention chosen after seeing the variables (Ex post attention).** The DM chooses the weight \( m_i \) after seeing the \( x_i \). Formally, this is a special case of the former, where the agent knows the magnitude of the variables (but not the sign), and hence sets \( \sigma_{x_i} = |x_i| \).

- Finally, a random variable \( \eta_a \), a “modulus” for the discernible changes in action \( a \). That will ensure that the model is invariant in the units in which \( a \) is measured. In practice, only the product \( \kappa \eta_a \) matters, but it’s useful to think of \( \kappa \) and \( \eta_a \) separately. This way, \( \kappa \) is unitless and the case \( \kappa = 0 \) corresponds exactly to the traditional, rational, non-sparse model.

In a static context, those objects are just exogenous. However, in a dynamic models, which have more structure, they become more endogenous (Gabaix 2012a). For instance, the “default model” may simply be the one that assumes that all variables are at their average value. The random variable \( \eta_a \) might represent the normal variability of the actions. At this stage, though, we shall just take them as exogenous. Still, much of the economics will come from the limited attention, not from playing with \( m^d \) and \( \eta_a \). In that sense, those parameters need to be specified, but do not matter too much.

We define \( a^d := \arg \max_a u (a, m^d, x) \), and will call it the default action. It’s the optimal action under the default model. We assume that function \( u (a, m, x) \) is strictly concave in \( a \), and twice continuously differentiable in \( (a, m) \) in a neighborhood of \( (a^d, m^d) \).
Indeed, as are de a simpler representation of it, namely a quadratic loss around the default. The utility loss from using the approximate model maximand, \( a^* \), written \( u^* = \text{smax}_{a|m^d,k,\eta,a,P} u(a, \mu, x) \) and \( a^* = \text{arg max}_{a|m^d,k,\eta,a,P} u(a, \mu, x) \), are defined by the following procedure: First, calculate the optimally sparse representation of the world, \( m^* \):

\[
m^* = \arg \min_m \frac{1}{2} (\mu - m)^T \mathbb{E}_P [u'_{am} (-u_{aa}^{-1}) u_{am}] (\mu - m) + \kappa \sum_i |m_i - m_i^d| \mathbb{E}_P [(u_{ma_i} \eta_a)]^2 \]

(4)

Second, define the sparse action \( a^* \): \( a^* = \text{arg max}_a u(a, m^*, x) \). The resulting utility \( u^* = u(a^*, \mu, x) \). In (4) derivatives are evaluated at the default model and action: \( m^d, a^d := \text{arg max}_a u(a, m^d, x) \).

In other terms, the agent solves for the optimal \( m^* \) that trades off a proxy for the utility losses (the first term in the right-hand side of equation 4) and a psychological penalty for deviations from a sparse model (the second term on the left-hand side of 4). Then, the agent maximizes over the action \( a \), taking \( m^* \) to be the true model.

When \( \kappa = 0 \), the sparse max is simply the regular max. Hence, the model continuously includes the traditional model with no cognitive friction.

The expression \( L^{\text{quad}}(m) = \frac{1}{2} (\mu - m)^T \mathbb{E}_P [u'_{am} (-u_{aa}^{-1}) u_{am}] (\mu - m) \) is the quadratic approximation of the expected loss as a result of an imperfect model \( m \). 11 One modeling decision in defining the sparse max is to use \( L^{\text{quad}}(m) \) rather than the exact loss, which would be very complex to use for both the decision maker and the economist.12 The decision maker uses a simplified representation of the loss from inattention. This is one way to circumvent Simon’s “infinite regress problem” — that optimizing the allocation of thinking cost can be even more complex than the original problem. I avoid that problem by assuming a simpler representation of it, namely a quadratic loss around the default.

Let us now analyze the model further.

2.2 Model Analysis

Let us first analyze the problem \( \min_m \frac{1}{2} (m - \mu)^2 + \kappa |m| \), which is the elementary component of problem (4). There, \( m \) wants to be close to the ideal weight \( \mu \), but there is a linear penalty

11 Indeed, consider a function \( u \) with no \( x \), and \( a(m) = \text{arg max}_a u(a, m) \), the best action under model \( m \). The utility loss from using the approximate model \( m \) rather than the true model \( \mu \) is \( L(m) := u(a(\mu), \mu) - u(a(m), \mu) \). A Taylor expansion shows that for \( m \) close to \( \mu \), \( L(m) = L^{\text{quad}}(m) \) to the leading order. Indeed, as \( a(m) \) solves \( V_a(a(m), m) = 0 \), the implicit function theorem gives \( V_{aa} \delta a + V_{am} \delta m = 0 \), i.e., \( \delta a = -V_{aa}^{-1} V_{am} \delta m \) with \( \delta m = m - \mu \). Hence, the loss is, up to \( o(\|\delta a\|^2) \) terms, \( L = -V_a \delta a - \frac{1}{2} \delta a V_{aa} \delta a = 0 + L^{\text{quad}}(m) \).

12 Tibshirani (1996, section 8) also recommends using a quadratic approximation in statistics.
for deviation from $0$, $\kappa |m|$. Its solution (formalized in Lemma 1 in the appendix) is:

$$m = \tau(\mu, \kappa)$$  \hspace{1cm} (5)

for the "anchoring and adjustment" function $\tau$ plotted in Figure 1 and defined as:

$$\tau(\mu, \kappa) = (|\mu| - |\kappa|)_+ \text{sign}(\mu),$$  \hspace{1cm} (6)

i.e., for $\kappa > 0$, $\tau(\mu, \kappa) = 0$ for $|\mu| \leq |\kappa|$, $\mu - \kappa$ for $\mu > \kappa$ and $\mu + \kappa$ for $\mu < -\kappa$.

When the ideal weight $\mu$ is small ($|\mu| \leq \kappa$), then the sparse weight is 0 ($m = 0$). All small components are replaced by 0. This confers sparsity on the model. Second, for $\mu > \kappa$, $m = \mu - \kappa$. This corresponds to a partial adjustment towards the correct value $\mu$. This motivates the term “anchoring and adjustment,” a phenomenon demonstrated by Tversky and Kahneman (1974) and discussed below.

Let us continue our analysis of the sparse max by consider the case where:

$$u(a, m, x) = v(a, m_1x_1, ..., m_nx_n)$$  \hspace{1cm} (7)

and calculate the sparse max $\text{smax}_a u(a, \mu, x)$. We consider the case where the DM views the $x_i$’s as uncorrelated. We analyze the cases of ex ante and ex post attention in turn.

**Proposition 1** When “attention is chosen before seeing the variables”, the $\text{smax}$ operator can be equivalently formulated as: $a^* = \arg \text{smax}_a|m^d, \kappa, \sigma_{x_i} v(a, (\mu; x_i)_{i=1...n})$ and $u^* = \text{smax}_a|m^d, \kappa, \sigma_{x_i} v(a, (\mu; x_i)_{i=1...n})$. The attentional policy is:

$$m^*_i = \mu_i \tau \left(1, \frac{\kappa \sigma_a}{\sigma_{x_i} \cdot \partial a/\partial x_i} \right)$$  \hspace{1cm} (8)
where $\partial a/\partial x_i = -\mu_i v_{aa}^{-1} v_{a,x_i}$ is the traditional marginal impact of a small change in $x_i$, evaluated at the default model. The action is:

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

(9)

and the utility is $u^* = u(a^*, \mu, x)$.

**Proposition 2** When “attention is chosen after seeing the variables”, the smax operator can be equivalently formulated as: $a^* = \arg\max_a v(a, (\mu_i x_i)_{i=1..n})$ and $u^* = \max_{a, (\mu_i x_i)_{i=1..n}} v(a, (\mu_i x_i)_{i=1..n})$. The agent uses the truncated version $x_i^*$ of variable $x_i$, according to:

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

(10)

where $\partial a/\partial x_i = -\mu_i v_{aa}^{-1} v_{a,x_i}$ is the traditional marginal impact of a small change in $x_i$, evaluated at the default model. The action is $a^* = \arg\max_a v(a, \mu_1 x_1^*, ..., \mu_n x_n^*)$ and the utility is $u^* = u(a^*, \mu, x)$.

To interpret the sparse max further, take the example $\kappa = 0.3$. Hence, the sparse max procedure in (8) implies:

“Eliminate each feature of the world (i.e., $i = 1..n$) that would change the action by less than a fraction $\kappa = 30\%$ of the standard deviation of that action” (i.e., eliminate the $x_i$ such that $|\sigma_i \times \partial a/\partial x_i| < \kappa \sigma_a$).

This is how a sparse agent sails through life: for a given problem, out of the thousands of variables that might be relevant, he takes into account only a few that are important enough to significantly change his decision.

Then, after doing the simplification and removing many variables (many $m_i$ are 0), the DM chooses the action based on its simplified set of variables, eq. 9. Let us illustrate this by an example.13

**Example 1** Take the quadratic loss problem, (3). The traditional, non-sparse answer is: $a^{NS} = \sum_{i=1}^n \mu_i x_i$. The sparse answer is, when with ex ante attention:

$$a_{ex\ ante}^* = \sum_{i=1}^n \tau \left( \mu_i, \frac{\kappa \sigma_a}{\sigma_{x_i}} \right) x_i$$

(11)

while with ex post attention, it is $a_{ex\ post}^* = \sum_{i=1}^n \tau (\mu_i x_i, \kappa \sigma_a)$.

13Also (normalizing $\sigma_{x_i} = \sigma_a = 1$), it is easy to see that $m$ has at most $\min\left( \|\mu_i/\kappa\|_1, \|\mu_i/\kappa\|_2^2 \right)$ non-zero components (because $m_i \neq 0$ implies $|\mu_i/\kappa| \geq 1$). Hence, even with infinite-dimensional $\mu$ and $m$, provided the norm of $\mu$ is bounded, $m$ has a finite number of non-zero components, and is therefore sparse.
2.3 Psychological Underpinnings

The model is based on very robust psychological facts: (i) limitedness of attention and working memory; (ii) use of defaults and anchor; (iii) anchoring and adjustment processes. I take them in turn.

**Limited attention and sparsity**   It is clear that we do not handle thousands of variables when dealing with a specific problem. For instance, research on working memory documents that people handle roughly “seven plus or minus two” items (Miller 1956, Kahneman 2011). At the same time, we do know – in our long term memory – about many variables, \(x\). The model, roughly, represents that selective use of information: In step 1, the mind contemplates thousands of \(x_i\), and decides on which handful it will bring up for conscious examination. Those are the variables with a non-zero \(m_i\). We simplify problem, and can attend to only a few things – this is what sparsity represents.\(^{14}\)

**Reliance on defaults**   What about when we have no time to think? What comes to mind? This is represented by the vector \(m^d\) – the default parametrization of variables taken into account when we have no time to think.\(^{15}\) This default model, and the default action \(a^d\) (which is the optimal action under the default model) corresponds to “system 1 under extreme time pressure”. The importance of default has been shown in a growing literature (Carroll et al. 2009). Here, the model default is very simple (basically, it is “do not think about anything”, \(m^d = 0\)), but it could be enriched, following other models (e.g. Gennaioli and Shleifer 2010).

**Anchoring and adjustment**   The mind, in the model, anchors on the default model. Then, it does a full or partial adjustment towards the truth. This is exactly the psychology of “anchoring and adjustment,” as dubbed by Tversky and Kahneman (1974). There

\(^{14}\)There is a rich literature in psychology featuring elimination of dimensions (e.g., Tversky’s “Elimination by aspect” 1972, and the literature reviewed in Payne, Bettman and Johnson 1993; see also Gabaix, Laibson, Moloche and Weinberg 2006). The theme is that, given processing cost, the DM must not (and indeed, cannot) consciously pay attention to many dimensions. This is very intuitive, and there is strong experimental support for this, e.g., by looking at active clicks in information lookups (Payne, Bettman and Johnson 1993). Partly unconsciously, the mind monitors many things in parallel; when they’re important or unusual enough they’re brought to consciousness. Most of the time, it is modelled in psychology as a process. The disadvantage is that the whole search process needs to be simulated to obtain actual predictions, so that the predictions are somewhat opaque and cumbersome. The sparsity-based model is a model of elimination of dimension, which eschews any step by step process, hence obtaining a closed-form representation of what agents will really take into account. I have tried to optimize it to make it tractable and widely applicable, while trying to capture some core psychology of inattention.

\(^{15}\)Any model of limited attention needs something akin to a “default”. Bayesian models, for instance, need a “prior.”
is anchoring on a default value and partial adjustment towards the truth (e.g., people pay only partial attention to the base rate when forming probability inferences); “People make estimates by starting from an initial value that is adjusted to yield the final answer [...]. Adjustments are typically insufficient. That is, different starting points yield different estimates, which are biased toward the initial value” (Tversky and Kahneman, 1974, p. 1129). It now has a plethora of experimental evidence for it, appears central to the ability of the mind to find its way in complex problems (Kahneman 2011, chapter 11).

In the model, this effect is generated by the anchoring and adjustment function $\tau$. It exhibits anchoring on the default model, and partial adjustment towards the truth. It would be interesting to investigate experimentally the $\tau$ function – perhaps to refine it. The comparative statics make sense (less important variables are less used). The quantitative forms would make sense too. When the variable has high values, it is largely taken into account. Hence, even though there is no specific experimental evidence on this function, the extend psychological evidence supports its basic elements.

**Purposeful attention, directed to a priori important things** In this model, the DM pays more attention to more important things. Recent models show the rich psychological implication of that basic fact. In Bordalo, Gennaioli and Shleifer (2012), agents choosing between two gambles pay more attention to states where the two gambles are most different. In Koszegi and Szeidl (2012), people focus more on features that most differ in the choice set.

**Discussion** When presenting any boundedly rational procedure, a certain amount of modelling decisions have to be made. Going away from the safe shores of rationality, we venture into the unknown. An example is the literature on learning in games (Fudenberg and Levine 1998), which features somewhat ad hoc algorithms, like fictitious play. Indeed, it may be instructive to note that many of the useful innovations in basic modelling have started without any axiomatic basis: prospect theory, hyperbolic discounting, learning in games, fairness models, Calvo pricing etc. The axioms came much later, if ever.

Criteria to judge a model include: 1. Usability of the model: Portability and tractability. 2. Formal properties (e.g. representation invariance); 3. Predictions. 4. Intrinsic necessity / axioms. I'll insist on criteria 1–3, while delegating criterion 4 to the online appendix. I’d like to contrast the model with other approaches such as noisy signals with Bayesian

---

16 The online appendix shows, however, that given the quadratic-linear structure of the model its scaling factors (e.g. the multiplication by $\|V_{nm,\eta}\|$ and the linke) are rendered necessary by an axiom that the model should be invariant to the choice of units.
updating and fixed costs, which may do better on criterion 4, but are quite problematic for criterion 1. The sparsity model, we shall see, is very easy to use. I will illustrate that in a core question of economics: consumer and equilibrium theory. It will make one prediction which I believe is distinctive and quite true, regarding the impact on limited attention on cross-elasticities (the Slutsky matrix).

Isn’t that a complex problem? One could object that it is simpler to optimize simply on $a$, like in the traditional model, than on $a$ and $m$, like in the sparse max. This is not the case when the situation is seen the following way: at time 0, so to speak the agent chooses an “attentional policy”, i.e. the $m^*$. Then, he is ready to react to many situations, with a precompiled, sparse, attention that allows him to focus on just a few variables. In that way, it is economical to use something like the sparse max.

In addition, the sparse max leads (at least in many situation) to a quite simple algorithm (for the agents), as shown in Proposition 2.

If you know $x$ and $\mu$, why not use them? One interpretation is that it is system 1 who, at some level, knows $x$ (i.e. has a sense that the volatility of the interest isn’t important), and chooses not to bring it to the attention of system 2, for a more thorough analysis. System 1, chooses the representation $m$, while System 2 takes care of the actual maximization, with a simpler problem.17

How does the agent know $V_{aa}, V_{am}$? This knowledge is that system 1 has a sense of what variables are important, and which are not, at the default model. It seems intuitive that for many problems at least, agents do have a sense of which variables are important or not. To keep the model simple, this sense is encoded by a knowledge of $V_{aa}$ and $V_{am}$.

Why not a fixed cost or a quadratic cost? The “taste for sparsity” in (4) features a linear cost, $|m_i|$ (normalizing $m_i^d = 0$). Why not a more general cost, say $|m_i|^\alpha$? First, that could well be a good idea, in some cases. Then (assuming that the $x_i$ are uncorrelated), we obtain a different different $\tau$ function. The online appendix works the case of fixed cost. However, in general, an $\alpha > 1$ would generate no sparsity (generically, all $m_i$ would be non-zero). An $\alpha < 1$ (including a fixed cost, which is $\alpha = 0$) would generate a non-convex problem ($|m|^\alpha$ is not convex then). This would generate a very intractable problem in general (an NP-complete problem).18 Hence, the case $\alpha = 1$ is the only one to generate both sparsity and tractability.

The agent’s decisions depend on the basis, i.e. of which $x_i$ are available. That is arguably a desirable feature of the model. For instance, take the Chetty, Looney and Kroft (CLK,

---

17See e.g. Fudenberg and Levine (2006), Brocas and Carillo (2008) for very different dual-self models.

18This special role of $\alpha = 1$ has been elucidated by Tibshirani (1996). Hassan and Mertens (2011) also use $\alpha = 1$. 

14
2009) results, where $x_1$ is the sticker price, and $x_2$ extra tax (all demeaned) and consumption should reflect the total price ($x_1 + x_2$), then a sparse agent will react with a high weight on the sticker price, and a low weight the tax (because the tax rate is expected to vary less). This conforms to the experimental evidence of CLK. To model that the sticker price is very “available”, we put it on the basis, and don’t put the total price on the basis: calculating the real price requires more effort. In contrast, a “pure” rational inattention-with-entropy agent (à la Sims 2003) has predictions that do not depend on the basis. This is one of the appealing features of the Sims model. However, this leads to counterfactual predictions, in this context at least. The Sims model will dampen equally the sticker price and the tax, i.e. people will perceive $\lambda(x_1 + x_2)$ with the same dampening. The experimental results of CLK support that lower weight on the seemingly less important part, the tax. To account for such empirical facts, the basis available to the agents does indeed matter.

Invariance by rescaling The model is invariant by rescaling; its predictions are independent of the units in which the components of $m$ and $a$ are measured. For instance, $|m_t| \|a_{m,a}|$ does not depend on the units of $m_t$. More generally, the model is invariant (for small changes) to reparametrizations of the action. For instance, if the agent picks consumption or log consumption, the representation chosen by the decision maker is the same. This adds some robustness and ease of use to the model. In addition, the model is invariant by (possibly nonlinear) transformation of the utility function (see online appendix, section 9.1)

Calibration We can venture a word about calibration. As a rough baseline, we can imagine that people will search for information that accounts for at least $\xi^2 = 10\%$ of the variance of the decision, i.e., if $|\mu_i| \sigma_i < \xi \sigma_a$. Then, using (101), we find $\kappa \simeq \xi$. That leads to the baseline of $\kappa \simeq 0.3$. The reader may find that, rather than 10%, $\xi^2 = 1\%$ is better (though this may be very optimistic about people’s attention), which corresponds to $\kappa = \sqrt{1\%} = 0.1$ – a number still in the same order of magnitude as $\kappa \simeq 0.3$. As it turns out, in subsequent work (Gabaix 2012b), the calibration $\kappa \simeq 0.3$ works quite well in predicting the subject’s behavior in experimental games.

Section 6.1 contains more justification. Given it involves more math, I defer it.

---

19 More generally, with the Sims (2003) model, the solution to the quadratic problem (3) is: $\mathbb{E}[a_{\text{Sims}} | x] = \lambda \sum_i \mu_i x_i$, for a $\lambda \in [0, 1]$ which increases with the attention budget. Hence, all dimensions are dampened equally. In contrast, in the sparse model (11), less important dimensions are dampened more.

20 This is why other researchers on rational inattention (Ma´ckowiak and Wiederholt 2009, Woodford 2012) also have basis-dependent models, or their counterpart in their framework.
3 Sparse Max with Constraints

3.1 Formulation of the Sparse Max with Constraints

We extend the model, so that it handles maximization under constraints. The decision maker wishes to maximize a utility \( u \) subject to \( K \) constraints, i.e. (2). For instance, \( B \) could be a budget constraint, \( B(a, m, x) = w - p(m) \cdot c \) where \( p(m) \) is a vector of prices under the model \( m \). A leading example is the following: We start from a default price \( p^d \). The new price is \( p_i(m) = p^d_i + x_i \), while the price perceived by the agent is \( p_i(m) = p^d_i + m_i x_i \). The ideal vector of weights is thus \( \mu = (1, ..., 1) \). This can be rewritten as:

\[
p_i(m) = m_i p_i + (1 - m_i) p^d_i
\]  

(12)

We assume that \( u \) and \(-B\) are concave in \( a \).

**Definition 2** (Sparse max operator with constraints). The sparse maximum of problem (2) is defined by the following two steps. We call \( L(a, m, x, \lambda) := u(a, m, x) + \lambda \cdot B(a, m, x) \) the Lagrangian associated with the problem.

1. i. Select the Lagrange multiplier \( \lambda^d \in \mathbb{R}_+^K \) associated with the problem (2) at the default model \( m^d \), so that the optimal action in under the default model is \( \max_a L(a, m^d, x, \lambda^d) \).

   ii. **Selection of model** \( m^* \). Use the sparse max operator of Definition 1, without constraints, for the function \( L(a, m, x, \lambda^d) \). That returns a representation \( m^* \).

2. **Action taking the constraints into account.**

   i. Form a function \( a(\lambda) = \arg \max_a L(a, m^*, x, \lambda) \), the optimal action under model \( m^* \) with Lagrange multiplier \( \lambda \).

   ii. Maximizes utility subject to the “true” constraint: \( \lambda^* = \arg \max_{\lambda \in \mathbb{R}_+^K} u(a(\lambda), m^*, x) \) s.t. \( B(a(\lambda), \mu, x) \geq 0 \). With just one binding constraint this is equivalent to choosing \( \lambda^* \) so that the constraint does bind: \( B(a(\lambda^*), \mu, x) = 0 \). The resulting sparse action is \( a^* := a(\lambda^*) \). Utility is \( u^* = u(a^*, m, x) \).

**Model justification** This device does not depend on the sparsity perspective to translate a BR maximum without constraints, into a BR maximum with constraints. It could be reused in other contexts. Let me show how it is a reasonable way to extend the operator to add constraints.

The first step seems quite natural. To replace a problem with constraints into an unconstrained problem, we add the “price” of the constraints to the utility. Step 1.i of Definition
2 picks a Lagrange multiplier $\lambda^d$, using the default representation $m^d$. For instance, in a consumption problem (1), $\lambda^d$ is the “marginal utility of a dollar”, at the default prices. This way, in Step 1.ii we can use Lagrangian $L \left(a, m^d, x, \lambda^d\right)$ to encode the importance of the constraints, and maximize it without constraints, so that the basic sparse max can be applied: that yields the chosen attention allocation, $m$.

The second step comes from the following intuition. Take the consumer problem (1): we would like to keep the usual psychological / economic reasoning that the ratio of marginal utilities is the ratio of perceived prices: $\frac{u'_1}{u'_2} = \frac{p_1(m^*)}{p_2(m^*)}$. Pretty much the only way to keep that intuition and budget constraint is to say that $\frac{w_0}{\mu} = \frac{\bar{p}_1(m^*)}{\bar{p}_2(m^*)}$, for some $\lambda$ that will maximize utility, subject to the budget constraint – i.e., makes the constraint bind. In step 2, the agent “hears clearly” if the budget constraint binds.23

**Advanced topic: Duality for the SparseMax** In addition, this formulation has a nice duality property that will be useful later. The reader is encouraged to skip this paragraph at the first reading.

**Proposition 3** (Duality is respected by the sparse max). Consider the two dual optimization problems (with $u$ concave in $a$, $w$ convex in $a$, and $\hat{u}$, $\hat{w}$ real numbers): (i) $\bar{w}(\hat{w}) := \text{smax}_a u(a, m, x) \text{ s.t. } w(a, m, x) \leq \hat{w}$, (ii) $\bar{w}(\hat{u}) := \text{smin}_a w(a, m, x) \text{ s.t. } u(a, m, x) \geq \hat{u}$. Then the two problems are dual of each other, i.e. $\bar{w}(\bar{u}(\hat{w})) = \hat{w}$ and $\bar{w}(\bar{u}(\hat{u})) = \hat{u}$, for all $\hat{u}, \hat{w}$. They also yield the same model $m^*$.

3.2 Application: consumption under a budget constraint

Let us apply the sparse max with constraints to the canonical problem of choosing a consumption bundle under a budget constraint. We rewrite (1) with more compact notations, $c = (c_1, ..., c_n)$: we define the Marshallian demand as $c(p, w) = \text{arg max}_{c \in \mathbb{R}^n} u(c)$ subject to $p \cdot c \leq w$. Hence the constraint is $B(c, m, p) = w - p(m) \cdot c$. Definition 2 has two parts: the first part states the optimal perceived price that we will call $p^m := p(m^*)$ as a short-hand. Let us defer its analysis, to focus the second part: given perceived the price $p^m$, what is the desired consumption $c^*(p, w)$ of a sparse agent? We call $c^*(p, w)$ the Marshallian demand under the traditional, or rational, model.

---

21 The equivalence with $B(a \lambda^*, \mu, x) = 0$, mentioned in Step 2, is justified in Lemma 2 in the appendix.

22 Otherwise, as usual, if we had $\frac{u'_1}{u'_2} > \frac{p_1(m^*)}{p_2(m^*)}$, the consumer could consume a bit more of good 1 and less of good 2, and project to be better off.

23 This model, with a general objective function and $K$ constraints, delivers, as a special case, the third adjustment rule discussed by the NBER WP version of Chetty, Looney and Kroft (2009) in the context of consumption between two goods and one tax.
Proposition 4 (Marshallian demand). Given the perceived price $p^m$ and the true price $p$, the demand of a sparse agent is

$$c^s(p, w) = c^r(p^m, w')$$  \tag{13}$$

where the as-if budget $w'$ solves $p \cdot c^r(p^m, w') = w$.

When rational demand is linear in wealth ($c^r(p, w) = c^r(p, 1)w$), the demand of a sparse agent is:

$$c^s(p, w) = \frac{c^r(p^m, w)}{p \cdot c^r(p^m, 1)}$$  \tag{14}$$
i.e. $c^s_i(p, w) = \frac{c^r_i(p^m, w)}{\sum_j p_j c^r_j(p^m, 1)}$.

The case of the demand linear in wealth is perhaps most intuitive. Let us say that the consumer goes to the supermarket, with a budget of $w = \$95$. Because of the lack of full attention to prices, the value of the basket in the cart is actually $c^r(p^m, w) = \$100$. Then, the consumer buys 5\% less of all the goods, to hit the budget constraint, and spend exactly $\$95$ (this is the adjustment factor $1/p \cdot c^r(p^m, 1) = 0.95$).

In the general case, the demand equation means: “Given an as-if budget $w'$, optimize with perceived prices $p^m$: that gives $c^r(p^m, w')$. If the budget constraint isn’t saturated, change the budget $w'$ so as to hit the budget constraint”.24

Here are three concrete examples. Recall that in $p^m = p(m^*)$ the $m$ indicates “in the agent’s model” rather than an exponent.

Example 2 (Demand by a sparse Cobb-Douglas agent). Take $u(c) = \sum_{i=1}^n \alpha_i \ln c_i$, with $\alpha_i \geq 0$. Demand is: $c^s_i(p, w) = \frac{\alpha_i}{\sum_j \alpha_j p_j} w$.

Example 3 (Demand by a sparse CES agent). Take $u(c) = \sum_i c_i^{1-1/\eta} / (1 - 1/\eta)$, with $\eta > 0$. Demand is: $c^s_i(p, w) = (p_i^m)^{-\eta} \times \frac{w}{\sum_j p_j (p_j^m)^{-\eta}}$.

Example 4 (Demand by a sparse agent with quasi-linear utility). Take $u(c) = \sum_{i=1}^{n-1} v_i(c_i) + c_n$, with $v_i$ concave. Demand for good $i < n$ is independent of wealth and is: $c^s_i(p) = c^r_i(p^m)$.

In the above example, the demand of the sparse agent is simply the rational demand given the perceived price. The budget constraint is “absorbed” by the residual good $n$. This is often the most plausible way budget is respected: extra over- and under-spending is absorbed in a general fund, “savings”, say.

---

24 At the margin, if the consumer is short $\$1$, say, the he increases his consumption by $\frac{\partial c^r(p^m, w)}{\partial w}$ times $\$1$: the adjustment is proportional to $\frac{\partial c^r(p^m, w)}{\partial w}$, not $c^r$. 

18
Determination of the attention to prices, $m^*$. The model endogenizes the weights in (12) as follows.

**Proposition 5** (Attention to prices). In the basic consumption problem, assuming either that price shocks are uncorrelated, or that utility is separable, attention to prices is:

$$m^*_i = \tau \left( 1, \frac{\kappa \sigma_{c_i}/c^d_i}{\psi_i \sigma_{p_i}/p^d_i} \right)$$

where $\psi_i = -u_{c_i}(u^\prime - 1)_{ii}/c^d_i$ is the elasticity of demand for good $i$.

Empirical work already measures something akin to $m^*_i$. For instance, Allcott and Wozny (2012) find that car buyers behave as if they put a weight $m = 0.72$ on gas prices, rather than a weight of 1. Chetty, Looney and Kroft (2009) find that people take the tax partially into account, with a $m = 0.35$. Proposition 5 states that attention to prices is greater for goods with more volatile price, and to goods with higher price elasticity\(^{25}\). These predictions seem sensible, though not extremely surprising.\(^{26}\) Still they might provide plausible hypotheses for empirical work, as they express attention in terms of terms that are observable in principle (except for the parameter $\kappa$, but it is hypothesized to be the same across goods). It would be interesting to test them. It is not easy, but we shall see the results below (esp. (17)) offer a way to do that.

For our theoretical issues, what is important is that we do have some procedure to pick the $m_i$, so that the model is closed; this allows us to derive the consequences of that limited attention to prices. More surprises happen here, as we shall now see.

### 4 Consumer Theory

#### 4.1 Basic consumer theory

We revisit the theory of consumer’s behavior, with a sparse flavor. Some notations: we call $c(p, w) = \arg \max_c u(c)$ s.t. $p \cdot c \leq w$ the Marshallian demand given price vector $p$ and budget

\(^{25}\)Here $\psi_i$ is, more precisely, the compensated own-price elasticity of demand for a rational agent, i.e. $\psi_i = \frac{1}{c_i} S_{ii}$, where $S_{ii} = \partial c_i/\partial p_i$ is the compensated price sensitivity in the Slutsky matrix (see the online appendix, section 9.1.2).

\(^{26}\)The term $\sigma_{c_i}/c_i$ can be directly measured in empirical work. In theory, it could just be a parameter, e.g. $\kappa \sigma_{c_i}/c_i = 0.1$ if only changes that make consumption move by 10% are deemed important enough for the DM. Another variant is to assume that, under the default model, the agent just pays attention to his wealth, wealth has standard deviation $\sigma_w$, and $\sigma_{c_i} = \frac{\partial u_i}{\partial c_i} \sigma_w$ is the standard deviation of $c_i$ under the default model. If the $\sigma_{c_i}$ is the actual standard deviation of consumption for the consumer, then there is a fixed point, and the solution is simple in the case where one good $n$ ("money") has linear utility: $m_i = 1/(1 + \kappa)$. 
$w$, and $V(p, w) = u(c(p, w))$, the indirect utility function. We call $e(p, v) = \min_c p \cdot c$ s.t. $u(c) \geq v$ the expenditure function — the minimum expenditure required to reach utility level $v$ — and the associated Hicksian demand $h(p, v) = \arg \min_c p \cdot c$ s.t. $u(c) \geq v$. Superscripts $s$ denotes a sparsely attentive version, and $r$ for the traditional, rational, model. Hence, we define $c^s$ and $e^s$ in the same way, with a smax and smin.

The Slutsky matrix is an important object, because it encodes both elasticities of substitution and welfare losses (eq. 18). It is defined to be:

$$S_{ij}(p, w) := \frac{\partial c_i(p, w)}{\partial p_j} + \frac{\partial c_i(p, w)}{\partial w} c_j(p, w)$$

(15)

With the traditional agent, the most surprising fact about it is that it is symmetrical: $S_{ij} = S_{ji}$, i.e., the (compensated) impact of the price of good $j$ on the consumption of good $i$, is exactly the same as the impact of the price of good $i$ on the consumption of good $j$. Mas-Colell, Whinston and Green (1995, p.70) comment “Symmetry is not easy to interpret in plain economic terms”. Varian (1992, p.123) concurs: “This is a rather nonintuitive result.”

We now present a less rational alternative, under the sparsity model. We use the price misperception discussed in (12). The consumer is partially inattentive, and “sees” only part of the price change. We first derive the marginal Marshallian demand.

**Proposition 6** The marginal Marshallian demand $c^s(p, w)$ is, at the default price $p^d$:

$$\frac{\partial c^s_i}{\partial p_j} = \frac{\partial c^r_i}{\partial p_j} \times m_j - \frac{\partial c^r_i}{\partial w} c_j \times (1 - m_j)$$

(16)

It comprises attenuated attention to the price $\frac{\partial c^r_i}{\partial p_j} m_j$, and compensation to satisfy the budget constraint, $\frac{\partial c^r_i}{\partial w}$. We are now ready to derive the Slutsky matrix.

**Proposition 7** (Slutsky matrix). Evaluated at the default price, the Slutsky matrix $S^s$ is, compared to the traditional matrix $S^r$:

$$S^s_{ij} = S^r_{ij} m_j$$

(17)

i.e. the sparse demand sensitivity to price $j$ is the rational one, times $m_j$, the salience of

---

27 Hence $V$ is only the “consumption utility” of the DM. One could imagine a richer notion using psychic costs, but that raises many other interesting issues (Bernheim and Rangler 2009).

28 In the traditional model, this comes from the fact that $S_{ij}(p, v) = \frac{\partial^2 e(p, v)}{\partial p_i \partial p_j}$, which is symmetric, as the order in which one takes derivatives does not matter. Not coincidentally, that mathematical property (“Young’s theorem”) is itself surprising at first.
price \( j \). As a result the sparse Slutsky matrix is not symmetric in general. Sensitivities corresponding to “non-salient” price changes (low \( \mu \)\( \sigma \)) are anomalously low.

Instead of looking at the full price change, the consumer just reacts to \( \mu \sigma \) of it. Hence, she’s typically less responsive than the rational agent. For instance, say that \( m_i > m_j \), so that the price of \( i \) is salient, but the price of good \( j \) isn’t very salient. The model predicts that \( S_{ij}^s \) lower than \( S_{ji}^s \). That’s because good \( j \)'s price isn’t very salient, so quantities don’t react much to it.

The non-symmetry of the Slutsky matrix indicates that, in general, a sparse consumer cannot be represented by a rational consumer who simply has different tastes or adjustment costs (e.g. a more inelastic demand). Such a consumer would feature a symmetric Slutsky matrix.29

**Proposition 8** (Estimation of limited attention) Choice data allow to recover the allocation of attention \( m \), up to a multiplicative factor \( \overline{m} \). Indeed, suppose that an empirical Slutsky matrix \( S_{ij}^s \) is available. Then, \( m \) can be recovered as \( m_j = \overline{m} \prod_i \left( \frac{S_{ij}^s}{S_{ji}^s} \right)^{\gamma_i} \), for any set of \( \gamma_i \) s.t. \( \sum \gamma_i = 1 \).

**Proof:** We have \( \frac{S_{ij}^s}{S_{ji}^s} = \frac{m_j}{m_i} \), so \( \prod_i \left( \frac{S_{ij}^s}{S_{ji}^s} \right)^{\gamma_i} = \prod_i \left( \frac{m_j}{m_i} \right)^{\gamma_i} = m_j / \overline{m} \), for \( \overline{m} := \prod_i m_i^{\gamma_i} \).

Equation (17) makes tight testable predictions, as the sparsity model is a parsimonious extension of the traditional one.30 It allows for the recovery of the attention terms \( m_j \), up to a multiplicative factor. The underlying “rational” matrix can be recovered at \( S_{ij}^c := S_{ij} / m_j \). A testable prediction is that \( S^c \) should be symmetric.31 There is literature on consumption, estimating Slutsky matrices (e.g. Deaton and Muellbauer 1980), but it does not have in mind a specific direction for a potential failure of symmetry, and does not seem to explore the potential role of inattention. It would be good to revisit this literature, emphasizing non-salient prices, using the comparative statics and specific functional form predicted by the model (Propositions 5 and 7).

---

29 Intuitively, the reason for this impossibility is the following: suppose that price 2 is not salient, so \( S_{22} \) is small. That might induce the modeler to represent the consumer as a rational agent with inelastic demand for good 2. However, that move would make \( S_{21} \) be small too. Hence, it is difficult (and indeed generally impossible, by Proposition 7) to represent the inattentive consumer by a rational consumer who is simply more price inelastic.

30 It complements the earlier parsimonious deviation from symmetry identified by Browning and Chiappori (1998), who have in mind a very different phenomenon: intra-household bargaining. Quantitatively, the \( m \) vector adds \( n - 1 \) extra degrees of freedom (d.f.), which may be compared to the \( 2n + O(1) \) d.f. added in the Browning-Chiappori model (I absorb low-dimensions restrictions in the \( O(1) \) term). The \( S^c \) matrix has \( n^2 + 3n + O(1) \) d.f., a good restriction compared to unrestricted symmetric matrix with \( n^2 \) d.f.

31 The Slutsky matrix, by itself, does not allow to recover the extra multiplicator factor: for any \( \overline{m} > 0 \), \( S^c \) admits a dilated factorization \( S_{ij}^c = (\overline{m}^{-1} S_{ij}^c) (\overline{m} m_j) \), so the Slutsky matrix can give \( m \) only up to a multiplicative factor \( \overline{m} \). To recover this factor, one needs to see how the demand changes as \( p^d \) varies.
Qualitatively, the extant evidence is encouraging. Besides the tax salience already mentioned, there is literature pointing to effects qualitatively consistent with Proposition 7: the literature on obfuscation and shrouded attributes (Gabaix and Laibson 2006, Ellison and Ellison 2009). Those papers (Abaluk and Gruber 2011, Anagol and Kim 2012, Brown, Hossain and Morgan 2010, Chetty, Looney and Kroft 2009) find direct field evidence that some prices are neglected (at least partially) by consumers, leading to an underreaction to prices of add-ons. However, I am not aware of direct tests of the quantitative structure predicted by Proposition 7— which makes sense, given this proposition and its functional form are new.

4.2 More advanced considerations on consumer theory

This subsection may be skipped at the first reading. It contains more advanced considerations on consumer theory. We define $M := \text{diag}(m_1, ..., m_n)$ the attention matrix, so that (17) becomes $S^s = S^r M$.

The Weak Axiom of Revealed Preferences is violated We shall see that the venerable weak axiom of revealed preferences (WARP) is violated, in a way that makes sense psychologically. First, recall that it implies for a small price change $\delta p$, the compensated change in consumption is $\delta c = S\delta p$. In the traditional model, it satisfies $\delta p \cdot \delta c^r \leq 0$, i.e. $\delta p^r S^r \delta p \leq 0$: the Slutsky matrix is negative semi-definite. Here is the version with a sparse agent.

Proposition 9 (Violation of the WARP). Suppose that $SM p^d \neq 0$, and reason at the default price. The agent’s decisions violate the weak axiom of revealed preferences: there is a price change $\delta p$, such that the corresponding change in consumption $\delta c^s = S^s \delta p$ satisfies $\delta p \cdot \delta c^s > 0$. In other terms, the Slutsky matrix $S^s$ fails to be negative semi-definitive. However, for all price changes $\delta p$, we have: $\delta p^m \cdot \delta c^s \leq 0$.

WARP fails, but something like it holds: at $p^d$, we have $\delta p^m \cdot \delta c^s \leq 0$, i.e. $\sum m_i \delta p_i \delta c^s \leq 0$. Hence we do preserve that “if prices go up, (compensated) demand goes down,” in a salience-weighted sense.

Here is the intuition, more detailed in the appendix. Suppose that the agent pays attention to the car price, but not gas. Suppose that the car price goes down, but gas price goes up by a lot. A rational agent will see that the total price of transportation (gas+price) has gone up, so he consumes less of it: $\delta c^r \cdot \delta p < 0$, with $c^r = (c_{\text{car}}, c_{\text{gas}}, c_{\text{food}})$. However, a sparse agent just sees that the car prices went down, so he consumes more transportation. As a result $\delta c^s \cdot \delta p > 0$. 
Proposition 9 shows that this logic is quite general. The condition \( SMP^d \neq 0 \) is quite weak – with two goods it essentially means that \( m_1 \neq m_2 \). Here is a simple example example with two goods: \( p^d = (1, 1) \) and \( S^r = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \) (which can come from \( u = \ln c_1 + \ln c_2 \)), and good 2 non-salient, \( m = (1, 0) \). Consider \( \delta p = (1, 2) \) (which the reader may wish to multiply by some small \( \varepsilon > 0 \)). As the price of good 2 increases more, the rational consumer changes his demand by \( \delta c^r = S^r \delta p = (1, -1) \), i.e. consumes less of good 2, and more of good 1. However, the sparse consumer perceives \( \delta p^m = M \delta p = (1, 0) \), so he perceives only the good 1’s price increases more. His consumption change is the opposite of the rational agents’: \( \delta c^s = S^s \delta p^m = S^s \delta p = (-1, 1) = -\delta c^r \). Hence, we have \( \delta p \cdot \delta c^r = -1 < 0 < \delta p \cdot \delta c^s = 1 \), a violation of WARP.

**Hicksian demand, expenditure function and welfare**  We start with an explicit expression for Hicksian demand and expenditure function.

**Proposition 10** The sparse Hicksian demand is: \( h^s (p, v) = h^r (p^m, v) \). The sparse expenditure function is \( e^s (p, v) = p \cdot h^r (p^m, v) \).

Indeed, this is the consumption chosen by a consumer attempting to minimize expenditure under the perception of prices \( p^m \). We next examine the link between expenditure function and Slutsky matrix, which is tight in the traditional case.

**Proposition 11** (Link between Slutsky matrix and expenditure function). The expenditure function satisfies, at the default price: \( e^s = e^r, e^s_p = e^r_p \), and: \( e^s_{pp} = e^r_{pp} - (I - M)' e^r_{pp} (I - M) \), i.e. \( e^s_{pp, p_j} = e^r_{pp, p_j} \times (m_i + m_j - m_i m_j) \). Note that \( e^s_{pp, p_j} = S^s_{ij} + S^s_{ji} - \frac{S^s_{ij} S^s_{ji}}{S^s_{ij}} \) rather than the traditional \( e^r_{pp, p_j} = S^r_{ij} \).

The expenditure function \( e^s_{pp, p_j} \) is “twisted” by the attention terms, \( m_i \). We have a relation somewhat more complex than in the traditional model (which gives \( e^r_{pp, p_j} = S^r_{ij} \)).

Let us now discuss the welfare losses from price misperception.

**Proposition 12** (Indirect utility function and welfare losses). At the default price \( p^d \), \( V, V_p, V_w, V_{pw}, V_{ww} \) are the same under the behavioral and the rational model, but \( V_{pp} \) differs:

\[
V^s_{pp} - V^r_{pp} = -V^r_w (e^s_{pp} - e^r_{pp}) = V^r_w (I - M) S^r (I - M) \tag{18}
\]

The intuition is simple: the utility loss \( (V^s_{pp} - V^r_{pp}) \) is equal to the extra expenditure \( e^s_{pp} - e^r_{pp} \) due to suboptimal behavior, times the utility value of money, \( V_w \). That suboptimal behavior is itself captured by a lack of substitution when reacting to prices, \( (I - M) S^r (I - M) \). As usual, we can expect the welfare losses to be quite small, as they are second order (e.g. Akerlof-Yelen 1985, Krusell-Smith 1996).
Shepard’s lemma, Roy’s identity  Their behavioral version is as follows.

Proposition 13 (Shepard’s lemma, Roy’s identity). Evaluated at the default price, we have Shepard’s lemma: $e_{pi}^s = h_i^s$ at $p = p^d$, and Roy’s identity: $c_i^r(p,w) + \frac{V_i(p,w)}{V_w(p,w)} = 0$. However, away from $p^d$, these expressions need to be modified. Indeed we have the modified Shepard’s lemma

$$e_{pi}^s(p,v) = h_i^s(p,v) + (p - p^m) \cdot h_p^r(p^m,v) m_i$$

(19)

and the modified Roy’s identity:

$$c_i^r(p,w) + \frac{V_{pi}}{V_w^s(p,w)} = [(p \cdot c_{w'}(p^m,w')) p^m - p] \cdot c_{w'}^r(p^m,w') m_i$$

(20)

Note that in (20) at $p^d = p = p^m$, we recover the traditional Roy’s identity because $p \cdot c_{w'}(p,w') = 1$.

Price perceptions in logs  Another interpretation works in logs, i.e. people pay attention to the price increases, rather than the dollar changes. Then, we obtain the loglinear model:

$$p_i^m = p_i^{m_i} (p_i^d)^{1-m_i}$$

(21)

rather than the linear model (12). The implications remain the same, as $\frac{\partial p_i^m(p,m)}{\partial p_i} = m_i$ at $p = p^d$, as in the linear case. Then, it is easy to verify that all properties in section 4.1 apply also with this log-linear formulation (except for the last two displayed equations of Proposition 13, where $m_i$ should be replaced by $\partial p_i^m/\partial p_i$). The loglinear model leads to slightly easier calculations in the competitive equilibrium.

5  Competitive Equilibrium and Edgeworth Boxes

5.1 Edgeworth Boxes: Extra-dimensional Offer Curves

Take a consumer with endowment $\omega \in \mathbb{R}^n$ (he is endowed with $\omega_i$ units of good $i$, $i = 1...n$). Given a price vector $p$, his wealth is $p \cdot \omega$, and so his demand is $D(p) := c(p,p \cdot \omega)$, which has values in $\mathbb{R}^n$. The offer curve (OC) is defined as the set of demands, as prices vary: $OC := \{D(p), p \in \mathbb{R}^n_{++}\}$.

Let us start with two goods ($n = 2$). Normally, the offer curve $OC$ is one-dimensional in the traditional model: because $c(p,p \cdot \omega)$ is homogeneous of degree 0 in $p$, only the relative price $p_1/p_2$ matters. However, in the sparsity model, $c(p,p \cdot \omega)$ is not homogeneous of degree 0 in $p$ any more (generally speaking – we will specify conditions later). Not just the relative
price \( p_1/p_2 \) matters. Hence, OC is described by two parameters \((p_1, p_2)\), so its image is 2-dimensional. We propose to call this an “extra-dimensional offer curve.” in the Edgeworth box.

The phenomenon is illustrated in Figure 2, which plots the offer curve of the rational (left panel) and sparse consumer (right panel), for a Cobb-Douglas utility.\(^{32}\) The gray area in the right panel is the offer curve, which is a 2-dimensional “ribbon”, with a pinch at the endowment \( \omega \).

We now formalize the notion that, if \( m \) does not have all equal components, then the offer curve has “one extra dimension” compared to the traditional model.

**Proposition 14** (Extra-dimensional offer curve). *Take a price \( p \) such that \( p^m \cdot \left( \frac{\partial p^m}{\partial p} \right)^{-1} (c - \omega) \neq 0 \). Then around \( D(p) \), the offer curve of the sparse agent has one extra dimension compared to the traditional model, i.e. it has dimension \( n \).*

\(^{32}\)It is instructive to stare at the result directly, with \( u(c_1, c_2) = \ln c_1 + \ln c_2 \) and \( p^d = (1, 1)' \). The BR agent’s offer curve is the set of \((c_1, c_2)\) such that: \( u'(c_2)/u'(c_1) = p_2^m / p_1^m \) and such that the budget constraint holds, \( p \cdot (c - \omega) = 0 \). Hence, the OC is the set of \((c_1, c_2)\) that satisfy the following two equations for some \( p_1 \) and \( p_2 \):

\[
\frac{c_1}{c_2} = \left( \frac{p_1}{p_2} \right)^{-m_1} p_2^{m_2 - m_1}, \quad \frac{p_1}{p_2} (c_1 - \omega_1) + c_2 - \omega_2 = 0
\]

Hence, when \( m_1 \neq m_2 \) (in the rational model, \( m_1 = m_2 = 1 \)), the offer curve is described by two parameters: \( p_1/p_2 \) and \( p_2^{m_1 - m_2} \). The offer curve is two-dimensional.
The restriction implies that \( c \neq \omega \): we do not start at the endowment (this can be seen in the “pinch” at \( \omega \) in Figure 2, right panel). It also implies that in the log linear model, \( m \) does not have all identical components — i.e., the consumer pays more attention to some goods then offers.  

Next, we show that when attention is limited, the offer curve is very wide indeed.

**Example 5 (Wide offer curves).** Suppose that there two goods, with different inattention \( (m_1 \neq m_2) \) in the loglinear specification (21). Then, any consumption that does not dominate the endowment nor is dominated it, is in the consumer’s offer curve.

This effect is of more mathematical interest, as it relies on potentially extreme prices and misperceptions. However, something is robust. The offer curve has “one extra dimension” compared to the traditional offer curve. The extra dimension might not cover the whole space, e.g. if there is “limited misperception.”

### 5.2 One-dimensional Set of Competitive Equilibria with Sparse Agents

There are \( A \) agents. Agent \( a \) has endowment \( \omega^a \in \mathbb{R}^n \), and the total endowment is \( \omega = \sum_{a=1}^{A} \omega^a \). We call \( Z (p) := \sum_{a=1}^{A} D^a (p) - \omega \) the economy’s aggregate excess demand function when prices are \( p \), with \( D^a (p) = c^a (p, p \cdot \omega^a) \). Call \( \mathcal{P}^* \) the set of equilibrium prices, i.e. of prices that lead zero excess demand:

\[
\mathcal{P}^* := \{ p : Z (p) = 0 \}.
\]

The resulting set of equilibrium allocations for consumer \( a \) is:

\[
\mathcal{C}^a := \{ \Delta^a (p) : p \in \mathcal{P}^* \}.
\]

Here, as in the traditional model, Walras’ law applies: equilibrium in the market for goods 1 to \( n - 1 \) implies equilibrium for good \( n \). So the set of equilibria is really \( \mathcal{P}^* = \{ p : Z_{-n} (p) = 0 \} \), where \( Z_{-n} = (Z_i)_{1 \leq i < n} \). As \( Z_{-n} \) is a function \( \mathbb{R}^n_{++} \to \mathbb{R}^{n-1} \), \( \mathcal{P}^* \) is still a 1-dimensional manifold (as in the traditional model).

In the traditional model with one equilibrium allocation, \( \mathcal{P}^* = \{ \lambda p^*, \lambda \in \mathbb{R}_{++} \} \), and \( \mathcal{C}^a \) is just a point, \( D^a (p^*) \) (more generally equilibria consists of a finite union of such sets, under

---

33 Indeed, if \( m = \mu 1 \), then \( \frac{\partial p^m}{\partial p_i} = \frac{p^m}{p_i} \delta_{ij} \), so \( p^m \cdot \left( \frac{\partial p^m}{\partial p} \right)^{-1} = \mu^{-1} p \), and the condition of Proposition 14 is equivalent to \( p \cdot (c - \omega) \neq 0 \), which is never satisfied.

34 Such a limited misperception would be model with “attention allocated ex post”: \( p_i (m) = p_i^L \cdot \exp \tau (\ln \frac{p_i}{p_i^L} \gamma, \omega \), where \( \alpha > 0 \) and \( \tau \) is as in (6). Then, \( \ln \frac{p_i (m)}{p_i^L} \leq \alpha \). With that model, we have a less extreme OC, but it remains extra-dimensional with dimension \( n \).

35 An equilibrium exists under a simple “non satiation” condition, see Debreu (1970). We simply need to assume Debreu’s Assumption A, which states that \( \| Z (p) \| \) becomes infinite when one price goes to 0, while the sum of prices remains constant. One can directly take Debreu’s existence Proposition, as he does not assume that \( Z (p) \) is homogeneous of degree 0. See also Shafer and Sonnenschein (1975) for equilibrium existence in non-standard economies.
weak conditions given in Debreu 1970). In the sparse setup without homogeneity, \( P^* \) is still 1-dimensional, but we have a one-dimensional set of equilibria \( C^a \) for the real allocations for a given consumer. This is in stark contrast to the traditional model, which typically ("generically") has only isolated equilibria – a one-dimensional set of equilibria. 36

Let us analyze the economics that give rise to those multiple equilibria.

**Economic forces leading to the one-dimensional continuum of equilibria** We say that good \( i \) is relatively less salient (or more obscure) than good \( j \) if its price is less salient, \( m_i < m_j \). A core effect is the following:

When the price level is high, the relative price of the non-salient good is high. To study this effect in its simplest form, take one representative agent (or, if the reader prefers, identical consumers \( a \) and \( b \), or several agents with same perceptions, homothetic preferences and endowments. It is easy to see that if \( p \) is an equilibrium price of the sparse economy, \( (Z (p) = 0) \), then any \( (p_k^{1/m_i})_{i=1...n} \) is an equilibrium price, with \( k > 0 \). So if there is just one \( p \in P^* \) with \( p_1 = 1 \), the set of equilibrium prices is:37

\[
P^* = \{ (p_k^{1/m_i})_{i=1...n} : k \in \mathbb{R}^+ \}
\]

(22)

This means that, “When the price level is high (\( k \) is high), the relative price of the obscure good is high.” When \( k \) is high, the consumer perceives the high price of the obscure good less, hence demands more of it. That increases the price of the obscure good. In general, if there were supply shocks, non-salient prices should move a lot, to compensate for inattention. 38

At the same time, in a representative agent model, there is just one equilibrium allocation, \( C^a = \{ c^a (p) \} \). The second effect relies on heterogeneity.

When the price level is high, the agent with the relatively higher endowment of the obscure good becomes relatively better off.

To illustrate this effect, we consider a polar case.

---

36In the traditional model, equilibria correspond to the intersection of offer curves. So naively one might think that the equilibrium set should be the intersection of the two 2-dimensional offer curves, so again a 2-dimensional surface. However, this is not the case here. The reason is the following. An intersection of the offer curves correspond to a \( p^a \) and \( p^* \) such that \( D^a (p^a) = \Omega - D^a (p^*) \). i.e. \( Z^a (p^a) + Z^a (p^*) = 0 \). But we could have \( p^a \neq p^* \) – though to have an equilibrium we need \( p^a = p^* \). The argument in the traditional model is the following: As \( p^a \cdot Z^a (p^a) = 0 \), we have \( p^a \cdot Z^a (p^*) = 0 \). We also have \( p^a \cdot Z^a (p^*) = 0 \). Either \( Z^a (p^a) = 0 \), or there is a real \( \xi \) such that \( p^a = \xi p^* \) (otherwise, \( \{p_a, p_b\} \) would span \( \mathbb{R}^2 \), and we would conclude that \( p \cdot Z^a = 0 \) for all \( p \), so \( Z^a (p^*) = 0 \). In the traditional model, \( Z \) is homogeneous of degree 0, so \( Z^a (p^a) = Z^a (\xi p^a) = 0 \), and \( Z^a (p^*) + Z^a (p^*) = 0 \). So the intersection of the offer curves is an equilibrium, at \( p^a \). However, in the sparse model, \( Z \) isn’t (typically) homogeneous of degree 0, so the intersection of offer curve isn’t necessarily (and typically) an equilibrium.

37If there is a set of such \( p \) with \( p_1 = 1 \), then the set of equilibria is the union of the corresponding sets.

38Gul, Pesendorfer and Strzalecki (2012) have a very different model where, to clear the market, prices need to move more than in the traditional model.
Figure 3: These Edgeworth boxes show competitive equilibria when both agents have Cobb-Douglas preferences. The left panel illustrates the traditional model with rational agents: there is just one equilibrium, $c^a = (1/2, 1/2)$. The right panel illustrates the situation when type $a$ is rational, and type $b$ is boundedly rational: there is a one-dimensional continuum of competitive equilibria. Agent’s $a$ share of the total endowment ($\omega^a$) is the same in both cases.

**Proposition 15** Assume that both agents have identical perceptions ($m^a = m^b$) and homothetic preferences ($c^a(p, w) = c^a(p, 1)w$). The set of equilibrium prices is given by (22), for some $\bar{p}$. The set of allocations becomes an interval: $C^a = [s^a_{\min}, s^a_{\max}] \omega$, where $s^a_{\min} = \min_i \frac{\omega^a_i}{\omega_i}$, $s^a_{\max} = \max_i \frac{\omega^a_i}{\omega_i}$. When the price level is high, the agent with the relatively higher endowment of the non-salient good (the good $i$ with lowest $m_i$) is relatively better off.

The emergence, stated here, of a continuum of real equilibria due to bounded rationality seems new. The most closely related may be those of Geanakoplos and Mas-Colell (1989), who analyze a two-period model with fewer assets than states, and show a (high-dimensional) continuum of equilibria (see also Magill and Quinzii (1992)). Those papers are quite far in their substance from the previous analysis (they study incomplete markets with full rationality, here I study complete markets with limited rationality). Still, they all stem from the fact that in the traditional model, prices have a one-dimensional indeterminacy.

**Pareto-Efficiency** It is discussed by the following Proposition.

**Proposition 16** (Efficiency of competitive equilibrium). Assume that competitive equilibria are interior. Then, an equilibrium is Pareto efficient if and only the perception of relative prices is identical across agents.

Hence, typically the equilibrium is not Pareto-efficient. However, the equilibrium is Pareto efficient if consumers have the same misperceptions. The argument is very simple: if consumers $a$ and $b$ have the same perceptions, then for two goods $i, j$, $u^a_{ci}/u^a_{cj} = p^m_i / p^m_j = u^b_{ci}/u^b_{cj}$, so that the ratio of marginal utilities is equalized across agents; there are no extra gains from
Figure 4: These Edgeworth boxes show competitive equilibria when both agents have Leontief preferences. The left panel illustrates the traditional model with rational agents: there is just one equilibrium, $c^a = (1/2, 1/2)$. The right panel illustrates the situation when type $a$ is rational, and type $b$ is boundedly rational: there is a one-dimensional continuum of competitive equilibria. Agent’s $a$ share of the total endowment ($\omega^a$) is the same in both cases.

trade. Hence, it is the heterogeneity in price misperceptions that create an inefficiency, rather than price misperceptions per se.

Some illustrative equilibria Note that in Proposition 15, we have a continuum of Pareto-efficient equilibria, as consumers have the same perceptions. Let us now consider the case of a rational consumer (consumer $a$) and a sparse consumer (consumer $b$), with $m_1 \neq m_2$ for that BR consumer. Let us draw some Edgeworth boxes, and the set of equilibrium consumptions $C^a$.

**Proposition 17** Suppose agent $a$ is rational, and the other agent is sparse with $m_1 \neq m_2$, and two goods. Consider the set $C^a$ of equilibrium allocations that agent $a$ obtains in competitive equilibria. It is equal to the rational agent’s offer curve. In particular, it is one-dimensional.

This example is a bit extreme, but illustrates the potential vastness of the set of equilibria. The economic intuition is as follows. Take an allocation $c^a$ a price $p$ such that $D^a(p) = c^a$. We want to see if consumer $B$ can be induced to consume the corresponding $c^s = \omega - c^a$. Consider prices $\mu p$, and take $m_1 < m_2$. Then when $\mu$ increases, the obscure good appears relatively cheaper, so the sparse consumer consumes more of it — arbitrarily more actually. That allows us to reach all the $c^s$ in the consumer’s budget curve corresponding to price $p$: \[
\{c \text{ s.t. } p \cdot c = p \cdot \omega^s\}.
\]

So, take a consumer $a$ who is Cobb-Douglas ($u(c_1, c_2) = \ln c_1 + \ln c_2$). Equilibrium set is his indifference curve, as shown in Figure 3, right panel. In contrast, in the traditional model, when the consumer $b$ has the same preferences, but is rational, there is just one equilibrium,
\[c^a = (1/2, 1/2).\] Figure 4 shows the same graphs, but for a Leontief consumers, i.e. have \(u(c_1, c_2) = \min_i c_i.\) 39 In the Cobb-Douglas case, across equilibria, a high consumption of good 1 corresponds to a low consumption of good 2. In the Leontief case, equilibria leads to a high (or low) consumption of both goods 1 and 2. This is because of the relative endowment effect mentioned above.

Finally, let us examine two sparse consumers, with heterogeneous attention. Differential attention leads to different allocation, even though agents have the same preferences and endowments. Some agents may pay no attention to some prices, as long as other agents do pay attention to that price.

**Example 6** Consider a case with identical preferences \(u(c_1, c_2) = v(c_1) + v(c_2),\) for some concave function \(v,\) and identical endowments, \(\omega^a = \omega^s = (1/2, 1/2),\) but asymmetric perceptions: Type \(a\) has attention \(m^a = (\alpha, \beta),\) and type \(b\) has attention \(m^b = (\beta, \alpha),\) with \(\alpha \neq \beta.\) Assume \(p^d = (1, 1)\) and \(\lim_{c \to 0} v'(c) = \infty.\) Then, the equilibrium set contains \(\{(c, 1-c), c \in (0, 1)\}\) and the set of equilibrium prices contains \(\{(p, p), p > 0\}\).

I conclude with some brief remarks.

### 5.3 Remarks on competitive equilibrium

**Are those multiple equilibria something real and important?** This question, actually, is much debated in macroeconomics. Standard macro deals with one equilibrium, conditioning on the price level (and expectations thereof). This is to some extent what we have here. Given a price level, there’s (locally) only one equilibrium, but the price level changes the equilibrium (when there are some frictions in the perception or posting of prices). This is akin to a (short-run) Philipps curve: when the price level goes up, the perceived wage goes up, and people supply more labor (this analysis is most naturally done in a production economy). Hence, we observe here the price-level dependent equilibria long theorized in macro (e.g. Lucas’ islands), but in the pristine and general universe of basic microeconomics.

**Equilibrium selection** What pins down the price? One approach is via a quantity theory of money \(M,\) e.g. chosen by the central bank: then the price \(p \in P^*\) is the (often unique) one that ensures that transactions have a value \(M (p \cdot \omega = M).\) Other approaches would rely on expectations in a dynamic model. It is best left to future research.

39 For the Cobb-Douglas case, the equilibrium set is curve is the set of \(c\) such that \(\prod_{i=1}^2 \left( \frac{2c}{\omega_i} - 1 \right) = 1\) and \(0 \leq c \leq \omega.\) For the Leontief, it’s the set of \(c\) such that \(\{c_1 = c_2\}\) and \(c_1 \in [\min_i \omega_i^a, \max_i \omega_i^a].\)
**Endogenous default price**  Endogeneizing the default price would be interesting. In the context of a static model, one might hypothesize that a good default price would be such that it is also an equilibrium price of the related sparse economy, i.e. such that $Z \left( p^d \mid p^d \right) = 0$ (rewrite excess demand as $Z \left( p \mid p^d \right)$). Then $p^d$ must be an equilibrium price of the underlying rational model. Alternatively, $p^d$ might be better thought of as some expectation of the price, given the past prices and recent shocks. The proper locus of the endogeneization of $p^d$ may be in an explicitly dynamic model, something I tackle in companion work.

**Generality of the effects**  Many effects simply stem from the fact that $c^a(p, w)$ is not homogenous of degree 0 in $(p, w)$, which is a form of “nominal illusion” – or really, a price illusion, as there is no “money”, simply a default price (Fehr and Tyran 2001, Shafir, Diamond and Tversky 1997 analyze money illusion). In many ways, many effects here do not depend on the specifics of money illusion. The “extra-dimensional offer curve” phenomenon and the one-dimensional continuum of equilibria can be expected to hold quite generally with nominal illusion. However, it is still useful to have a specific model, the sparsity model, to investigate it, as the economic intuition for many effects do depend on the spirit of the sparsity model (e.g. salient prices), and some conditions need to be specified (e.g. Proposition 14).

### 6 Discussion

#### 6.1 Two models that generate a noisy sparse max

Under some conditions, some models offer a noisy microfoundation for the sparse max, i.e. their representative agent version is an agent doing a sparse max. We emphasize the basic, quadratic case.

**Heterogeneous fixed costs**  Consider next an agent with an “all or nothing” attention: each $m_i$ is equal to 0 (no attention) or $\mu_i$ (full attention). This can be represented as a fixed cost for the penalty in (4), rather than a linear cost (the online appendix fills in the details related to scaling). That leads the agent to use the “hard thresholding” function $\tau^H(\mu, \kappa) = \mu \cdot 1_{|\mu| \geq |\kappa|}$, rather than the “soft thresholding” function $\tau$. One obtains a sparse max with that function $\tau^H$: we’d replace $\tau$ by $\tau^H$ throughout.

---

40 Function $c^a(p, w, p^d)$ is homogenous of degree 0, but function $c^a(p, w)$, with a fixed $p^d$, is generally not homogeneous of degree 0.
Some economists will prefer that fixed cost model. The following Proposition indicates that the sparse max is, in some sense, the “representative agent equivalent” of many agents with heterogeneous fixed costs.  

Proposition 18 (Fixed costs models as a microfoundation of sparse max in the basic case). Suppose that agents use fixed costs \( k \), and the distribution of \( k \)'s is \( f(k | \kappa) = \kappa_1 k_\kappa/k^2 \). Then, the soft-threshold \( \tau \) is the average values of the hard-threshold \( \tau^H \): \( \tau(\mu, \kappa) = \int_0^\infty \tau^H(\mu, k) f(k | \kappa) dk \). Also, in quadratic problems with a one-dimensional action, the average behavior of those “fixed cost” agents is described by the sparse max.

Noisy signals Here is a version of this idea with the noisy signals model.

Proposition 19 (Signal-extraction models as a noisy microfoundation of sparse max in the basic case). Consider a model with quadratic loss (3), where the signals are \( s_i = x_i + \varepsilon_i \), with \( \varepsilon_i \) noises with relative precision \( T_i = \text{var}(x_i) / \text{var}(\varepsilon_i) \), with \( (x, \varepsilon) \) Gaussian with uncorrelated components. The agent: (i) decides on signal precision \( T_i \) (ii) receives signals \( s = (s_1, ..., s_n) \), and then (iii) takes action \( a(s) \), to maximize: \( \max_{T_i \geq 0} \max_{a(s)} \mathbb{E}[u(a(s), \mu, x) | s] - \frac{K^2}{2} \sum_i T_i \). Then, for a given \( x \), averaging over the signals, the optimal action \( a(s) \) is the sparse max \( \alpha_{ec \ante}^s(x) \):

\[
\mathbb{E}[a(s) | x] = \alpha_{ec \ante}^s(x) = \sum_i \tau \left( \mu_i, \frac{K}{\sigma_{x_i}} \right) x_i
\]

This assumption that the cost of precision is linear is quite common (e.g. Geanakoplos and Milgrom 1991). Precision is proportional to the number of i.i.d. signals collected, and the cost can be proportional to the number.

Hence, economists who think that deep down, agents perform Bayesian updated with costly signals (which is simply a hypothesis) have a way to understand the sparse max in the following way: The sparse max is the representative agent version of a model with noisy signals.

I should say that this holds only under definite assumptions. In fact, it is extremely hard to go beyond the Gaussian, Linear-Quadratic with that noisy signal + Bayes paradigm (Matejka and Sims 2010).

6.2 Links with Themes of the Literature

Sparsity another line of attack on the polymorphous problem of confusion, inattention, simplification, and bounded rationality (Kahneman 2003). This paper is best viewed as a com-
plement rather than a substitute for existing models. For instance, there is a vast literature on learning (Sargent 1993, Fudenberg and Levine 1998) that sometimes generates a host of stylized facts because agents may not set up their models optimally (Fuster, Laibson, and Mendel 2010). One could imagine joining those literatures in a model of “sparse learning.” Some of the most active themes are the following.42

Inattention and information acquisition. This paper is related to the literature on modeling inattention (Dellavigna 2009, Veldkamp 2011). One strand uses fixed costs, paid over time (Grossman and Laroque 1990, Gabaix and Laibson 2002, Mankiw and Reis 2002, Reis 2006, Abel Eberly and Panageas 2010, Schwarzstein 2012). Those models are instructive, but they quickly become hard to work with as the number of variables increases. I have argued that fixed costs lead to quite complex papers. Also, those papers all require a time dimension – it is hard to use them to study a static problem, such as the basic consumption problem (1).

Closer to this paper, perhaps, is the theory present in the NBER working paper version of Chetty, Looney and Kroft (1999), but deleted from the published version. They study a consumption problem with two goods, where the DM may not think about the tax. Attention is modeled as paying a fixed cost, or a distribution of those costs. This paper proposes a general sparse max with constraints. Also, it derives the whole of basic consumer and equilibrium theory with several goods.

An influential proposal made by Sims (2003) is to use an entropy-based penalty for the cost of information; this literature is progressing impressively (e.g., Maćkowiak and Wiederholt 2011, 2012, Woodford 2012). It has the advantage of a nice mathematical foundation. However, it leads to non-deterministic models (agents take stochastic decisions), and the modeling is very complex when it goes beyond the linear-Gaussian case: the solutions require a computer to be solved, and there’s no closed form (Matejka and Sims 2010). As a result, no one has (yet) been able to work out the basic consumption problem (1), much less derive its implications for basic consumption and equilibrium theory.43 In contrast, this paper’s techniques allow for it. The sparse max presents some important differences of substance discussed in section 2.3. Still a reader might ask, “why not do Bayesian signal extraction all the time, with quadratic approximations, and average over the agent to look at the average outcomes?” One answer is that it may be better to directly work with the sparse

42I omit many models here, in particular “process” models, e.g., Bolton and Faure-Grimaud (2009), Gabaix, Laibson, Moloche, and Weinberg (2006). They are instructive but yield somewhat complex mappings between situations and outcomes. Automata models (Rubinstein 1998) lead to complex models that are not (yet) ready to handle basic problems such as the consumption choice problem.

43See however Maćkowiak and Wiederholt 2011 for a macro model, which uses quadratic approximations and source-specific costs, thus way departing from the Sims doctrine.
max, without signals, rather than always do the detour via an allegorical signal extraction with hypothesized Gaussian signal. In addition, the sparse max handles constraints easily. Constraints are much harder to handle by limited-information papers.

Near-rational approach. The near-rational approach of Akerlof and Yellen (1985) is based on the premise that agents will tolerate decisions that make them lose some ε utility, and still proves useful for empirical work (e.g., Chetty 2012). However, it does not yield definite predictions about precisely what actions the decision maker will take. The present paper does generate precise action.

Limited understanding of strategic interactions. In several types of models, the BR comes from the interactions between the decision maker and other players, (see Camerer, Ho, and Chong (2004), Eyster and Rabin’s (2005), Jéhiel’s (2005), Crawford and Iribari (2007), Crawford, Costa-Gomes, and Iribari (2013)). These models prove very useful for capturing naïveté about strategic interactions. However, in a single-person context, they (typically) model the decision maker as fully rational.

Uncertainty aversion and concern for robustness. Hansen and Sargent (2007) model that agents understand that they do not know the right model and have a concern for “robustness”. In contrast, in the present model, the decision maker is biased towards simplicity, not robustness towards bad events. It would be interesting to study the interaction between taste for sparsity and concern for robustness.

It may be interesting to note the Sims framework is based on Shannon’s information theory of the 1940s. The Hansen-Sargent framework is influenced by the engineering literature of the 1970s. The present framework is inspired by the sparsity-based literature of the 1990s–2000s (Tibshirani 1996, Candès and Tao 2006, Donoho 2006, Mallat 2009), which shows that sparsity has many properties of tractability and near-optimality. The present paper is the first paper in economic theory to use the recent sparsity-based literature from statistics.

7 Conclusion

This paper proposes a tractable enrichment of the traditional “max” operator, with some boundedly rational features: the “sparse max” operator. This formulation is quite tractable.

44 Relatedly, an interesting literature studies BR in organizations (e.g., Radner and Van Zandt 2001), and aims at predictions on the level of large organizations rather than individual decision making. See also Madarász and Prat (2010) for a recent interesting advance in BR in a strategic context.

45 For instance, somewhat miraculously, one can do a regression with fewer observations than regressors (like in genetics, or perhaps growth empirics) by assuming that the number of non-zero regressors is sparse and using an $\ell_1$ penalty for sparsity (see Belloni and Chernozhukov 2010), as in $\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta' x_i)^2 + \lambda \|\beta\|_1$. 

34
At the same time, it arguably features some psychological realism. We all simplify reality, and this model represents one way to do that. It features the “anchoring with partial adjustment” mechanism, which is a pervasive feature of psychology. It gives a quantitative version of when the adjustment will be large or small.

The simplicity of the core model allows for the formulation of a sparse, limited attention version of important building blocks of economics: the pure theory of consumer and producer behavior. We could bring behavioral enrichment to venerable and often-used concepts such as Marshallian and Hicksian demand, Edgeworth boxes, and competitive equilibrium sets. Some surprises have emerged. The model allowed us to understand what is robust and non-robust in basic microeconomics.

This paper argued that other models with a more traditional form (e.g. extraction of noisy signals) might lead to fairly similar features, but would be quite intractable. The sparse max allows us to explore features of economic life that hopefully apply to other models. It does it with relatively little effort.

No doubt, the model could and should be greatly enriched. As a work in progress, I extend the model to include multi-agent models and dynamic programming (Gabaix 2012a,b). In addition, the model is silent about some difficult operations such as Bayesian (or non-Bayesian) updating and learning (see Gennaioli and Shleifer 2010 for recent progress in that direction), and memory management (Mullainathan 2002). As is, it is silent about mental accounts (Thaler 1985), though that extension appears to be within reach.

Despite these current limitations, the sparse max might be a useful tool for thinking about the impact of bounded rationality in economic situations.

**Appendix: Proofs**

More technical proofs are in the online appendix. We start with the following lemma.

**Lemma 1** For $\kappa \geq 0$, the solution of $\min_m \frac{1}{2} (m - \mu)^2 + \kappa |m - m_*|$ is $m = m_* + \tau (\mu - m_*, \kappa)$ where $\tau$ is the anchoring and adjustment function given in (6).

**Proof.** By shifting $m \rightarrow m - m_*$, $\mu \rightarrow \mu - m_*$, it is enough to consider the case $m_* = 0$. The f.o.c. is: $m - \mu + \kappa \text{sign}(m) = 0$, where $\text{sign}(m)$ is the sign of $m$ if $m \neq 0$, and an indefinite quantity in $[-1, 1]$ if $m = 0$. When $\mu > 0$ and $m > 0$, that yields $m = \mu - \kappa$. Consistency with $m > 0$ requires $\mu > \kappa$. If $0 \leq \mu \leq \kappa$, the solution is $m = 0$: the loss function is decreasing for $m < 0$, and increasing for $m > 0$. That yields $m = \tau (\mu, \kappa)$ for $\mu \geq 0$. The case $\mu < 0$ is symmetrical. ■
Proof of Proposition 1  We have \( u_a(a, m, x) = v_a(a, m_1x_1, ..., m_nx_n) \), \( u_{aa} = v_{aa} \) and \( u_{am_i} = v_{ai}x_i \), where \( v_{ai} \) is the cross-derivative of \( v(a, y_1, ..., y_n) \) with respect to (w.r.t.) \( a \) and \( y_i \). Hence, the problem becomes:

\[
m^* = \arg \min_m \sum_{i,j} \frac{1}{2} (m_i - \mu_i) (m_j - \mu_j) \left( -v_{aa}^{-1} \right) v_{ai}^2 \mathbb{E}_P [x_i | x_j] + \kappa \sum_i |m_i| |v_{ai}| \mathbb{E}_P \left[ x_i^2 \eta_{a_i}^{1/2} \right]
\]

\[
= \arg \min_m \sum_i \frac{1}{2} (m_i - \mu_i)^2 \left( -v_{aa}^{-1} \right) v_{ai}^2 \sigma_{x_i}^2 + \kappa \sum_i |m_i| |v_{ai}| \sigma_{x_i} \sigma_a
\]

as under probability \( P \), \( x_i \) and \( x_j \) are uncorrelated for \( i \neq j \). We have \( n \) decoupled problems: for \( i = 1...n \),

\[
m_i^* = \arg \min_{m_i} \frac{1}{2} (m_i - \mu_i)^2 \left( -v_{aa}^{-1} \right) v_{ai}^2 \sigma_{x_i}^2 + \kappa |m_i| |v_{ai}| \sigma_{x_i} \sigma_a
\]

Let use the notations \( \sigma_a := \mathbb{E}_P \left[ \eta_a^{1/2} \right] \) and \( \kappa_i := \frac{\kappa \sigma_a}{|v_{aa}v_{ai}| \sigma_{x_i}} \) (which can be infinite). As \( v_{aa} < 0 \), the problem becomes:

\[
m_i^* = \arg \min_{m_i} \frac{1}{2} (m_i - \mu_i)^2 + \kappa_i |m_i|
\]

(24)

Using Lemma 1 just before this proof, this leads to \( m_i^* = \tau(\mu_i, \kappa_i) \). As \( a^d(x) = \arg \max_a v(a, \mu_i, x_i) \), we have \( v_a(a, \mu_i, x_i) = 0 \), and the implicit function theorem gives: \( v_{aa} \partial a / \partial x_i + v_{ai} \mu_i = 0 \), and finally \( \partial a / \partial x_i = -\mu_i v_{aa}^{-1} v_{ai} \). Hence, \( \kappa_i = \left| \frac{\mu_i \sigma_a}{\sigma_{x_i} \partial a / \partial x_i} \right| \), and

\[
m_i^* = \tau(\mu_i, \kappa_i) = \tau \left( \mu_i, \mu_i \left| \frac{\kappa \sigma_a}{\sigma_{x_i} \partial a / \partial x_i} \right| \right) = \mu_i \tau \left( 1, \frac{\kappa \sigma_a}{\sigma_{x_i} \partial a / \partial x_i} \right),
\]

using \( \tau(\mu, \kappa) = \tau(\mu, -\kappa) \) and the fact that \( \tau \) is homogenous of degree 1 in \((\mu, \kappa)\).

Proof of Proposition 2  We apply Proposition 1 to the case \( \sigma_{x_i} = |x_i| \). That leads to:

\[
m_i^* x_i = \mu_i \tau \left( 1, \frac{\kappa \sigma_a}{x_i \partial a / \partial x_i} \right) x_i = \mu_i \tau \left( x_i, \frac{\kappa \sigma_a}{\partial a / \partial x_i} \right) = \mu_i x_i^s
\]

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

Proof of Proposition 3  We shall first use a Lemma.

Lemma 2  Consider the case with 1 constraint. In the setup of Definition 2, Step 2, utility \( u(a(\lambda), m^*, x) \) is weakly decreasing in \( \lambda \). If \( \lambda^* > 0 \), then \( B(a(\lambda^*)) = 0 \).

Proof. \( L \) is concave in \( a \). Let us assume it is strictly concave in \( a \), as the general result obtains by continuity. We have \( L_a(a(\lambda), \lambda) = 0 \), so \( L_{aa}a'(\lambda) + L_{\lambda \lambda} = 0 \), i.e. \( (L_{\lambda \lambda} = B_a) \),
\(a'(\lambda) = -L_{aa}^{-1}B_a\). Hence, utility \(V(\lambda) = u(a(\lambda))\) satisfies (observing that \(u_a + \lambda B_a = 0\),
\[V'(\lambda) = u_a a'(\lambda) = \lambda B_a' L_{aa}^{-1}B_a \leq 0\) as \(L_{aa}^{-1}\) is negative definite, and \(\lambda\) is nonnegative. 

The Lagrangeans associated with the two problems are (adding a constant strategically, and calling \(\Lambda\) the Lagrange multiplier for the second problem) \(L(a, m, \lambda) = u(a, m, x) - \tilde{u} + \lambda(\tilde{w} - w(a, m, x))\) and \(M(a, m, \Lambda) = \tilde{w} - w(a, m, x) + \Lambda(u(a, m, x) - \tilde{u})\). Hence

\[L(a, m, \lambda) = \lambda M(a, m, 1/\lambda)\]  
(25)

which is a sort of duality relation. Let us follow Definition 2. In step 1.i, we obtain \(\Lambda^d = 1/\lambda^d\).

Next, see step 1.ii. Consider \(m^\ast\). It comes from \(\text{smax}_a L(a, m, \lambda^d)\), which yields the same \(m^\ast\) as \(\text{smax}_a \lambda^d M(a, m, \Lambda^d)\), as the two functions are the same. Hence, in step 1 applied to \(L\) or \(M\), the two \(m^\ast\) selected are the same.

Move on to step 2. We define \(a^L(\lambda) := \text{arg max}_a L(a, m^\ast, \lambda)\) and \(a^M(\Lambda) := \text{arg max}_a M(a, m^\ast, \Lambda)\). Then (25) implies: \(a^M(\lambda) = a^L(1/\lambda)\).

Finally, as \(\lambda^\ast > 0\), Lemma 2 shows that the constraint binds, i.e. \(w(a^L(\lambda^\ast), \mu) = \tilde{w}\). It yields a utility \(u(a^L(\lambda^\ast), \mu) = \tilde{w}\). But this is the same solution as for problem \(M\), which seeks a solution \(a^M(\Lambda^\ast)\), that should solve \(u(a^M(\Lambda^\ast), \mu) = \tilde{w}\) and also yield a cost \(w(a^M(\Lambda^\ast), \mu) = \tilde{w}\). As the two problems are the same, we get \(a^L(\lambda^\ast) = a^M(\Lambda^\ast)\) and \(\Lambda^\ast = 1/\lambda^\ast\).

**Proof of Proposition 4** We follow Definition 2. The Lagrangian is \(L(c, m, x, \lambda) = u(c) + \lambda(w - p^m \cdot c)\), with \(p = p^d + x\) and \(p^m_i = p^d_i + m_ix_i\). We just need step 2 here. The function \(c^\lambda = \text{arg max}_c L(c, m^\ast, x, \lambda)\) satisfies \(u'(c(\lambda)) = \lambda p^m\) i.e. \(c(\lambda) = u''(1) (\lambda p^m)\). Hence using the solution \(\lambda^\ast\) satisfies the budget constraint \(p \cdot c(\lambda^\ast) = w\) (Lemma 2 shows formally it binds), and the chosen consumption is \(c^\ast = c(\lambda^\ast)\). Calling \(w' = p^m \cdot c^\ast\), we have that \(c^\ast = c^\ast (p^m, w')\), as it satisfies \(p^m \cdot c^\ast = w\) and \(u'(c^\ast) = \lambda p^m\).

When demand is linear, the equation for \(w'\) writes: \(w = p \cdot c^\ast (p^m, w') = p \cdot c^\ast (p^m, 1) w'\), so \(w' = w/p \cdot c^\ast (p^m, 1)\), and

\[c^\ast (p, w) = c^\ast (p^m, w') = c^\ast (p^m, 1) w' = \frac{c^\ast (p^m, 1)}{p \cdot c^\ast (p^m, 1)} w.\]

**Proof of Proposition 5** The step 1 problem is: (with \(\lambda = \lambda^d\))

\[
\min_m \sum_i (m_i - 1)^2 \sigma_{pi}^2 (-u''(1))_{ii} \lambda^2 + \kappa \sum_i |m_i| \lambda |\eta_i \sigma_{pi}| 
\]
so \( m_i = \tau (1, \kappa_i) \) with

\[
\kappa_i = \frac{\lambda |\eta_{i1}| \sigma_{pj}}{\lambda^2 \sigma_p^2 (-u^u)^{-1}} = \kappa \frac{|\eta_{i1}/c_i^d| \cdot c_i^d}{\lambda \sigma_{pj} (-u^u -1)} = \kappa \frac{|\eta_{i1}/c_i^d| \cdot c_i^d}{u_{ci} \cdot \frac{\sigma_{pj}}{p_i^d} (-u^u -1)} \text{ using } \lambda p_i^d = u_{ci}
\]

\[
= \kappa \frac{|\eta_{i1}/c_i^d|}{\psi_{i1} \sigma_{pj} / p_i^d} \text{ using } \psi_i := u_{ci} (-u^u -1) \cdot c_i^d.
\]

**Proof of Proposition 6** We have: \( c^*(p, w) = c^* ((p_i^d + m_i (p_i - p_i^d))_{i=1...n}, w' (p)) \).
We differentiate w.r.t. \( p_j \), and the point \( p = p_i^d \):

\[
\frac{\partial c^s}{\partial p_j} = \frac{\partial c^r}{\partial p_j} \cdot m_j + \frac{\partial c^r}{\partial w} \frac{\partial w'}{\partial p_j}
\]

(26)

Proposition 4 implies \( p \cdot c^r (p^m, w' (p)) = w \), and differentiating w.r.t. \( p_j \)

\[
0 = c^r_j + p \cdot \frac{\partial c^r}{\partial p_j} m_j + p \cdot \frac{\partial c^r}{\partial w} \frac{\partial w'}{\partial p_j}
\]

As for the rational demand, \( p \cdot c^r (p, w) = w \), we have (respectively differentiating w.r.t. \( p_j \) and \( w \)) \( c^r_j + p \cdot \frac{\partial c^r}{\partial p_j} m_j + p \cdot \frac{\partial c^r}{\partial w} \frac{\partial w'}{\partial p_j} = 0 \) and \( p \cdot \frac{\partial c^r}{\partial w} = 1 \), so \( 0 = c^r_j - c^r_j m_j + \frac{\partial w'}{\partial p_j} \), and \( \frac{\partial w'}{\partial p_j} = (m_j - 1) c^r_j \).
Finally (26) gives:

\[
\frac{\partial c^s}{\partial p_j} = \frac{\partial c^r}{\partial p_j} \cdot m_j + \frac{\partial c^r}{\partial w} \cdot (m_j - 1) c^r_j.
\]

**Proof of Proposition 7** We can now evaluate the Slutsky term (15) at \( p = p_i^d \):

\[
S^s_{ij} = \frac{\partial c_i^s}{\partial p_j} + \frac{\partial c_i^s}{\partial w} c_j^s = \frac{\partial c_i^s}{\partial p_j} + \frac{\partial c_i^r}{\partial w} c_j^r, \text{ as for all } w, \ c^s_j (p_i^d, w) = c^r_j (p_i^d, w)
\]

\[
= \frac{\partial c_i^r}{\partial p_j} \cdot m_j - \frac{\partial c_i^r}{\partial w} c_j^r (1 - m_j) + \frac{\partial c_i^r}{\partial w} c_j^r, \text{ by (16)}
\]

\[
= \left( \frac{\partial c_i^r}{\partial p_j} + \frac{\partial c_i^r}{\partial w} c_j^r \right) m_j = S^s_{ij} m_j, \text{ by (15)}
\]

**Proof of Proposition 9** Call \( f (x) = x^t S^r x = x^t S^r M \). Then, \( f' (x) = S^r M x + x^t S^r M \), and as \( p^d S^r = 0 \), \( f' (p^d) = S^r M p^d = 0 \), while \( f (p^d) = 0 \). As \( q \neq 0 \), \( f' (p^d) q = q^t q > 0 \). Hence, a vector \( x = p^d + \varepsilon q \) satisfies \( f (x) > 0 \) for \( \varepsilon > 0 \) small enough (as \( f (x (\varepsilon)) = q^t q \varepsilon + O (\varepsilon^2) \)).

The part on \( \delta p \cdot \delta c^s \leq 0 \) comes from: \( \delta c^s = S^r M \delta p \), so \( \delta p' M \delta c^s = \delta p' M S^r M \delta p \leq 0 \). □

Here is more analytics for the car + gas example. Goods 1, 2, 3 are car, gas and food. The default price is \( p_i^d = (1, 1, 2) \), and expenditure shares are \((1/4, 1/4, 1/2)\). The Slutsky matrix is:

\[
S^r = \begin{pmatrix}
-1 & -1 & 1 \\
-1 & -1 & 1 \\
1 & 1 & -2
\end{pmatrix}
\]

times a constant that we normalize to 1 (verify that the matrix
does satisfy $S^r p^d = 0$. This can be rationalized by a utility function $\ln \left( \min (c_1, c_2) \right) + \ln c_3$: car and gas are both needed for transportation. Hence, for “transportation”, the price is $p_1 + p_2$, the price of car plus gas.

Let’s say that attention is $m = (1, 0, 1)$, i.e. people pay attention to the car price and food, but not to gas: this is meant to capture lower attention to energy consumption, when the agent buys the car. Suppose now that there is a decrease in the car price, and an increase in the price of gas, say $\delta p = (-1, 2, 0)$. The rational agent sees that the total price of transportation has increased by $-1 + 2 = 1$, so he should consume less transportation – less car and gas. However, a sparse agent perceives a price $\delta p^m = (m_i \delta p_i)_{i=1..3} = (-1, 0, 0)$: he rejoices as price of a car has decreased, but he does not see the increase of the gas price. He thinks that the price of transportation has decreased, so consumes more of car+gas – the price of transportation has truly increased, but he consumes more of it. Mathematically, the $\delta c^r = S^r \delta p = (-1, -1, 1)$, while $\delta c^s = S^r \delta p^m = (1, 1, -1)$: so $\delta p \cdot \delta c^r = -1$, while $\delta p \cdot \delta c^s = 1$.

**Proof of Proposition 11** We have: $e^s (p, v) = p \cdot h^s (p, v) = p \cdot h^r (p^m, v)$. As the default, $p = p^d = p^m$, hence $e^s (p^d) = e^r (p^d)$ and $e^s_p (p^d) = p^d \cdot h^r (p^d, v) + h^r_p (p^d, v) = e^r_p (p^d, v)$.

We have $e^s_p = h^s + p \cdot h^s_p (p)$ and, taking the derivative at $p^d$,

$$ e^s_{pp} = h^s + h^r_p + p \cdot h^s_{pp} = h^r_p M + M h^s_p + p \cdot h^s_{pp} $$

(Here, the expression $p \cdot h^s_{pp}$ is understood as the matrix with terms $p \cdot h^s_{pp,ij}$). As

$$ h^s = h^r \left( (I - M) p^d + Mp \right), \quad h^s_p = h^r_p \left( (I - M) p^d + Mp \right) \cdot M $$

$$ h^s_{pp} = M h^r_{pp} \cdot M $$

and as $h^r (\lambda p) = h^r (p)$ for all $\lambda > 0$, we have (differentiating w.r.t. $p$), $\lambda h^r_p (\lambda p) = h^r_p (p)$, and (differentiating w.r.t. $\lambda$ at $\lambda = 1$), $h^r_p + h^r_{pp} \cdot p = 0$, so

$$ p \cdot h^s_{pp} = M \left( p \cdot h^r_{pp} \right) \cdot M = -M h^r_p M = -Me^r_{pp} M \quad \text{where} \quad h^r_p = e^r_{pp} \quad \text{is standard} $$

$$ e^s_{pp} = e^r_{pp} M + Me^r_{pp} - Me^r_{pp} M $$

This expression can be written coordinate-wise: $e^s_{p,p,j} = e^r_{p,p,j} (m_i + m_j - m_i m_j)$, and $e^s_{pp} = e^r_{pp} - (I - M) e^r_{pp} (I - M)$. The last expression can be understood as: the sparse expenditure function (in its second derivative) is the rational one, plus an extra cost due to the lack of
perfect optimization. That cost is \((I - M) e_{pp}^r (I - M)\).

Relation \(e_{p_ip_j}^s = S_{ij}^s + S_{ji}^s - \frac{S_{ij}^s S_{ji}^s}{S_{ij}^s}\) follows immediately from that, and Proposition 7, which implies \(S_{ij}^s = e_{ij}^s m_j\).

**Proof of Proposition 12** The fact that \(c^s (p^d, w) = c^r (p^d, w)\) implies that \(V^s (p^d, w) = V^r (p^d, w)\), hence \(V, V_p, V_{ww}\) is identical for both models. To go further, we use Proposition 32 in the online appendix. The fact that \(V_p^s\) is the same for both models comes from (70). Then, apply (71) to the shift \(s = (w, p)\) gives:

\[
V_{wp}^s - V_{wp} = (c^s_w - c_w) L_{cc} (c^s_w - c_w) = 0
\]

where we use \(c^s_w - c_w = 0\), which comes from \(c^s (p^d, w) = c^r (p^d, w)\) for all \(w\).

For \(V_{pp}^s\) we rely on duality: we have, for all \(p\) and \(w\), and for both the behavioral and traditional model: \(V (p, e (p, w)) = w\), so

\[
V_p (p, e (p, w)) + V_w (p, e (p, w)) \cdot e_p (p, w) = 0
\]

and taking the derivative at \(p = p^d\),

\[
V_{pp}^s + 2V_{pw}^s e_p + e_p V_{ww}^s e_p + V_w e_{pp}^s = 0
\]

This is true for both the \(B\) and \(R\) models: Also, note that \(e_p, V_w, V_{ww}\) and \(V_{wp}\) are the same for both models. Hence

\[
V_{pp}^s + 2V_{pw}^s e_p + e_p V_{ww}^s e_p + V_w e_{pp}^s = 0
\]

The common values are left without a superscript, e.g. \(e_p = c_p^r = c_p^s\). Subtracting those two equations yields: \(V_{pp}^s - V_{pp}^r = -V_w (e_{pp}^s - e_{pp}^r)\).

Finally, the expression in \(S\) comes from: \(e_{pp}^s = c_{pp}^r - (I - M) e_{pp}^r (I - M)\).

**Proof of the Proposition 13** Shepard’s lemma. We have: \(e^s (p, v) = p \cdot h^r (p^m, v)\), so

\[
e_{p_j}^s = h_j^r + p \cdot h_{p_j}^r (p^m, v) m_j = h_j^r + (p - p^m) \cdot h_{p_j}^r (p^m, v) m_j
\]

as \(h^r (p', v)\) is homogenous of degree 0 in \(p'\), so that \(p' \cdot h_{p_j}^r (p', v) = 0\), and \(p^m \cdot h_{p_j}^r (p^m, v) = 0\).

**Roy’s identity.** Let us first calculate \(\frac{\partial e^r}{\partial w}\) and \(\frac{\partial e^r}{\partial p_j}\). Proposition 4 gives \(p \cdot c^r (p^m, w') = w\), so \(p \cdot c^r \frac{\partial e^r}{\partial w} = 1\), and \(\frac{\partial e^r}{\partial w} = \frac{1}{p c_{w_j}^r (p^m, w_j)}\). Also, taking the derivative w.r.t. \(p_j\), \(c_j^r + p \cdot c_{p_j}^r m_j + \)
\[ p \cdot c^r_{w'} \frac{\partial w'}{\partial p_j} = 0, \text{ so} \]
\[ \frac{\partial w'}{\partial p_j} = \frac{-c^r_j - p \cdot c^r_{p_j}(p^m, w') m_j}{p \cdot c^r_{w'}(p^m, w')} \]

Now, note that \( c^s_w = c^r_{w'}(p^m, w') \frac{\partial w'}{\partial w} \) (from eq. 13) implies:
\[ c^s_w = \frac{c^r_{w'}(p^m, w')}{p \cdot c^r_{w'}(p^m, w')} \]

Next, given \( V^s(p, w) = u(c^s(p, w)) \) we have: \( V^s_w = u'(c^s) \cdot c^s_w \) hence \( V^s_w = \lambda p^m \cdot \frac{c^r_{w'}(p^m, w')}{p \cdot c^r_{w'}(p^m, w')} \), and as \( q \cdot c^r_{w'}(q, w') = 1 \) (which comes from \( q' \cdot c(q', w') = w' \)),
\[ V^s_w = \frac{\lambda}{p \cdot c^r_{w'}(p^m, w')} \] (27)

Finally, \( c^s(p, w) = c^r(p^m, w') \) gives:
\[ c^s_{p_j} = c^r_{p_j}(p, w') m_j + c^r_{w'}(p^m, w') \frac{\partial w'}{\partial p_j} \]
\[ c^s_{p_j} = c^r_{p_j}(p, w') m_j - c^r_{w'}(p^m, w') \frac{c^r_j + p \cdot c^r_{p_j}(p^m, w') m_j}{p \cdot c^r_{w'}(p^m, w')} \] (28)

so
\[ V^s_{p_j} = u'(c) \cdot c^s_{p_j}(p, w) \]
\[ = \lambda p^m \cdot \left[ c^r_{p_j}(p, w') m_j - c^r_{w'}(p^m, w') \frac{c^r_j + p \cdot c^r_{p_j}(p^m, w') m_j}{p \cdot c^r_{w'}(p^m, w')} \right] \]
\[ = \lambda p^m \cdot c^r_{p_j}(p, w') m_j - \lambda \frac{c^r_j + p \cdot c^r_{p_j}(p^m, w') m_j}{p \cdot c^r_{w'}(p^m, w')} \]
\[ = \lambda p^m \cdot c^r_{p_j}(p, w') m_j - \left( c^r_j + p \cdot c^r_{p_j}(p^m, w') m_j \right) V^s_w \]

Hence
\[ \frac{V^s_{p_j}}{V^s_w} = p^m \cdot c^r_{p_j}(p, w') m_j \times p \cdot c^r_{w'}(p^m, w') - c^r_j - p \cdot c^r_{p_j}(p^m, w') m_j \]
\[ = -c^r_j + \left[ p^m \cdot c^r_{p_j}(p, w') \times p \cdot c^r_{w'}(p^m, w') - p \cdot c^r_{p_j}(p^m, w') \right] m_j \]
\[ = -c^r_j + \left[ (p \cdot c^r_{w'}(p^m, w')) p^m - p \right] c^r_{p_j}(p^m, w') m_j \]

**Proof of Proposition 14**  It is enough to show that for small \( \delta c \), we can find a small \( \delta p \) such that \( D(p + \delta p) = c + \delta c \), i.e. there is a \( \delta \lambda \) such that \( \Phi(p + \delta p, \lambda + \delta \lambda, c + \delta c) = 0 \),

41
with
\[
\Phi(p, \lambda, c) = \left( u_c(c) - \lambda p^m(p) \right. \left. \quad p \cdot (c - \omega) \right)
\]

It is enough that \( \partial \Phi \) (the derivative w.r.t. \( (p, \lambda) \)) has rank \( n + 1 \). We calculate \( \partial \Phi = \left( \begin{array}{cc} \frac{\partial p^m}{\partial p} & -p^m \\ c - \omega & 0 \end{array} \right) \). To see if \( \partial \Phi \) has rank \( n + 1 \), we show that, given arbitrary \( \delta u, \delta b \), we can find \( \delta p, \delta \lambda \) such that \( \partial \Phi(\delta p) = (\delta \omega) \), i.e. \( \left( \begin{array}{cc} \frac{\partial p^m}{\partial p} & -p^m \\ c - \omega & 0 \end{array} \right) \) \( (\delta p) = (\delta \omega) \). The first equation gives \( \delta p = (\frac{\partial p^m}{\partial p})^{-1} (\delta u + p^m \delta \lambda) \), and the second equation gives \( (c - \omega) \left( \frac{\partial p^m}{\partial p} \right)^{-1} \right) p^m \neq 0.

**Example 5**  Call \( c = (c_1, c_2) \) the candidate consumption. Suppose that \( (c_1, c_2) \neq (\omega_1, \omega_2) \). We’re looking to see if there are prices \( p \) such that \( D(p) = c \), i.e. if there is a \( \lambda > 0 \) and a \( p \) such that: \( u_{c_1}(c) = \lambda p^m_i \) and \( p \cdot (\omega - c) = 0 \). Given \( \lambda \), we just set \( p_i = (u_{c_i}(c) / \lambda)^{1/m_i} \), so the equation to solve is \( f(\lambda) = 0 \) where \( f(\lambda) = p(\lambda) \cdot (\omega - c) \), i.e. \( f(\lambda) = \sum_i \lambda^{-1/m_i} u_{c_i}(c)^{1/m_i} (\omega_i - c_i) \).

Suppose, without loss of generality, that \( m_1 > m_2 \). Then as \( \lambda \to 0 \), \( f(\lambda) \sim \lambda^{-1/m_1} u_{c_1}(c)^{1/m_1} (\omega_1 - c_1) \), and \( \text{sign} \ f(\lambda) \) = \( \text{sign} \ (\omega_1 - c_1) \). Likewise, as \( \lambda \to \infty \), \( \text{sign} \ f(\lambda) \) = \( \text{sign} \ (\omega_2 - c_2) \). By lack of domination, \( \omega_1 - c_1 \) and \( \omega_2 - c_2 \) have opposite signs. Hence, \( f(\lambda) \) has opposite signs in near 0 and near infinity. By the intermediate value theorem, there is a \( \lambda^* \) such that \( f(\lambda^*) = 0 \).

The proof suggests a way this property might fail with more than two goods. Order them such that \( m_1 > \ldots > m_n \). Then, if \( \omega_1 - c_1 \) and \( \omega_n - c_n \) have the same sign, then the \( \lambda \) might fail to exist. On the other hand, if \( \omega_1 - c_1 \) and \( \omega_n - c_n \) have different signs, then \( c \) is on the offer curve.

**Proof of Proposition 15**  Given \( p \), the share of total income that goes to consumer \( a \) is \( s^a = \frac{p a^a}{p \omega} \). We clearly have \( s^a \in [s_{\min}^a, s_{\min}^a] \). With two goods, say that \( \frac{a^1}{\omega_1} \geq \frac{a^2}{\omega_2} \).

Now, let p vary according to \( p(k) = \overline{p} \lambda^{1/m_1} \). For \( k > 0 \), so as \( p_1/p_2 \to 0 \), \( s^a \to s_{\min}^a = \frac{a^a}{\omega_2} \), while as \( p_1/p_2 \to \infty \), \( s^a \to \frac{a^a}{\omega_1} = s_{\max}^a \). So, as \( p \) varies, \( s^a \) covers the whole range \([s_{\min}^a, s_{\max}^a]\).

We now prove the uniqueness part. Given consumers have homothetic preferences and identical perceptions, \( u'(\omega) = k p^m \) for some \( k \). So (22) holds, for \( \overline{p} \) s.t. \( \overline{p}^n = u'(\omega) \) (i.e. \( \overline{p}_1^{m_1} (p_1^d)^{1-m_1} = u'(\omega) \)).

**Proof of Proposition 16**  Efficiency implies common price misperception. An agent \( a \)'s consumption features \( u'(c_a) = \lambda_a p(m_a) \), where \( p(m_a) \) is the price he perceives. Suppose

42
two consumer $a, b$ don’t perceive the same relative prices. So there are two goods – that we can call 1 and 2, such that $\frac{p_1(m_a)}{p_2(m_a)} < \frac{p_1(m_b)}{p_2(m_b)}$. That implies $\frac{u^a_{c_1}}{u^a_{c_2}} < \frac{u^b_{c_1}}{u^b_{c_2}}$. Hence, as is well-known we can design a Pareto improvement by having $a$ and $b$ trade some of good 1 for good 2.\footnote{This is well-known. Take $\zeta \in (\frac{u^a_{c_1}}{u^a_{c_2}}, \frac{u^b_{c_1}}{u^b_{c_2}})$, and consider, for $\varepsilon \geq 0$, the change in allocation $\delta c_a (\varepsilon) := (-\varepsilon, \xi \varepsilon, 0, ..., 0) := -\delta c_b (\varepsilon)$. Then, define $v^a (\varepsilon) := u(c^a + \delta c_a (\varepsilon))$, and $v^a (\varepsilon)$ similarly. We have $v^a (0) = -u^a_{c_1} + \xi u^a_{c_2} > 0$, and $v^b (0) = u^b_{c_1} - \xi u^b_{c_2} > 0$. So, there is a small $\varepsilon > 0$ such that $\delta c_a (\varepsilon), \delta c_b (\varepsilon)$ lead to a Pareto-improvement.}

**Common price misperception implies efficiency.** Call $p^m$ the vector of perceived (relative) prices. Consider an alternative allocation $(\bar{c}^a)_{a \in A}$ that would strictly Pareto-dominate $(c^a)_{a \in A}$. Consumer $a$ picked allocation $c^a$ to $\bar{c}^a$, so, following the sparse max, (and calling $\lambda^*_a$ the Lagrange multiplier used by $a$), we have $u(c^a) - \lambda^*_a p^m \cdot c^a \geq u(\bar{c}^a) - \lambda^*_a p^m \cdot \bar{c}^a$, which implies, $\lambda^*_a p^m \cdot (\bar{c}^a - c^a) \geq u(\bar{c}^a) - u(c^a) \geq 0$, and $p^m \cdot (\bar{c}^a - c^a) \geq 0$, with at least one inequality strict. Summing over the $a$’s, and using $\sum_a \bar{c}^a = \sum_a c^a = \omega$, we obtain $p^m \cdot (\omega - \omega) > 0$, a contradiction.

**Proof of Proposition 17** Clearly the set of equilibrium allocations for consumer $a$ and in $a$’s offer curve, $OC^a$. Let us now show the converse, i.e. $OC^a \subset C^a$. Take a point in $c \in OC^a$, a price $p$ that supports it ($D^a (p) = c$) and $c^a = \omega - c$ the corresponding allocation for consumer $b$. Because $p \cdot c = p \cdot \omega^a$, and $\omega = \omega^a + \omega^b$, we have $p \cdot c^a = p \cdot \omega^b$, so $c^a$ is on consumer $b$’s budget set for any price $kp$ with $k > 0$. We have to see if there is a $k > 0$ such that price $kp$ leads consumer $b$ to demand consumption $c^a$. This is the case iff there is a Lagrange multiplier $\lambda$ such that $u' (c) = \lambda (kp)^m$ where $(kp)^m$ is the perceived price corresponding to price $kp$, i.e.

$$u' (c^a) = \left( \frac{\lambda k^{m_1} p_1^{m_1}}{\lambda k^{m_2} p_1^{m_2}} \right) \left( \frac{p_1^{d_1}}{p_2^{d_2}} \right)^{1-m_1} \left( \frac{p_1^{d_1}}{p_2^{d_2}} \right)^{1-m_2}$$

(29)

But as $m_1 \neq m_2$, the right-hand side spans $\mathbb{R}^2_+$ as $\lambda, k$ vary in $\mathbb{R}^2_+$. Hence, there are $\lambda, k$ such that (29) holds, i.e. $c^a$ is indeed the consumption demanded by consumer $b$ at price $kp$. That means that $c \in C^a$, and finally $OC^a \subset C^a$.

**Example 6** Call $(c_1, c_2)$ the allocation of consumer $a$, so that consumer $b$’s allocation is $(1 - c_1, 1 - c_2)$. We look for equilibria with prices $(p, p)$ and allocations $c^a = (c, 1 - c)$. The f.o.c. for consumer $a$ is: $\frac{v' (c_1)}{v' (c_2)} = \frac{p_1^\alpha}{p_2^\beta} = p^{\alpha - \beta}$, and for consumer $b$: $\frac{v' (1 - c_1)}{v' (1 - c_2)} = \frac{p_1^{\alpha}}{p_2^{\beta}} = p^{\beta - \alpha}$. This is possible if $\frac{v' (c)}{v' (1 - c)} = p^{\alpha - \beta}$. 

46
Proof of Proposition 18  Direct calculation shows \( \mathbb{E} \left[ \tau^H (\mu, k) \right] = \int_{-\infty}^{\infty} \mu 1_{|\mu|>k} \frac{k^2}{dk} \, dk = \tau (\mu, \kappa) \). So, in the quadratic problem, \( a^{NS} (x) = \sum_i \mu_i x_i \), then the action of a fixed cost agent with cost \( k_i \) is: \( a^{FC} (x) = \sum_i \tau^H (b_i, k_i) x_i \). If \( k_i \) is distributed as \( f (k_i \mid \kappa_i) \),

\[
\mathbb{E} [a^{FC} (x)] = \sum_i \mathbb{E} \left[ \tau^H (b_i, k_i) \right] x_i = \sum_i \tau (b_i, \kappa_i) x_i
\]

Proof of Proposition 19  Given the precision \( T_i \), it is well-known that \( \mathbb{E} [x_i | s_i] = \lambda_i s_i \) with \( \lambda_i = \frac{\text{cov}(x_i, s_i)}{\text{var}(s_i)} = \frac{\sigma_{x_i}^2}{\sigma_{x_i}^2 + \sigma_{s_i}^2} = \frac{T_i}{1 + T_i} \). The optimal action is \( a (s) = \mathbb{E} [\sum_i \mu_i x_i | s] = \sum_i \mu_i \lambda_i s_i \), utility is

\[
\mathbb{E} [u (a (s), \mu, x)] = \frac{1}{2} \mathbb{E} \left[ \left( \sum_i \mu_i (x_i - \lambda_i s_i) \right)^2 \right] = \frac{1}{2} \sum_i \frac{\mu_i^2 \sigma_{x_i}^2}{1 + T_i}
\]

after a bit of algebra. Hence, the problem is: \( \max_{T_i \geq 0} \sum_i \frac{-1}{2} \frac{\mu_i^2 \sigma_{x_i}^2}{1 + T_i} - K^2 T_i \). If the solution is interior (i.e., if \( T_i > 0 \)), we can take the f.o.c.:

\[
\frac{\mu_i^2 \sigma_{x_i}^2}{(1 + T_i)^2} - K^2 = 0, \text{ i.e. } \lambda_i = 1 - \frac{1}{1 + T_i} = 1 - \frac{K}{|\mu_i| \sigma_{x_i}}.
\]

The corner solution \( T_i = 0 \) corresponds to \( \lambda_i = 0 \), so finally: \( \lambda_i = \left( 1 - \frac{K}{|\mu_i| \sigma_{x_i}} \right)^+ \), and the action is \( a (s) = \sum_i \mu_i (1 - \frac{K}{|\mu_i| \sigma_{x_i}})^+ s_i \), i.e. \( a (s) = \sum_i \tau \left( \mu_i, \frac{K}{\sigma_{x_i}} \right) s_i \) and \( \mathbb{E} [a (s) | x] = \sum_i \tau \left( \mu_i, \frac{K}{\sigma_{x_i}} \right) x_i \). The precision is \( T_i = \tau (\mu_i \sigma_{x_i}, K) / K \).

References


Belloni, Alexandre and Victor Chernozhukov, “High Dimensional Sparse Econometric


Donoho, David, “Compressed Sensing,” IEEE Transactions on Information Theory, 52


Hansen, Lars, and Thomas Sargent, 


