Online Appendix for "Optimal Taxation with Behavioral Agents"

Emmanuel Farhi and Xavier Gabaix

August 2015

This online appendix contains additional results and extensions of the paper; proofs that were omitted in the paper; and complements on behavioral welfare theory.

12 Additional Results

12.1 Complements to basic consumer theory

12.1.1 Hybrid Model: Agent maximizing the wrong utility function with the wrong prices

Suppose now an agent with true problem $\max_{\boldsymbol{c}} u(\boldsymbol{c})$ s.t. $B(\boldsymbol{p}, \boldsymbol{c}) \leq w$, but who maximizes $\text{smax}_{\boldsymbol{c}|\boldsymbol{p}^s} u^s(\boldsymbol{c})$ s.t. $B(\boldsymbol{p}, \boldsymbol{c}) \leq w$ with both the wrong utility and the wrong prices. This is hybrid of the two previous models.

In terms of decision (if not welfare), the agent is a misperceiving agent with utility u^s and perceived prices $\boldsymbol{p}^s$. Call $v^s(\boldsymbol{p}, w) = u^s(\boldsymbol{c}(\boldsymbol{p}, w))$ and $\boldsymbol{h}^{r,s}(\boldsymbol{p}^s, \widehat{u}) = \arg\min_{\boldsymbol{c}} B(\boldsymbol{p}^s, \boldsymbol{c})$ s.t. $u^s(\boldsymbol{c}) \geq \widehat{u}$ the indirect utility function (of that misperceiving agent) and the rational compensated demand of that agent with utility u^s. Then, our agent has demand:

$$\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{h}^{r,s}(\boldsymbol{p}^s(\boldsymbol{p}, w), v^s(\boldsymbol{p}, w)) \tag{64}$$

Proposition 12.1 (Agent misperceiving both utility and prices) *Take the model of an agent maximizing the wrong utility function $u^s(\boldsymbol{c})$, with the wrong perceived prices $\boldsymbol{p}^s$. Call $\boldsymbol{S}^{r,s}(\boldsymbol{p}, w) = \boldsymbol{h}^{r,s}_{\boldsymbol{p}^s}(\boldsymbol{p}^s(\boldsymbol{p}, w), v^s(\boldsymbol{p}, w))$ the Slutsky matrix of the underlying rational agent who has utility u^s, and define*

$$\boldsymbol{S}^s_j(\boldsymbol{p}, w) = \boldsymbol{S}^{r,s}(\boldsymbol{p}, w) \cdot \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w) \tag{65}$$

i.e. $S^s_{ij} = \sum_k S^{r,s}_{ik} \frac{\partial p^s_k(\boldsymbol{p}, w)}{\partial p_j}$, *where* $\frac{\partial p^s_k(\boldsymbol{p}, w)}{\partial \boldsymbol{p}_j}$ *is the matrix of marginal perception. Then,*

$$\boldsymbol{S}^C_j(\boldsymbol{p}, w) = \boldsymbol{S}^s_j(\boldsymbol{p}, w) + \boldsymbol{c}_w \left(\frac{v^s_{p_j}}{v^s_w} + B_{p_j} \right)$$

$$\boldsymbol{S}^H_j(\boldsymbol{p}, w) = \boldsymbol{S}^s_j + \boldsymbol{c}_w \left(\frac{v^s_{p_j}}{v^s_w} - \frac{v_{p_j}}{v_w} \right)$$

$$\boldsymbol{S}^s_j(\boldsymbol{p}, w) = \boldsymbol{c}_{p_j} - \boldsymbol{c}_w \frac{v^s_{p_j}}{v^s_w}$$

We can write

$$-D_j = \overline{\boldsymbol{\tau}}^b \cdot \boldsymbol{S}_j^s$$

with:

$$\overline{\boldsymbol{\tau}}^b = \left(B_{\boldsymbol{c}}\left(\boldsymbol{c}, \boldsymbol{p}\right) - B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right)\right) + \left(\frac{u_{\boldsymbol{c}}^s}{v_w^s} - \frac{u_{\boldsymbol{c}}}{v_w}\right) \tag{66}$$

Finally, $B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right) \cdot \boldsymbol{S}_j^s = 0$.

This tax $\overline{\boldsymbol{\tau}}^b$ is the sum of two gaps: between the prices $\left(B_{\boldsymbol{c}}\left(\boldsymbol{c}, \boldsymbol{p}\right) - B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right)\right)$ and perceived prices, and between true utility and perceived utility $\left(\frac{u_{\boldsymbol{c}}^s}{v_w^s} - \frac{u_{\boldsymbol{c}}}{v_w}\right)$.

Proof. So, with $\boldsymbol{M}_j = \frac{\partial \boldsymbol{p}^s(\boldsymbol{p},w)}{\partial p_j}$, and use $\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{h}^{r,s}\left(\boldsymbol{p}^s, v^s\left(\boldsymbol{p}, w\right)\right)$:

$$\begin{aligned}
\boldsymbol{c}_w\left(\boldsymbol{p}, w\right) &= \boldsymbol{h}_u^{r,s} v_w^s \\
\boldsymbol{c}_{\boldsymbol{p}_j}\left(\boldsymbol{p}, w\right) &= \boldsymbol{h}_{p^s}^{r,s} \cdot \boldsymbol{M}_j + \boldsymbol{h}_u^{r,s} v_j^s = \boldsymbol{S}_j^s + \boldsymbol{c}_w \frac{v_j^s}{v_w^s}
\end{aligned}$$

Using (50) and (51) gives:

$$\begin{aligned}
\boldsymbol{S}_j^C &= \boldsymbol{c}_{p_j}\left(\boldsymbol{p}, w\right) + \boldsymbol{c}_w\left(\boldsymbol{p}, w\right) B_{p_j}\left(\boldsymbol{c}, \boldsymbol{p}\right) = \boldsymbol{S}_j^s + \boldsymbol{c}_w\left(\frac{v_j^s}{v_w^s} + B_{\boldsymbol{p}_j}\right) \\
\boldsymbol{S}_j^H &= \boldsymbol{c}_{p_j}\left(\boldsymbol{p}, w\right) - \boldsymbol{c}_w\left(\boldsymbol{p}, w\right) \frac{v_{p_j}\left(\boldsymbol{p}, w\right)}{v_w\left(\boldsymbol{p}, w\right)} = \boldsymbol{S}_j^s + \boldsymbol{c}_w\left(\frac{v_j^s}{v_w^s} - \frac{v_{p_j}\left(\boldsymbol{p}, w\right)}{v_w\left(\boldsymbol{p}, w\right)}\right)
\end{aligned}$$

We have

$$B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right) \cdot \boldsymbol{S}_j^s = B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right) \cdot \boldsymbol{h}^r\left(\boldsymbol{p}^s, v^s\right) \cdot \boldsymbol{M}_j = 0 \text{ as } B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right) \cdot \boldsymbol{h}^r\left(\boldsymbol{p}^s, v^s\right) = 0$$

Recall also that $\frac{u_{\boldsymbol{c}}^s}{v_w^s} = \Lambda B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right)$ as the agent maximizes with perceived prices $\boldsymbol{p}^s$. Hence,

$$\begin{aligned}
-D_j &= \left(B_{\boldsymbol{c}}\left(\boldsymbol{c}, \boldsymbol{p}\right) - \frac{u_{\boldsymbol{c}}}{v_w}\right) \boldsymbol{S}_j^s \\
&= \left(\left(B_{\boldsymbol{c}}\left(\boldsymbol{c}, \boldsymbol{p}\right) - B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right)\right) - \left(\frac{u_{\boldsymbol{c}}}{v_w} - \Lambda B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right)\right)\right) \boldsymbol{S}_j^s \text{ as } B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right) \cdot \boldsymbol{S}_j^s = 0 \\
&= \left(\left(B_{\boldsymbol{c}}\left(\boldsymbol{c}, \boldsymbol{p}\right) - B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right)\right) - \left(\frac{u_{\boldsymbol{c}}}{v_w} - \frac{u_{\boldsymbol{c}}^s}{v_w^s}\right)\right) \cdot \boldsymbol{S}_j^s = \overline{\boldsymbol{\tau}}_b \cdot \boldsymbol{S}_j^H
\end{aligned}$$

□

12.1.2 Representation lemma for behavioral models

The following Lemma means that the demand function of a general abstract consumer can be represented as that of a misperceiving consumer with perceived prices $\boldsymbol{p}^s\left(\boldsymbol{p}, w\right)$.

Lemma 12.1 *(Representing an abstract demand by a misperception). Given an abstract demand* $\boldsymbol{c}(\boldsymbol{p}, w)$*, and a utility function* $u(\boldsymbol{c})$*, we can define the function:*

$$\boldsymbol{p}^s(\boldsymbol{p}, w) = \frac{u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))}{v_w(\boldsymbol{p}, w)} \tag{67}$$

Then, the demand function can be represented as that of a sparse agent with perceived prices $\boldsymbol{p}^s(\boldsymbol{p}, w)$*.*

$$\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{c}^s(\boldsymbol{p}, \boldsymbol{p}^s(\boldsymbol{p}, w), w). \tag{68}$$

Proof The demand of a sparse agent $\boldsymbol{c}^s(\boldsymbol{p}, \boldsymbol{p}^s, w)$ is characterized by $u_{\boldsymbol{c}}(\boldsymbol{c}^s(\boldsymbol{p}, \boldsymbol{p}^s, w)) = \lambda \boldsymbol{p}^s$ for some λ, and $\boldsymbol{p} \cdot \boldsymbol{c} = w$. By construction, we have $u_{\boldsymbol{c}}(\boldsymbol{c}^s(\boldsymbol{p}, \boldsymbol{p}^s, w)) = \lambda \boldsymbol{p}^s$ for $\boldsymbol{p}^s = \overline{\boldsymbol{p}}^s(\boldsymbol{p}, w)$. Hence, the representation is valid. We make a mild assumption, namely that given a $\boldsymbol{c} = \boldsymbol{c}(\boldsymbol{p}, w)$, there's no other $\boldsymbol{c}'$ with $\boldsymbol{p} \cdot \boldsymbol{c}' = w$, $u_{\boldsymbol{c}}(\boldsymbol{c}') = u_{\boldsymbol{c}}(\boldsymbol{c})$, and $u(\boldsymbol{c}') > u(\boldsymbol{c})$. Otherwise, we would need to consider another "branch" of the sparse max, namely a solution $u_{\boldsymbol{c}}(\boldsymbol{c}^s) = \lambda \boldsymbol{p}^s$ with $\boldsymbol{c}^s \cdot \boldsymbol{p} = w$ with λ not necessarily the lowest value possible.□

We note that for any $\boldsymbol{p}^s(\boldsymbol{p}, w) = k u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))$ for some $k > 0$, we have $\frac{u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))}{v_w(\boldsymbol{p}, w)} = \frac{\boldsymbol{p}^s}{\boldsymbol{p}^s \cdot \boldsymbol{c}_w}$ (indeed, both are equal to $\frac{u_{\boldsymbol{c}}}{u_{\boldsymbol{c}} \cdot \boldsymbol{c}_w}$).

By contrast, the general model cannot in general be represented by a decision vs. experienced utility model. Indeed, a decision utility model always generates a symmetric Slutsky matrix $\boldsymbol{S}^H(\boldsymbol{q}, w)$, and this property does not hold in general for the general model. For example, the misperception model with exogenous perception $M_{ij}(\boldsymbol{q}, w) = m_j 1_{\{i=j\}}$ features $S^H_{ij}(\boldsymbol{q}, w) = S^r_{ij}(\boldsymbol{q}, w) m_j$. Since $S^r(\boldsymbol{q}, w)$ is symmetric, $S^H(\boldsymbol{q}, w)$ is not symmetric as long as there exists i and j with $m_i \neq m_j$.

12.2 Complements on Endogenous Attention: Attention as a good

12.2.1 Interpreting attention as a good

We propose a simple abstract way to think about attention, including its potentially suboptimal allocation. Call $\boldsymbol{c} = (\boldsymbol{C}, \boldsymbol{m})$ the "meta-good" made of both regular commodities $\boldsymbol{C}$ and attention $\boldsymbol{m}$, which are both vectors. Utility is $u(\boldsymbol{c}) = u(\boldsymbol{C}, \boldsymbol{m})$. The framework applies $u(\boldsymbol{c})$. There is a demand $\boldsymbol{c}(\boldsymbol{p}, w)$. We can think that attention has market price 0 (it could have a non-zero price, for instance if $\boldsymbol{m}$ is the amount of computer power one uses to optimize).

The concrete example of misperception framework is worth keeping in mind for concreteness. We have the demand $\boldsymbol{C}^s(\boldsymbol{p}, \boldsymbol{p}^s, w, \boldsymbol{m}) = \arg \mathrm{smax}_{\boldsymbol{C}|p^s}\, u(\boldsymbol{C}, \boldsymbol{m})$ s.t. $p \cdot (\boldsymbol{C}, \boldsymbol{m}) \leq w$, which depends on perceived prices $\boldsymbol{p}^s$. Then, given $\boldsymbol{p}^s(\boldsymbol{p}, w, \boldsymbol{m})$, demand is $\boldsymbol{C}(\boldsymbol{p}, w, \boldsymbol{m}) = \boldsymbol{C}^s(\boldsymbol{p}, \boldsymbol{p}^s(\boldsymbol{p}, \boldsymbol{m}, w), w, \boldsymbol{m})$. If $\boldsymbol{m}$ is optimally allocated, $\boldsymbol{m}(\boldsymbol{p}, w) = \arg\max_{\boldsymbol{m}} u(\boldsymbol{C}(\boldsymbol{p}, w, \boldsymbol{m}), \boldsymbol{m})$. In general (even with non-optimal attention), given an attention policy $\boldsymbol{m}(\boldsymbol{p}, w)$, demand is $\boldsymbol{C}(\boldsymbol{p}, w) = \boldsymbol{C}(\boldsymbol{p}, w, \boldsymbol{m}(\boldsymbol{p}, w))$.

Characterizing optimal allocation of attention Suppose that we have a constraint: $\boldsymbol{c} = \boldsymbol{c}(\boldsymbol{p}, w, \theta)$ for some parameter θ. For instance, suppose that $\boldsymbol{c}(\boldsymbol{p}, w, \theta) = (\boldsymbol{C}(\boldsymbol{p}, w, \boldsymbol{m}(\theta)), \boldsymbol{m}(\theta))$; when $\boldsymbol{m}(\theta) = \theta$, we're considering the potentially optimal allocation of attention, as attention affects directly the choice of goods. If $\boldsymbol{m} = (m_1, m_2, m_3) = (\theta_1, \theta_2, \theta_2)$, we captures that the attention to goods 2 and 3 have to be the same.[58]

Proposition 12.2 (Characterizing optimal allocation of attention) *The first order condition for the optimal allocation of parameter θ (i.e., $\theta(\boldsymbol{p}, w) = \arg\max_\theta u(\boldsymbol{c}(\boldsymbol{p}, w, \theta))$) is:*

$$\boldsymbol{\tau}^b \cdot \boldsymbol{c}_\theta(\boldsymbol{p}, w, \theta) = 0 \tag{69}$$

Proof The FOC is $u_{\boldsymbol{c}} \boldsymbol{c}_\theta = 0$. We note that $B_{\boldsymbol{c}} \cdot \boldsymbol{c}_\theta = 0$ by budget constraint: $B(\boldsymbol{c}(\boldsymbol{p}, w, \theta)) = w$. So,

$$\boldsymbol{\tau}^b \cdot \boldsymbol{c}_\theta = \left(B_{\boldsymbol{c}} - \frac{u_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p})}{v_w(\boldsymbol{p}, w)}\right) \cdot \boldsymbol{c}_\theta = -\frac{u_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}) \cdot \boldsymbol{c}_\theta}{v_w(\boldsymbol{p}, w)}$$

so that $\boldsymbol{\tau}^b \cdot \boldsymbol{c}_\theta = 0$ if and only if $u_{\boldsymbol{c}} \cdot \boldsymbol{c}_\theta = 0$. $\square$

Proposition 12.3 (Value of D_j when attention is optimal). *When attention is of the form $\boldsymbol{c}(\boldsymbol{p}, w, \theta) = (\boldsymbol{C}(\boldsymbol{p}, w, \boldsymbol{m}(\theta)), \boldsymbol{m}(\theta))$, and is optimally chosen, then*

$$\begin{aligned}
-D_j &= \boldsymbol{\tau}^b_{\boldsymbol{C}} \cdot \boldsymbol{C}_{p_j}(\boldsymbol{p}, w, \boldsymbol{m})_{|\boldsymbol{m}=\boldsymbol{m}(\theta(\boldsymbol{p},w))} \\
&= \boldsymbol{\tau}^b_{\boldsymbol{C}} \cdot \boldsymbol{S}^H_{j|\boldsymbol{m}}(\boldsymbol{p}, w, \boldsymbol{m})_{|\boldsymbol{m}=\boldsymbol{m}(\theta(\boldsymbol{p},w))} = \boldsymbol{\tau}^b_{\boldsymbol{C}} \cdot \boldsymbol{S}^C_{j|\boldsymbol{m}}(\boldsymbol{p}, w, \boldsymbol{m})_{|\boldsymbol{m}=\boldsymbol{m}(\theta(\boldsymbol{p},w))}
\end{aligned}$$

where $\boldsymbol{\tau}^b_{\boldsymbol{C}} = B_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{p}) - \frac{u_{\boldsymbol{C}}(\boldsymbol{C}, \boldsymbol{m})}{v_w(\boldsymbol{p}, w)}$ is the misoptimization wedge restricted to goods consumption, and $\boldsymbol{S}^H_{j|\boldsymbol{m}}$ and $\boldsymbol{S}^C_{j|\boldsymbol{m}}$ are the Slutsky matrices $\boldsymbol{S}^H_j$ and $\boldsymbol{S}^C_j$ holding attention constant, i.e. associated to decision $\boldsymbol{C}(\boldsymbol{p}, w, \boldsymbol{m})$ with constant $\boldsymbol{m} = \boldsymbol{m}(\theta(\boldsymbol{p}, w))$.

Proof We have

$$\begin{aligned}
-D_j &= \boldsymbol{\tau}^b \cdot \boldsymbol{c}_{p_j}(\boldsymbol{p}, w, \theta) = \boldsymbol{\tau}^b \cdot \left[\left(\boldsymbol{C}_{p_j}(\boldsymbol{p}, w, \boldsymbol{m}), 0\right) + \boldsymbol{c}_\theta(\boldsymbol{p}, w, \theta)\, \theta_{p_j}(\boldsymbol{p}, w)\right] \\
&= \boldsymbol{\tau}^b \cdot \left(\boldsymbol{C}_{p_j}(\boldsymbol{p}, w, \boldsymbol{m}), 0\right) \text{ as } \boldsymbol{\tau}^b \cdot \boldsymbol{c}_\theta = 0 \\
&= \left(\boldsymbol{\tau}^b_{\boldsymbol{c}}, \boldsymbol{\tau}^b_{\boldsymbol{m}}\right) \cdot \left(\boldsymbol{C}_{p_j}(\boldsymbol{p}, w, \boldsymbol{m}), 0\right) = \boldsymbol{\tau}^b_{\boldsymbol{c}} \cdot \boldsymbol{C}_{p_j}(\boldsymbol{p}, w, \boldsymbol{m}). = \boldsymbol{\tau}^b_{\boldsymbol{c}} \cdot \boldsymbol{S}^{C,\boldsymbol{m}}_j = \boldsymbol{\tau}^b_{\boldsymbol{c}} \cdot \boldsymbol{S}^H_{j|\boldsymbol{m}}
\end{aligned}$$

"No attention cost in welfare" benchmark Another benchmark is the "no attention cost in welfare", i.e. the cost of attention is not taken into account in the welfare analysis. Suppose

[58] In a model of noisy decision-making à la Sims (2003), the same logic exactly applies, except that quantities are generally stochastic. The consumption is a random variable $c(p, w, \widetilde{\varepsilon})$, where $\widetilde{\varepsilon}$ indexes noise, rather than a deterministic function. Then, utility is $U(c(p, w)) := \mathbb{E}[u(c(p, w, \widetilde{\varepsilon}))]$, $\boldsymbol{S}^H(p, w)$ is likewise a random variable. We do not pursue this framework further here, at it is hard to solve beyond linear-quadratic settings, e.g. with Gaussian distribution of prices – which in turn generates potentially negative prices.

that attention $\boldsymbol{m}$ just moves with prices, but as an automatic process whose "cost" is not counted: this is, $u(\boldsymbol{C},\boldsymbol{m}) = u(\boldsymbol{C})$ and attention has 0 price, $\boldsymbol{p_m} = 0$. This is the way it is often done in behavioral economic (see however Bernheim and Rangel 2009): people choose using heuristics, but the "cognitive cost" associated with a decision procedure isn't taken into account in the agent's welfare (largely, because it is very hard to measure, and that revealed preference techniques do not apply).

Proposition 12.4 (Value of D_j in the case of fixed attention, and the case of "No attention cost in welfare"). *In the "fixed attention" case and the "No attention cost in welfare"case*

$$-D_j = \left(\boldsymbol{\tau}_C^b, 0\right) \cdot \boldsymbol{S}_j^H(\boldsymbol{p}, w) = \sum_{i=1}^{n} \boldsymbol{\tau}_{C_i}^b \boldsymbol{S}_{ij}^H = \left(\boldsymbol{\tau}_{\boldsymbol{C}}^b, 0\right) \cdot \boldsymbol{S}_j^C(\boldsymbol{p}, w) = \left(\boldsymbol{\tau}_{\boldsymbol{C}}^b, 0\right) \cdot \boldsymbol{c}_j(\boldsymbol{p}, w)$$

This is, only the components of $\boldsymbol{\tau}^b$ and the Slutsky matrix linked to commodities matter.

Proof

We have $\boldsymbol{\tau}^b = \left(\boldsymbol{\tau}_{\boldsymbol{c}}^b, \boldsymbol{\tau}_{\boldsymbol{m}}^b\right) = \left(\boldsymbol{\tau}_{\boldsymbol{c}}^b, 0\right)$ as $u_{\boldsymbol{m}} = 0$. So, $-D_j = \boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^H(\boldsymbol{p}, w) = \boldsymbol{\tau}_{\boldsymbol{C}}^b \cdot \boldsymbol{S}_{C,j}^H$. $\square$

Misperception example In the misperception model with attention policy $\boldsymbol{m}(\boldsymbol{p}, w)$, we have:

$$\boldsymbol{c}(\boldsymbol{p}, w) = \left(\boldsymbol{C}^s\left[\boldsymbol{p}, \boldsymbol{p}^s(\boldsymbol{p}, w, \boldsymbol{m}(\boldsymbol{p}, w)), v(\boldsymbol{p}, w)\right], \boldsymbol{m}(\boldsymbol{p}, w)\right)$$

When attention is optimally chosen, we can apply Proposition 12.3 with $\boldsymbol{m}(\theta) = \theta$. This gives: $-D_j = \boldsymbol{\tau}_{\boldsymbol{C}}^b \cdot \boldsymbol{S}_{\boldsymbol{C}j}^{H,\boldsymbol{m}}$ with

$$\boldsymbol{S}_{\boldsymbol{C}j|\boldsymbol{m}}^H = \boldsymbol{S}^r \boldsymbol{p}_{p_j}^s(\boldsymbol{p}, w, \boldsymbol{m}) \tag{70}$$

i.e. the Slutsky matrix has the sensitivity with *fixed* attention. Hence, we have both $-D_j = \boldsymbol{\tau}_{\boldsymbol{C}}^b \cdot \boldsymbol{S}_j^{H,\boldsymbol{m}}$ when attention is optimal.

In the "no attention cost in welfare" case, $\boldsymbol{\tau}_{\boldsymbol{m}}^b = 0$ and

$$-D_j = \boldsymbol{\tau}_{\boldsymbol{C}}^b \cdot \boldsymbol{S}_{\boldsymbol{C}j}^H$$

When attention is not necessarily optimal, we also have (from (62)), using again decomposition $\boldsymbol{\tau}^b = \left(\boldsymbol{\tau}_{\boldsymbol{c}}^b, \boldsymbol{\tau}_{\boldsymbol{m}}^b\right)$:

$$-D_j = \boldsymbol{\tau}^b \cdot \boldsymbol{S}_j = \boldsymbol{\tau}_{\boldsymbol{C}}^b \cdot \boldsymbol{S}_{\boldsymbol{C}j}^H + \boldsymbol{\tau}_{\boldsymbol{m}}^b \frac{\partial \boldsymbol{m}}{\partial p_j}$$

where $\boldsymbol{S}_{\boldsymbol{C}j}^H = \boldsymbol{S}^r \cdot \boldsymbol{p}_{p_j}^s(\boldsymbol{p}, w)$, where now the total derivative matters, including the variable attention. The online appendix (section 12.2.2) presents more examples.

12.2.2 Attention as a good: examples

A linear-quadratic example To illustrate the situation, we work out completely a linear-quadratic example. Take decision utility have $u^s(c_0, c_1, m) = c_0 + U(c_1) - g(m)$, with $U(c) = \frac{ac - \frac{1}{2}c^2}{\Psi}$ and attention technology $p_1^s(p_1, m) = p_1^d + m\tau_1$, where τ_1 is a tax. Full utility is $u(c_0, c_1, m) = c_0 + U(c_1) - Ag(m)$, where $A = 0$ in the "no attention cost in welfare" case, and $A = 1$ in the "optimally allocated attention" case.

We assume $p_0 = 1$, $\Psi > 0$. Given attention m, demand satisfies $U'(c_1) = p^s$, so $c_1^r(p^s) = a - \Psi p^s$. The perceived tax is:

$$\tau_1^s = m(\tau_1)\tau_1$$

and demand is

$$c_1 = a - \Psi\left(p_1^d + m(\tau_1)\tau_1\right)$$

The losses from inattention are $\frac{1}{2}u_{c_1c_1}(c_1^r - c_1)^2 = -\frac{1}{2}\Psi\tau^2(1-m)^2$. (This is always true to the leading order, and here this is exact as the function is quadratic). Hence, the optimal attention problem is:

$$m(\tau_1) = \arg\max_m -\frac{1}{2}\Psi\tau_1^2(1-m)^2 - g(m),$$

whose first order condition is:

$$g'(m) = \Psi(1-m)\tau_1^2 \tag{71}$$

The Slutsky matrix with constant m has:

$$S_{11|m}^H = \frac{\partial c_1\left(p_1^d + m\tau_1, m\right)}{\partial \tau_1} = -\Psi m$$

while with variable attention $m(p)$, we have:

$$S_{11}^H = \frac{dc_1\left(p_1^d + m\tau_1, m(\tau_1)\right)}{d\tau_1} = -\Psi(m + \tau m'(\tau_1))$$
$$S_{21}^H = \frac{\partial m}{\partial \tau_1} = m'(\tau_1)$$

Then, we have: $\tau^b = (0, q - q^s, Ag'(m)) = (0, \tau_1(1-m), Ag'(m))$, and given $\tau = (0, \tau_1, 0)$, so

$$\tau - \tau^b = (0, \tau_1 m, -Ag'(m))$$
$$S_1^H = (0, -\Psi(m + \tau_1 m'(\tau_1)), m'(\tau_1))$$

Applying Proposition 3.1 gives:

$$\begin{aligned}\frac{\partial L(\boldsymbol{\tau},w)}{\partial\tau_1} &= (\lambda-\gamma)\,c_1+\lambda\left(\tau-\tau^b\right)\cdot S_1^H\\ &= (\lambda-\gamma)\,c_1-\Psi\tau_1 m\,(m+\tau_1 m'(\tau_1))-Ag'(m)\,m'(\tau_1)\end{aligned}$$

Normalize $\lambda=1,\gamma=1-\Lambda$, and define $\psi_1(c_1)=\Psi/c_1$. First, when m_1 is exogenous, we verify our formula from section 2

$$\frac{\partial L(\boldsymbol{\tau},w)}{\partial\tau_1}=\Lambda c_1-\Psi m\tau_1^s$$

i.e. $\tau_1^s=\frac{\Lambda}{m\psi_1}$, $\tau_1=\frac{\Lambda}{m^2\psi_1}$.

Next, in the "no attention cost in welfare" case, $A=0$

$$\frac{\partial L(\boldsymbol{\tau},w)}{\partial\tau_1}=\Lambda c_1-\Psi\tau_1^s\frac{d\tau_1^s(\tau_1,m_1(\tau_1))}{d\tau_1}=\Lambda c_1-\Psi\tau_1^s(m_1+\tau_1 m'(\tau_1))$$

so

$$\tau_1^s=\frac{-c_1\Lambda}{S_{11}^H}=\frac{\Lambda}{(m+\tau_1 m'(\tau_1))\,\psi_1},\qquad \tau_1=\frac{\Lambda}{(m^2+\tau_1 mm'(\tau_1))\,\psi_1}$$

Finally, in the "optimally allocated attention" case, $A=1$. First, we verify:

$$\begin{aligned}-D_1=\tau^b\cdot\boldsymbol{S}^H &= (0,\tau_1(1-m),g'(m))\cdot(0,-\Psi(m+\tau_1 m'(p)),m'(p_1))\\ &= -\Psi(m+\tau_1 m'(p))\,\tau_1(1-m)+g'(m)\,m'(p_1)=-\Psi m\tau_1(1-m)=-(\tau_1-\tau_1^s)\,\Psi m\\ &= \tau_C^s\cdot\boldsymbol{S}_{j|m}^H(\boldsymbol{p},w,m)\end{aligned}$$

with $\tau_C^s=\tau_1-\tau_1^s=(1-m)\,\tau_1$ and $\boldsymbol{S}_{j|m}^H(\boldsymbol{p},w,m)=-\Psi m$.

$$\begin{aligned}\frac{\partial L(\boldsymbol{\tau},w)}{\partial\tau_1}-\Lambda c_1 &= \left(\tau-\tau^b\right)\cdot\boldsymbol{S}_1^H=\boldsymbol{\tau}\cdot\boldsymbol{S}_1^H-\tau^b\cdot\boldsymbol{S}_1^H=\boldsymbol{\tau}\cdot\boldsymbol{S}_1^H-\tau_C^s\cdot\boldsymbol{S}_{1|m}^H\\ &= -\Psi\tau(m+\tau_1 m'(\tau_1))+\Psi\tau(1-m)\,m=-\Psi\tau\left(m^2+\tau m'(\tau)\right)\end{aligned}$$

$$\frac{\partial L(\boldsymbol{\tau},w)}{\partial\tau_1}=\Lambda c_1-\Psi\tau\left(m^2+\tau m'(\tau)\right)$$

so

$$\tau=\frac{\Lambda/\psi_1}{m(\tau)^2+\tau m'(\tau)}.$$

Optimal tax with endogenous, optimally chosen attention There is just one taxed good. The case with many, independent taxed goods follows.

Recall that the consumer chooses: $m(\tau)=\arg\min\frac{-1}{2}\psi y\tau^2(1-m)^2-\kappa g^1(m)$ and the planner chooses: $\tau(\Lambda) := \arg\max_\tau L(\tau,\Lambda)$ with

$$L(\tau,\Lambda) := -\frac{1}{2}\psi y m(\tau)^2 \tau^2 - \kappa g(m(\tau)) + \Lambda y \tau$$

A lemma on scaling We show that it is enough to compute the solution in the case $\psi = y = \kappa = 1$.

Lemma 12.2 *Suppose that when $\psi = y = \kappa = 1$ the optimal tax is $\tau' = f(\Lambda)$ and optimal attention is $m^1(\tau')$. Then, in the general case it is:*

$$\tau(\Lambda) = \sqrt{\frac{\kappa}{\psi y}} f\left(\Lambda \sqrt{\frac{y}{\kappa\psi}}\right)$$

and the attention is $m(\tau) = m^1\left(\tau\sqrt{\frac{\psi y}{\kappa}}\right)$.

For instance, in the basic rational case, $f(\Lambda) = \Lambda$ and $m^1(\tau') = 1$.

Proof This is a simple scaling argument. We define

$$\begin{aligned} m^1(\tau') &:= \arg\min \frac{-1}{2}\tau'^2(1-m)^2 - g^1(m) \\ L^1(\tau',\Lambda') &= -\frac{1}{2}m^1(\tau')^2\tau'^2 - g\left(m^1(\tau')\right) + \Lambda'\tau' \end{aligned}$$

$$\begin{aligned} \tau' &:= \tau\sqrt{\frac{\psi y}{\kappa}} \\ \Lambda' &:= \Lambda\sqrt{\frac{y}{\kappa\psi}} = \frac{\Lambda y}{\kappa}\frac{\tau}{\tau'} \end{aligned}$$

Then, we have:

$$\begin{aligned} m(\tau) &= \arg\min \frac{-1}{2}\frac{\psi y}{\kappa}\tau^2(1-m)^2 - g^1(m) \\ &= m^1(\tau') \\ L(\tau,\Lambda) &= -\frac{1}{2}\psi y m(\tau)^2\tau^2 - \kappa g(m) + \Lambda y\tau \\ &= \kappa\left[-\frac{1}{2}\frac{\psi y}{\kappa}\tau^2 m(\tau)^2 - g(m) + \frac{\Lambda y}{\kappa}\tau\right] \\ &= \kappa L^1(\tau',\Lambda') \end{aligned}$$

Hence, as $\tau' = f(\Lambda')$ at the optimum.$\square$

Example with continuously adjusting attention We have $g(m) = -\kappa \ln(1-m)$, so that attention is $m(\tau) = \left(1 - \frac{1}{\sqrt{\psi y \tau}}\right)_+$. Indeed, $\arg\min \frac{\sigma^2}{2}(1-m)^2 + g(m)$ is $m = \left(1 - \frac{1}{\sigma}\right)_+$.

Proposition 12.5 *In the above setup with optimal attention, the optimal tax is* $\tau_i = \sqrt{\frac{\kappa}{\psi_i y_i}} f\left(\Lambda \sqrt{\frac{y_i}{\kappa \psi_i}}\right)$, *for the continuous function*

$$\begin{aligned} f(\Lambda) &= \frac{\Lambda + 1 + \sqrt{(\Lambda+1)^2 - 4}}{2} \text{ for } \Lambda \geq 1 \\ &= 1 \text{ for } \Lambda < 1. \end{aligned}$$

Also, $m^1(\tau') = \left(1 - \frac{1}{\tau'}\right)_+$.

Proof. *We first reason in the case* $\psi = y = \kappa = 1$. Then,$m(\tau) = \left(1 - \frac{1}{\tau}\right)_+$ and

$$L(\tau) = -\frac{1}{2} m(\tau)^2 \tau^2 - g(m) + \Lambda \tau$$

Then, for $\tau > 1$,

$$L'(\tau) = 1 - \frac{1}{\tau} - \tau + \Lambda$$

so τ is the greater root of:

$$\tau + \frac{1}{\tau} = \Lambda + 1$$

which exists provided $\Lambda \geq 1$, i.e.:

$$\begin{aligned} f(\Lambda) &= \frac{\Lambda + 1 + \sqrt{(\Lambda+1)^2 - 4}}{2} \text{ for } \Lambda \geq 1 \\ &= 1 \text{ for } \Lambda < 1. \end{aligned}$$

□

An example with 0-1 attention A concrete example of attention choice is:

$$m(\tau) = \arg\max_m -\frac{1}{2}\psi \tau^2 (1-m)^2 - g(m)$$

with

$$g(m) = \frac{1}{2}\kappa^2 \left[1 - (1-m)^2\right].$$

Then, the solution is

$$m(\tau) = 1_{\tau > \tau_*}, \qquad \tau_* := \frac{\kappa}{\sqrt{\psi}} \tag{72}$$

As an aside, a fixed cost $g(m) = \frac{\kappa^2}{2} 1_{m>0}$ gives the same result.

Proposition 12.6 *The optimal tax is* $\tau_i = \sqrt{\frac{\kappa}{\psi_i y_i}} f\left(\Lambda\sqrt{\frac{y_i}{\kappa\psi_i}}\right)$*, for* $f(\Lambda) = 1$ *if* $\Lambda \leq \sqrt{2}+1$ *and* $f(\Lambda) = \Lambda$ *if* $\Lambda > \sqrt{2}+1$*. Also,* $m^1(\tau') = 1_{\tau'>1}$.

In that case, the optimal tax has a discontinuity. When Λ is low enough, the planner keeps the taxes at $\tau_i = \sqrt{\frac{\kappa}{\psi_i y_i}}$, just below the "detectability threshold" and agents do not pay attention to the tax.

Proof We start with the case $\psi = y = \kappa = 1$. Then, $m(\tau) = 1_{\tau>1}$. For $\tau \leq 1$, $L(\tau) = \Lambda\tau$, so the optimum for $\tau \in [0,1]$ is $\tau = 1$.

$$L(1) = \Lambda$$

For $\tau > \tau_*$, $m(\tau) = 1$, so $L(\tau) = -\frac{1}{2}\tau^2 - g(1) + \Lambda\tau$, and the optimum is $\tau = \Lambda$. We have

$$L(\tau) = \frac{\Lambda^2 - 1}{2}$$

So $L(\tau) > L(\tau_*)$ if and only if $\frac{\Lambda^2-1}{2} > \Lambda$, i.e. if and only if$\frac{\Lambda^2-1}{2} > \Lambda$, i.e. if and only if $\Lambda > \sqrt{2}+1$.$\square$

12.3 Mental Accounts: Complements

The optimal tax formulas in Propositions 3.1 and 3.2 corresponding to the many-person Ramsey problem without and with externalities can then be applied without modifications to this simple model of mental accounting. However, it is also enlightening to write these formulas in a slightly different way by leveraging the specific structure of the simple mental accounting model. We define the "income k-compensated" Slutstky matrix for the extended demand function as

$$\boldsymbol{S}_j^{C,k}(\boldsymbol{q}, \boldsymbol{\omega}) = \boldsymbol{c}_{q_j}(\boldsymbol{q}, \boldsymbol{\omega}) + \boldsymbol{c}_{\omega^k}(\boldsymbol{q}, \boldsymbol{\omega}) c_j(\boldsymbol{q}, \boldsymbol{\omega}). \tag{73}$$

This Slutsky matrix corresponds to a decomposition of price effects into income of substitution effects where the latter are compensated using with an adjustment of mental account k. In the traditional model without behavioral biases, this decomposition is independent of the mental account k which is used for this decomposition, since the marginal utility of income is equalized across all accounts: $\boldsymbol{c}_{\omega^k}(\boldsymbol{q}, \boldsymbol{\omega}^r(\boldsymbol{q}, w)) = \boldsymbol{c}_{\omega^{k'}}(\boldsymbol{q}, \boldsymbol{\omega}^r(\boldsymbol{q}, w))$. It follows that the "income k-compensated" Slutstky matrix is also independent of k.[59] By contrast, with behavioral biases in mental accounting, the marginal utility of income is not equalized across all accounts so that in general $\boldsymbol{c}_{\omega^k}(\boldsymbol{q}, \boldsymbol{\omega}(\boldsymbol{q}, w)) \neq \boldsymbol{c}_{\omega^{k'}}(\boldsymbol{q}, \boldsymbol{\omega}(\boldsymbol{q}, w))$. As a result, the decomposition of price effects on income effects

[59] However the "income k-compensated" Slutstky matrix $S_j^{C,k,r}(\boldsymbol{q}, \omega^r(\boldsymbol{q}, w))$ of the extended demand function is in general different from the "income compensated" Slutstky matrix $S_j^{C,r}(\boldsymbol{q}, w)$ of the demand function, which is defined as in Section 3.1. Indeed, the latter also reflects the subsitution effects associated with the adjustments $\omega_{q_j}^{k,r}(\boldsymbol{q}, w)$ in the mental accounts in response to changes in the price q_j of commodity j.

and substitution effects depends on which mental account k is used for this income compensation, and the "income k-compensatedq" Slutstky matrix depends on k.[60]

Reintroducing h superscripts to index agent heterogeneity, we define the social marginal utility of k-income for agent h as

$$\gamma^{k,h} = \beta^{k,h} + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}^h_{\omega^{k,h}} \text{ where } \beta^{k,h} = W_{v^h} v^h_{\omega^{k,h}}.$$

We also define the income-k based misoptimization wedges for the extended demand and utility function as

$$\boldsymbol{\tau}^{b,k} = \boldsymbol{q} - \frac{u_{\boldsymbol{c}}}{v_{\omega_k}}, \qquad \tilde{\boldsymbol{\tau}}^{b,k,h} = \beta^{k,h} \boldsymbol{\tau}^{b,k}.$$

Finally, for every commodity i, we denote by $k(i)$ the mental account to which this commodity is associated with. We can then rewrite the tax formula in the following way. Note that this is simply a re-expression of Proposition 3.1.

Proposition 12.7 (Many-person Ramsey with mental accounting) *If commodity i can be taxed, then at the optimum*

$$\frac{\partial L(\boldsymbol{\tau})}{\partial \tau_i} = 0 \quad \textit{with} \quad \frac{\partial L(\boldsymbol{\tau})}{\partial \tau_i} = \sum_h [\left(\lambda - \gamma^{k(i),h}\right) c^h_i + \lambda (\boldsymbol{\tau} - \tilde{\boldsymbol{\tau}}^{b,k(i),h}) \cdot \boldsymbol{S}^{C,k(i),h}_i + \sum_k \gamma^{k,h} \omega^{k,h}_{q_i}]. \quad (74)$$

This alternative expression of the many-person Ramsey optimal tax formula for commodity i features the "income $k(i)$-compensated" Slutsky matrix corresponding the the mental account to which commodity i is associated, the social marginal utilities of k-income $\gamma^{k.h}$, the a$k(i)$ based misoptimization wedges, and the price derivatives of the mental accounting functions $\omega^{k,h}_{q_i}$. Writing the optimal tax formula in this way will prove useful to derive specific results below in the context of further specializations of the model.

We could also derive a similar alternative expression for the many-person Ramsey optimal tax formula in the presence of externalities along very similar lines. In the interest of space, we do not include it in the paper.

[60] The price theory concepts introduced in Section 3.1 are still defined in the same way. They can be related to the corresponding concepts that we have introduced in this section. In particular, we have

$$c_w(\boldsymbol{q}, w) = \sum_k \omega^k_w \boldsymbol{c}_{\omega^k}(\boldsymbol{q}, \omega),$$

and

$$\boldsymbol{S}^C_j(\boldsymbol{q}, w) = \boldsymbol{c}_{q_j}(\boldsymbol{q}, \omega) + \sum_k \omega^k_{q_j} \boldsymbol{c}_{\omega^k}(\boldsymbol{q}, \omega) + \sum_k \omega^k_w \boldsymbol{c}_{\omega^k}(\boldsymbol{q}, \omega) c_j(\boldsymbol{q}, \omega).$$

12.3.1 Roy's identity with mental accounts

We consider the extended indirect utility function $v(\boldsymbol{p},\boldsymbol{\omega}) = u(\boldsymbol{c}(\boldsymbol{p},\boldsymbol{\omega}))$. The budget constraint is $B(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega}) \leq 0$. A leading case is the linear budget constraint, $B(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega}) = \max_k \boldsymbol{C}^k \cdot \boldsymbol{p}^k - \omega^k$. We define the misoptimization wedge linked to account k as:

$$\boldsymbol{\tau}^{b,k} := -\frac{u_{\boldsymbol{c}}}{v_{\omega_k}} - \frac{B_{\boldsymbol{c}}(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega})}{B_{\omega_k}(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega})}$$

With the linear budget constraint

$$\boldsymbol{\tau}^{b,k} = \boldsymbol{p} - \frac{u_{\boldsymbol{c}}}{v_{\omega_k}}$$

Proposition 12.8 (Roy's identity with mental accounts) *With mental account, the modified Roy's identity is:*

$$\frac{v_{p_i}(\boldsymbol{p},\boldsymbol{\omega})}{v_{\omega_k}(\boldsymbol{p},\boldsymbol{\omega})} = \frac{B_{p_i}}{B_{\omega_k}} - \boldsymbol{\tau}^{b,k} \cdot \boldsymbol{c}_{p_i} = \frac{B_{p_i}}{B_{\omega_k}} - \boldsymbol{\tau}^{b,k} \cdot \boldsymbol{S}_i^{C,k} \tag{75}$$

With a linear budget constraint,

$$\frac{v_{p_i}(\boldsymbol{p},\boldsymbol{\omega})}{v_{\omega_k}(\boldsymbol{p},\boldsymbol{\omega})} = -c_i - \boldsymbol{\tau}^{b,k} \cdot \boldsymbol{S}_i^{C,k} \tag{76}$$

Proof We first note a few simple identities. As $B(\boldsymbol{c}(\boldsymbol{p},\boldsymbol{\omega}),\boldsymbol{p},\boldsymbol{\omega}) = 0$ and $v(\boldsymbol{p},\boldsymbol{\omega}) = u(\boldsymbol{c}(\boldsymbol{p},\boldsymbol{\omega}))$, we have:

$$B_{\boldsymbol{c}}\boldsymbol{c}_{p_i} + B_{p_i} = 0, \qquad B_{\boldsymbol{c}}\boldsymbol{c}_{\omega^k} + B_{\omega^k} = 0, \qquad v_{\omega^k} = u_{\boldsymbol{c}}\boldsymbol{c}_{\omega^k} \tag{77}$$

We calculate:

$$\begin{aligned}\boldsymbol{\tau}^{b,k} \cdot \boldsymbol{c}_{\omega_l} &= \left(-\frac{u_{\boldsymbol{c}}}{v_{\omega_k}} - \frac{B_c}{B_{\omega_k}}\right) \cdot \boldsymbol{c}_{\omega_l} \\ &= -\frac{v_{\omega_l}}{v_{\omega_k}} + \frac{B_{\omega_l}}{B_{\omega_k}}\end{aligned}$$

using (77). So typically $\boldsymbol{\tau}^{b,k} \cdot \boldsymbol{c}_{\omega_l} \neq 0$, except when $l = k$:

$$\boldsymbol{\tau}^{b,k} \cdot \boldsymbol{c}_{\omega^k} = 0 \tag{78}$$

We are now ready to study Roy's identity. We have:

$$\begin{aligned}
\frac{v_{p_i}(\boldsymbol{p},\boldsymbol{\omega})}{v_{\omega_k}(\boldsymbol{p},\boldsymbol{\omega})} &= \frac{u_{\boldsymbol{c}}\boldsymbol{c}_{p_i}(\boldsymbol{p},\boldsymbol{\omega})}{v_{\omega_k}} = \left(\frac{u_{\boldsymbol{c}}}{v_{\omega_k}} + \frac{B_c(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega})}{B_{\omega_k}(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega})} - \frac{B_c(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega})}{B_{\omega_k}(\boldsymbol{c},\boldsymbol{p},\boldsymbol{\omega})}\right)\boldsymbol{c}_{p_i} \\
&= \left(\frac{u_{\boldsymbol{c}}}{v_{\omega_k}} + \frac{B_c}{B_{\omega_k}}\right)\boldsymbol{c}_{p_i} - \frac{B_c}{B_{\omega_k}}\boldsymbol{c}_{p_i} \\
\frac{v_{p_i}(\boldsymbol{p},\boldsymbol{\omega})}{v_{\omega_k}(\boldsymbol{p},\boldsymbol{\omega})} &= -\boldsymbol{\tau}^{b,k}\cdot\boldsymbol{c}_{p_i} + \frac{B_{p_i}}{B_{\omega_k}} \qquad (79)
\end{aligned}$$

using (77).

Using (78) gives:

$$\boldsymbol{\tau}^{b,k}\cdot c_{p_j} = \boldsymbol{\tau}^{b,k}\cdot\boldsymbol{S}_j^{C,k}$$

so

$$\frac{v_{p_i}(\boldsymbol{p},\boldsymbol{\omega})}{v_{\omega_k}(\boldsymbol{p},\boldsymbol{\omega})} = -\boldsymbol{\tau}^{b,k}\cdot\boldsymbol{S}_j^{C,k} + \frac{B_{p_i}}{B_{\omega_k}}$$

□

12.3.2 How mental accounts modify demand elasticities

We take the quasilinear case $u(\boldsymbol{c}) = c_0 + U^1(c_1) + U(c_2,...,c_n)$ with good 1 in its own mental account, ω_1, and default ω_1^d. How much will be attributed to the mental account? We will haQve $c_1 = \frac{\omega_1}{q_1}$

$$\omega_1 = \arg\max_{\omega_1} U^1\left(\frac{\omega_1}{q_1}\right) - \omega^1 - g^1\left(\omega_1 - \omega_1^d\right) \qquad (80)$$

$$c_1 = \arg\max_{c_1} U^1(c_1) - q_1 c_1 - g^1\left(c_1 q_1 - \omega_1^d\right)$$

Then, we can calculate the sensitivity to the tax.

Lemma 12.3 *With a flexible mental account, the empirical elasticity is:*

$$\alpha_1 := -\frac{q_1}{c_1}\frac{\partial c_1}{\partial q_1} = \frac{1 + g'\left(\omega_1 - \omega_1^d\right) + \omega_1 g''\left(\omega_1 - \omega_1^d\right)}{\frac{1}{\psi_1} + \omega_1 g''\left(\omega_1 - \omega_1^d\right)}$$

with $\omega_1 = q_1 c_1$.

Proof The first order condition for consumption is:

$$f(c_1, q_1) := U^{1\prime}(c_1) - q_1 - q_1 g'\left(c_1 q_1 - \omega_1^d\right) = 0$$

Hence

$$\alpha_1 := -\frac{q_1}{c_1}\frac{\partial c_1}{\partial q_1} = -\frac{q_1}{c_1}\frac{f_{q_1}}{-f_{c_1}} = \frac{q_1}{c_1}\frac{1+g'+q_1c_1g''}{-U''+q_1^2g''}$$
$$= \frac{1+g'+q_1c_1g''}{-\frac{c_1U''}{q_1}+q_1c_1g''}$$

The rational elasticity is the one that would occur with $g=0$,

$$\psi_1 = \frac{1}{-\frac{c_1U_1''(c_1)}{q_1}} \tag{81}$$

so

$$\alpha_1 = \frac{1+g'+q_1c_1g''}{\frac{1}{\psi_1}+q_1c_1g''}$$

□

Next, we suppose that at $q_1 = p_1$, the account is optimal: $\omega_1^d = \arg\max_{\omega_1} U^1\left(\frac{\omega_1}{q_1}\right) - \omega^1$. We suppose that we are near $\omega^1 = \omega_1^d$, and $g'(0) = 0$. We have

$$\alpha^1 = \frac{1+\omega_1 g''\left(\omega_1-\omega_1^d\right)}{\frac{1}{\psi_1}+\omega_1 g''\left(\omega_1-\omega_1^d\right)} \tag{82}$$

In the traditional case, $g''=0$, so $\alpha^1 = \psi_1$. In the completely rigid case, $g'' = +\infty$, so $\alpha^1 = 1$ (indeed, we have then $c_1 = \frac{\omega_1^d}{q_1}$, so the elasticity of demand is 1).

This allows to calculate the $\omega_{q_1}^1$, the derivative of the account value as a function of the price. Starting from $\omega^1 = q_1c_1$, we have

$$\omega_{q_1}^1 = c_1 + q_1\frac{\partial c_1}{\partial q_1} = c_1 - c_1\alpha_1 = c_1\left[1-\frac{1+\omega_1 g''\left(\omega_1-\omega_1^d\right)}{\frac{1}{\psi_1}+\omega_1 g''\left(\omega_1-\omega_1^d\right)}\right]$$

so finally

$$\omega_{q_1}^1 = c_1\frac{\frac{1}{\psi_1}-1}{\frac{1}{\psi_1}+\omega_1 g''\left(\omega_1-\omega_1^d\right)} \tag{83}$$

By the budget constraint,

$$\omega_{q_1}^0 = -\omega_{q_1}^1$$

12.3.3 Summarizing the effects of misperceptions and mental accounts

We call $\alpha_i := -\frac{q^i}{c^i}c_{\tau_i}^i$ the empirical elasticity, which is $\alpha_i = \psi_i$ in the traditional model, and $\alpha_i = m_i\psi_i$ in the misperception model with attention m_i to the tax. We call $\psi_i = -\frac{U^{i\prime}(c_i)}{c_iU^{i\prime\prime}(c_i)}$, which is simply the inverse of the curvature of the utility function U^i for good i. This is also the "rational" elasticity, the demand elasticity that the agent would have if he was fully attentive; it might be the elasticity elicited in a careful procedure that makes the agent attentive to the tax.

We have the following Lemma.

Lemma 12.4 *As explained just above, call ψ_i the "rational" demand elasticity, and α_i the "behavioral" elasticity. In the limit of small taxes, welfare is:*

$$L(\tau) - L(0) = -\frac{1}{2}\sum_i -\frac{\alpha_i^2}{\psi_i} y_i \tau_i^2 + \Lambda \sum_i \tau_i y_i + o\left(\|\tau\|^2\right) + o\left(\|\tau\| \Lambda\right) \tag{84}$$

Proof of Lemma 12.4 Demand is:

$$c_i(\tau_i) = y_i \left(1 - \alpha_i \tau_i + O\left(\tau_i^2\right)\right), \tag{85}$$

Hence have:

$$\begin{aligned} L &= \sum_i U^i(c_i(\tau_i)) - (1+\tau_i) c_i(\tau_i) + (1+\Lambda)\tau_i c_i(\tau_i) \\ &= \sum_i U^i(c_i(\tau_i)) - c_i(\tau_i) + \Lambda \tau_i c_i(\tau_i) \\ &= \sum_i f^i(\tau_i) + \Lambda \tau_i c_i(\tau_i) \end{aligned}$$

with

$$f_i(\tau_i) = F^i(c_i(\tau_i)), \qquad F^i(c) = U^i(c) - c$$

We have

$$f_i(\tau_i) - f_i(0) = f_i'(0)\tau_i + \frac{1}{2} f_i''(0)\tau_i^2 + o\left(\tau^2\right)$$

$$\begin{aligned} f_i'(\tau_i) &= F^{i\prime}(c_i(\tau_i))\, c_i'(\tau_i) \\ f_i''(\tau_i) &= F^{i\prime\prime}(c_i(\tau_i))\, c_i'(\tau)^2 + F^{i\prime}(c_i(\tau_i))\, c_i''(\tau_i) \end{aligned}$$

As $F_i'(0) = 0$, we have

$$\begin{aligned} f_i'(0) &= 0 \\ f_i''(0) &= F''(c_i(\tau_i))\, c_i'(\tau)^2 = U_i''(c_i)\, c_i^2 \alpha_i^2 \text{ using (85)} \\ &= \frac{-1}{\psi_i} q_i c_i \alpha_i^2 \text{ using (81)} \\ &= -\frac{\alpha_i^2}{\psi_i} y_i \end{aligned}$$

so

$$
\begin{aligned}
L(\tau) - L(0) &= \sum_i \left[\frac{1}{2} f''(0)\, \tau_i^2 + \Lambda \tau_i q_i c_i \right] + o\left(\tau^2\right) + o(\tau \Lambda) \\
&= \sum_i \left[\frac{1}{2} \frac{\alpha_i^2}{\psi_i} q_i c_i + \Lambda \tau_i q_i c_i \right] + o\left(\tau^2\right) + o(\tau \Lambda)
\end{aligned}
$$

So the objective function is:

$$
L = -\frac{1}{2} \sum_i -\frac{\alpha_i^2}{\psi_i} y_i \tau_i^2 + \Lambda \sum_i \tau_i c_i + o\left(\tau^2\right) + o(\tau \Lambda) \tag{86}
$$

□

This implies the following.

Proposition 12.9 *In the basic Ramsey model, the optimal tax is*

$$
\tau_i = \Lambda \frac{\psi_i}{\alpha_i^2} \tag{87}
$$

where ψ_i is the underlying elasticity of true preferences, and α_i is the behavioral elasticity.

Proof We can also use the general formulas (Proposition 3.1) to verify the result. However, it is also instructive to use the following derivation. Maximizing over τ_i, the result from Lemma 12.4

$$
L = -\frac{1}{2} \sum_i -\frac{\alpha_i^2}{\psi_i} y_i \tau_i^2 + \Lambda \sum_i \tau_i y_i
$$

we find: $\tau_i = \frac{\Lambda \psi_i}{\alpha_i^2}$.□

For instance, in the traditional case $\alpha_i = \psi_i$, and we recover the traditional formula $\tau_i = \frac{\Lambda}{\psi_i}$.

Proposition 12.10 *In the basic Pigou model, the optimal tax is $\tau_i = \xi_i \frac{\psi_i}{\alpha_i}$.*

Proof We would like this to be the first best allocation, so that $u'(c_i) = 1 + \xi_i$, i.e. $c_i = c_i^d (1 - \psi_i \xi_i)$. The response to the tax is: $c_i = c_i^d (1 - \alpha_i \tau_i)$. So optimal tax satisfies: $\alpha_i \tau_i = \psi_i \xi_i$, i.e. $\tau_i = \xi_i \frac{\psi_i}{\alpha_i}$. □

The Table shows the link between different models. We use $\omega_i = c_i q_i$.

	Ramsey problem	Pigou problem	Elasticity α_i to tax rate
General	$\tau_i = \Lambda \frac{\psi_i}{\alpha_i^2}$	$\tau_i = \xi_i \frac{\psi_i}{\alpha_i}$	$\alpha_i := -\frac{q^i}{c^i} c^i_{\tau_i}$
Traditional model	$\tau_i = \frac{\Lambda}{\psi_i}$	$\tau_i = \xi_i$	$\alpha_i = \psi_i$
Misperception model	$\tau_i = \frac{\Lambda}{m_i^2 \psi_i}$	$\tau_i = \frac{\xi_i}{m_i}$	$\alpha_i = m_i \psi_i$
Mental account: rigid	$\tau_i = \Lambda \psi_i$	$\tau_i = \xi_i \psi_i$	$\alpha_i = 1$
Mental account: flexible	$\tau_i = \frac{\Lambda}{\psi_i} \left(\frac{1+\psi_i \omega_i g_i''}{1+\omega_i g_i''} \right)^2$	$\tau_i = \xi_i \frac{1+\psi_i \omega_i g_i''}{1+\omega_i g_i''}$	$\alpha_i = \frac{1+\omega_i g_i''}{\frac{1}{\psi_i}+\omega_i g_i''}$
Hybrid model: Flexible mental account with misperceptions	$\tau_i = \frac{\Lambda}{\psi_i} \frac{1}{m_i^2} \left(\frac{1+\psi_i \omega_i g_i''}{1+\omega_i g_i''} \right)^2$	$\tau_i = \frac{\xi_i}{m_i} \frac{1+\psi_i \omega_i g_i''}{1+\omega_i g_i''}$	$\alpha_i = m_i \frac{1+\omega_i g_i''}{\frac{1}{\psi_i}+\omega_i g_i''}$

12.3.4 Cross-Effects of Attention

How does attention to one good affect the optimal tax on another? To answer this question, we use the specialization of the general model developed in Section 3.5, assuming a representative consumer (so that we drop the index h), no internality/externality so that $\boldsymbol{\tau}^X = 0$, and in the limit of small taxes. Defining $\Lambda = \frac{\lambda}{\gamma} - 1$, we can rewrite formula (25) as

$$\boldsymbol{\tau} = -\Lambda \left(\boldsymbol{M}' \boldsymbol{S}^r \boldsymbol{M} \right)^{-1} \boldsymbol{c}.$$

This is a generalization of Proposition 2.1, which assumed a diagonal matrix $\boldsymbol{S}^r$.

To gain intuition, we take $n = 2$ goods, $\boldsymbol{M} = diag\,(m_1, m_2)$, we normalize prices to $p_1 = p_2 = 1$, and we write the rational Slutsky matrix as $S^r_{ii} = -c_i \psi_i$ for $i = 1, 2$, and $S^r_{12} = S^r_{21} = -\sqrt{c_1 c_2 \psi_1 \psi_2}\rho$.

Proposition 12.11 (Impact of cross-elasticities on optimal taxes with inattentive agents) *With two taxed goods, the optimal tax on good 1 is* $\tau_1 = \frac{\Lambda}{m_1^2 \psi_1} \frac{1-\rho\sqrt{\frac{m_1^2 \psi_1 c_2}{m_2^2 \psi_2 c_1}}}{1-\rho^2}$. *When attention to the tax of good 2 m_2 falls, the optimal tax on good 1 increases (respectively decreases) if goods 1 and 2 are substitutes (respectively complements).*

Suppose for example that the goods are substitutes with $\rho < 0$.[61] When m_2 falls, the optimal tax on good 2 increases by the effects in Proposition 2.1, and optimal taxes on substitute goods also increase.[62]

[61] We have $\rho^2 < 1$ since $\boldsymbol{S}^r$ is a 2×2 negative definite matrix so that $0 < \det \boldsymbol{S}^r = c_1 c_2 \psi_1 \psi_2 \left(1 - \rho^2\right)$.

[62] Perhaps curiously, we can have $\frac{\partial \tau_1}{\partial m_1} > 0$ with complement goods $\rho > 0$. This happens if and only if $2 < \frac{m_1}{m_2} \rho \sqrt{\frac{\psi_1 c_2}{\psi_2 c_1}}$. That latter condition is quite extreme, and would imply that $\tau_1 < 0$ even though the planner wants to raise revenues. This is because the planner wants to increase consumption of the low elasticity (low m_2, ψ_2), good 2, he wants to subsidize good 1 if it is a strong complement of good 2.

12.4 Nudges vs Taxes with Redistributional Concerns

The following complements the example in Section 4.3. We assume that $\left(\boldsymbol{\tau}-\boldsymbol{\tau}^{\xi h}\right)\cdot \boldsymbol{c}_w^h=0$, so that $\beta^h=\gamma^h$. As in Section 3.5, we call $\xi^h:=\tau^{X,h}=\frac{\gamma^{\xi,h}}{\lambda}\boldsymbol{\tau}^{I,h}+\boldsymbol{\tau}^{\xi,h}$ as the sum of the internality plus externality.

$$c^h\left(\tau,\tau^{\chi,h}\right)=c_0^h-\Psi\left(m^h\tau+\tau^{\chi,h}\right)$$

Individual h creates an externality plus internality. We have successively:

$$\begin{aligned}
\Xi &= -\mathbb{E}\left[\beta^{h\prime}\right]=-\mathbb{E}\left[\gamma^{h\prime}\right]=-\lambda\\
\tau^{\xi h} &= \frac{-\Xi}{\lambda}\tau^{\xi,h}=\tau^{\xi,h}\\
\tau^{bh} &= \left(1-m^h\right)\tau+\tau^{I,h}-\tau^{\chi,h}\\
\widetilde{\boldsymbol{\tau}}^{b,\xi,h} &= \frac{\gamma^{\xi,h}}{\lambda}\boldsymbol{\tau}^{b,h}\\
\lambda\left(\tau-\tau^{\xi,h}-\widetilde{\tau}^{b,\xi,h}\right) &= \lambda\left(\tau-\tau^{\xi,h}\right)-\gamma^h\left(\left(1-m^h\right)\tau+\tau^{I,h}-\tau^{\chi,h}\right)\\
&= \left(\lambda-\gamma^h\left(1-m^h\right)\right)\tau-\lambda\tau^{X,h}+\gamma^h\tau^{\chi,h}
\end{aligned}$$

Proposition 3.2 gives, using $\boldsymbol{S}_i^{H,h}=-\Psi m^h$,

$$\begin{aligned}
\frac{\partial L}{\partial\tau} &= \mathbb{E}\sum_h\left(\lambda-\gamma_h\right)c^h-\lambda\left(\tau-\tau^{\xi,h}-\widetilde{\tau}^{b,\xi,h}\right)\Psi m^h\\
\frac{\partial L}{\partial\tau} &= \mathbb{E}\sum_h\left(\lambda-\gamma_h\right)c^h-\left[\left(\lambda-\gamma^h\left(1-m^h\right)\right)\tau-\lambda\tau^{X,h}+\gamma^h\tau^{\chi,h}\right]\Psi m^h
\end{aligned}\tag{88}$$

We use the notation

$$\sigma_{Y,Z}:=cov\left(Y_h,Z_h\right)$$

Using$\mathbb{E}\left[\gamma^h\right]=\lambda$, we have:

$$-\mathbb{E}\left[\left(\lambda-\gamma^h\right)m^h\right]=\mathbb{E}\left[\gamma^h m^h\right]-\mathbb{E}\left[\gamma^h\right]\mathbb{E}\left[m^h\right]=\sigma_{\gamma,m}$$

Hence, using $\tau^{\chi,h}=\chi\eta^h$

$$\begin{aligned}
\frac{1}{\Psi}\frac{\partial L}{\partial\tau} &= -\mathbb{E}\left[\left(\lambda-\gamma^h\left(1-m^h\right)\right)m^h\right]\tau-\mathbb{E}\left[\gamma^h\eta^h m^h\right]\chi+\mathbb{E}\left[\left(\lambda-\gamma_h\right)\frac{c^h}{\Psi}+\lambda\tau^{X,h}m^h\right]\\
&= -\mathbb{E}\left[\left(\lambda-\gamma^h\left(1-m^h\right)\right)m^h\right]\tau-\mathbb{E}\left[\gamma^h\eta^h m^h\right]\chi+\mathbb{E}\left[\left(\lambda-\gamma^h\right)\frac{c^h}{\Psi}+\lambda\tau^{X,h}m^h\right]\\
&= -\mathbb{E}\left[\gamma^h {m^h}^2-\sigma_{\gamma m}\right]\tau-\mathbb{E}\left[\gamma^h\eta^h m^h\right]\chi+\mathbb{E}\left[\lambda\tau^{X,h}m^h-\sigma_{\gamma,\frac{c}{\Psi}}\right]
\end{aligned}$$

Proposition 3.3 gives:

$$\frac{\partial L}{\partial \chi} = \sum_h \left[\lambda\left(\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi h}\right) - \beta^h \tau^{b,\xi,h}\right] \boldsymbol{c}_\chi^h$$
$$= -\mathbb{E}\sum_h \left[\lambda\left(\tau - \tau^{X,h}\right) - \gamma^h\left(\left(1-m^h\right)\tau - \tau^{\chi,h}\right)\right]\Psi\tau_\chi^{\chi,h}$$
$$\frac{\partial L}{\partial \chi} = -\mathbb{E}\sum_h \left[\left(\lambda - \gamma^h\left(1-m^h\right)\right)\tau - \lambda\tau^{X,h} + \gamma^h\tau^{\chi,h}\right]\Psi\tau_\chi^{\chi,h}$$

Using $\tau^{\chi,h} = \chi\eta^h$ gives $\tau_\chi^{\chi,h} = \eta^h$ hence:

$$\frac{1}{\Psi}\frac{\partial L}{\partial \chi} = -\mathbb{E}\sum_h \left[\left(\lambda - \gamma^h\left(1-m^h\right)\right)\tau - \lambda\tau^{X,h} + \gamma^h\tau^{\chi,h}\right]\eta^h$$
$$= -\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]\tau - \mathbb{E}\left[\gamma^h {\eta^h}^2\right]\chi + \mathbb{E}\left[\lambda\tau^{X,h}\eta^h\right]$$

This implies

$$\frac{1}{\Psi}\frac{\partial^2 L}{\partial\tau\partial\chi} = -\mathbb{E}\sum_h \left[\left(\lambda - \gamma^h\left(1-m^h\right)\right)\eta^h\right]$$

Hence, at the optimum:

$$\mathbb{E}\left[\gamma^h {m^h}^2 - \sigma_{\gamma m}\right]\tau + \mathbb{E}\left[\gamma^h\eta^h m^h\right]\chi = \mathbb{E}\left[\lambda\tau^{X,h}m^h - \sigma_{\gamma,c/\Psi}\right]$$
$$\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]\tau + \mathbb{E}\left[\gamma^h {\eta^h}^2\right]\chi = \mathbb{E}\left[\lambda\tau^{X,h}\eta_h\right]$$

Solving for the two unknowns τ and χ gives the following.

Proposition 12.12 *The optimal tax and nudge satisfy*

$$\tau = \frac{\mathbb{E}\left[\gamma^h {\eta^h}^2\right]\mathbb{E}\left[\lambda\tau^{X,h}m^h - \sigma_{\gamma,c/\Psi}\right] - \mathbb{E}\left[\gamma^h\eta_h m^h\right]\mathbb{E}\left[\lambda\tau^{X,h}\eta^h\right]}{\mathbb{E}\left[\gamma^h {\eta^h}^2\right]\mathbb{E}\left[\gamma^h {m^h}^2 - \sigma_{\gamma m}\right] - \mathbb{E}\left[\gamma^h\eta^h m^h\right]\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]}$$
$$\chi = \frac{\mathbb{E}\left[\lambda\tau^{X,h}\eta^h\right]\mathbb{E}\left[\gamma_h {m^h}^2 - \sigma_{\gamma m}\right] - \mathbb{E}\left[\lambda\tau^{X,h}m^h - \sigma_{\gamma,c/\Psi}\right]\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]}{\mathbb{E}\left[\gamma^h {\eta^h}^2\right]\mathbb{E}\left[\gamma^h {m^h}^2 - \sigma_{\gamma m}\right] - \mathbb{E}\left[\gamma_h\eta^h m^h\right]\mathbb{E}\left[\gamma^h\eta^h m^h - \sigma_{\gamma,\eta}\right]}$$

12.5 Quadratic losses from imperfect tax instruments

We introduce the Lagrangian that allows for agent-specific lump-sum transfers w^h and taxes $\tau^h, \tau^{s,h}$:

$$L\left(\left\{\tau^h\right\}, \left\{\tau^{s,h}\right\}, \left\{w^h\right\}\right) := W\left(v^h\left(p+\tau, p+\tau^{s,h}, w^h, \xi\right)\right) + \lambda\sum_h \left[\tau\cdot c^h\left(p+\tau, p+\tau^{s,h}, w^h, \xi\right) - w^h\right]$$

with $\xi = \xi\left(\left\{c^h\right\}\right)$ as a fixed point. We also define:

$$g\left(\left\{\tau^{s,h}\right\}\right) := \max_{\left\{w^h\right\}} L\left(\left\{\tau^{s,h}\right\}, \left\{\tau^{s,h}\right\}, \left\{w^h\right\}\right) \tag{89}$$

which is the Lagrangian with rational agents perceiving $\tau^{s,h}$ and with optimum agent-specific lump-sum transfer.

The social utility achieved with agent-specific taxes $\tau^{s,h}$, and optimum agent-specific lump-sum transfers, with a rational agent.

Proposition 12.13 *In general, in the Ramsey problem with externalities and redistribution, the social loss (realized social minus first best) is:*

$$L = L^{distribution} + L^{distortions}$$

with

$$\begin{aligned} L^{distribution} &= \frac{1}{2}\sum_{h,h'} \left(\gamma^{\xi,h} - \overline{\gamma}\right)\left(L_{ww}\left(w,\tau\right)^{-1}\right)_{h,h'}\left(\gamma^{\xi,h} - \overline{\gamma}\right) \\ L^{distortions} &= \frac{1}{2}\sum_{h,h'} \left(\tau^{s,h} - \tau^{*s,h}\right) g_{\tau^{s,h}\tau^{s,h'}} \left(\tau^{s,h\prime} - \tau^{*s,h'}\right) \end{aligned}$$

This reflects that at the optimum, the $\gamma^{\xi h}$ should be the same (and equal to λ), and we should have $\tau^{s,h} - \tau^{*s,h}$.

Proof of Proposition 12.13 We note that for any tax system,

$$L\left(\left\{\tau^h\right\}, \left\{\tau^{s,h}\right\}, \left\{w^h\right\}\right) := L\left(\left\{\tau^{s,h}\right\}, \left\{\tau^{s,h}\right\}, \left\{w^h + \left(\tau^{s,h} - \tau^h\right)\cdot c^h\left(p+\tau^h, p+\tau^{s,h}, w^h, \xi\right)\right\}\right).$$

and

$$L\left(\left\{\tau^{s,h}\right\}, \left\{\tau^{s,h}\right\}, \left\{w^h\right\}\right) = W\left(v^{h,r}\left(p+\tau^{s,h}, w^h, \xi\right)\right) + \lambda\sum_h\left[\tau\cdot\overline{\overline{c}}^{r,h}\left(p+\tau^{s,h}, w^h, \xi\right) - w^h\right]$$

Here $w, w^* \in \mathbb{R}^H$. Call $y = \left(\left\{\tau^h\right\}, \left\{\tau^{s,h}\right\}\right) \in \mathbb{R}^{2nH}$ (with n the number of goods). The first best (in a world with externalities) has (w^*, y^*). We call $w^{**}(y)$ the optimal redistribution given a tax system y. So, $w^* = w^{**}(y^*)$.

$$\begin{aligned}
L^{\text{tot}} &= L(w,y) - L(w^*,y^*) \\
&= [L(w,y) - L(w^{**}(y),y)] + [L(w^{**}(y),y) - L(w^{**}(y^*),y^*)] \\
&= \frac{1}{2}(w - w^{**}(y)) \cdot L_{ww}(w^{**}(y),y) \cdot (w - w^{**}(y)) + \frac{1}{2}(y-y^*) g_{yy}(y-y^*) \text{ by Lemma 16.3} \\
&= L^{\text{distribution}} + L^{\text{distortion}} \\
L^{\text{distribution}} &= \frac{1}{2}(w - w^{**}(y)) \cdot L_{ww}(w^{**}(y),y) \cdot (w - w^{**}(y)) \\
L^{\text{distortion}} &= \frac{1}{2}(y-y^*) g_{yy}(y-y^*)
\end{aligned}$$

Redistribution terms

From Lemma 16.2, the expression of the loss involves $L_{w_h}(w,\tau) = \gamma^{\xi h} - \lambda$, the social marginal utility. Applying that Lemma 16.2 gives a loss:

$$L^{\text{distribution}} = \frac{1}{2}\sum_{h,h'} \left(\gamma^{\xi h} - \overline{\gamma}\right)\left(L_{ww}(w,\tau)^{-1}\right)_{h,h'}\left(\gamma^{\xi h} - \overline{\gamma}\right) \tag{90}$$

Tax distortion terms

We have $L^{\text{distortion}} = \frac{1}{2}(y-y^*) g_{yy}(y-y^*)$. Note that $g(y) := g(\{\tau^s\})$.

$$\begin{aligned}
g(\tau^s) &= \max_{w_1,\dots,w_n} L(w,\tau^s) = L(w^*(\tau^s),\tau^s) \\
g_{\tau^s\tau^s} &= L_{\tau^s\tau^s} - L_{\tau^s w} L_{ww}^{-1} L_{\tau^s w} \text{ by Lemma 16.3.}
\end{aligned}$$

□

Understanding the redistribution term For instance, take the case: $W = \sum v^h\left(q, q^{s,h}, w^h, \xi\right)$ and ξ is independent of w^h, then $L_{w^h w^{h'}} = v^h_{ww}$, so that

$$L^{\text{distribution}} = \frac{1}{2}\sum_{h,h'} \frac{\left(\gamma^h - \overline{\gamma}\right)^2}{v^h_{ww}}$$

The losses come from the lack of equalization of γ's.

Understanding the $g_{\tau^s\tau^s}$ better

Lemma 12.5 *We have*

$$g_{\tau^{s,h}\tau_{s,h'}} = \lambda S^{hr}\left(1_{h=h'} - \frac{d\tau^{\xi,h}}{d\tau^{s,h'}}\right)$$

When utility is quasi linear and the externality enters additively, $u(c,\xi) = u(c_1,\dots,c_n) + \lambda c_0 + \frac{1}{H}\xi$, *we have:*

$$g_{\tau_s^h \tau_{s,h'}} = \lambda S^{rh} 1_{h=h'} + S^{rh} \xi_{c^h c^{h'}} S^{rh'}. \tag{91}$$

Proof We observe that a tax τ^{sh} modifies the externality as:

$$\frac{d\xi\left(\left\{w^h\right\},\left\{\tau^{s,h}\right\}\right)}{d\tau^{s,h}} = \xi_{c^h} c^h_{\tau^{s,h}}\left(q,w,\xi\right) + \sum_{h'} \xi_{c^{h'}} c^{h'}_{\xi} \frac{d\xi}{d\tau^{s,h}}$$

so

$$\frac{d\xi\left(\left\{w^h\right\},\left\{\tau^{s,h}\right\}\right)}{d\tau^{s,h}} = \frac{\xi_{c^h} c^h_{\tau^{s,h}}}{1-\sum_{h'} \xi_{c^{h'}} c^{h'}_{\xi}} \tag{92}$$

Also $\frac{d\xi}{dw^h} = \xi_{c^h} c^h_{w^h}\left(q,w,\xi\right) + \sum_{h'} \xi_{c^{h'}} c^{h'}_{\xi} \frac{d\xi}{dw^h}$, so

$$\frac{d\xi}{dw^h} = \frac{\xi_{c^h} c^h_{w^h}}{1-\sum_{h'} \xi_{c^{h'}} c^{h'}_{\xi}} \tag{93}$$

We note that the FOC of (89) in w^h is

$$\begin{aligned}
0 &= v^h_{w_h} + \lambda\left(\tau^{s,h}\cdot c^{rh}_{w_h} - 1\right) + \frac{d\xi}{dw^h}\sum_{h'} v^{h'}_{\xi} + \lambda\tau^{s,h'}\cdot \overline{\overline{c}}^{rh'}_{\xi} \\
&= v^h_{w^h} + \lambda\left(\tau^{sh}\cdot c^{rh}_{w^h} - 1\right) + \frac{\xi_{c^h} c^h_{w^h}}{1-\sum_{h'} \xi_{c^{h'}} c^{h'}_{\xi}}\sum_{h'} v^{h'}_{\xi} + \lambda\tau^{sh'}\cdot \overline{\overline{c}}^{rh'}_{\xi} \text{ by (93)} \\
&= v^h_{w_h} + \lambda\left(\tau^{sh}\cdot c^{rh}_{w_h} - 1\right) + \xi_{c^h} c^h_{w^h}\Xi = v^h_{w^h} + \lambda\left(\tau^{s,h}\cdot c^{rh}_{w^h} - 1\right) - \lambda\tau^{\xi h} c^h_{w^h} \\
&= \gamma^{h,\xi} - \lambda
\end{aligned}$$

which confirms that at the optimum $\gamma^{h,\xi} = \lambda$ for all agents – even if the tax hasn't been optimized upon.

$$\begin{aligned}
g\left(\left\{\tau_{s,h}\right\}\right) &:= \max_{\left\{w^h\right\}} \sum_h v^r\left(p+\tau^{s,h}, w^h, \xi\right) + \lambda\sum_h\left[\tau^{s,h}\cdot \overline{\overline{c}}^r\left(p+\tau^{s,h}, w^h, \xi\right) - w^h\right] \\
&= \max_{\left\{w^h\right\}} \sum_h v^r\left(p+\tau^{s,h}, w^h, \xi\right) + \lambda\sum_h\left[p\cdot \overline{\overline{c}}^r\left(p+\tau^{s,h}, w^h, \xi\right)\right] \text{ [probably not useful]}
\end{aligned}$$

$$\left(p+\tau^h\right)c = w^h$$

We take the derivatives (89):

$$
\begin{aligned}
g_{\tau_{s,h}}\left(\{\tau_{s,h'}\}\right) &= \left(\lambda - v_w^h\right) c^{hr} + \lambda \tau^{s,h} c_p^{hr} + \frac{d\xi}{d\tau^{s,h}} \sum_{h'} \left[v_\xi^{h'} + \lambda \tau^{s,h'} \cdot \overline{\overline{c}}_\xi^r\right] \\
&= \left(\lambda - v_w^h\right) c^{hr} + \lambda \tau^{s,h} c_p^{hr} + \frac{\xi_{c^h} c_{\tau^{s,h}}^h}{1 - \sum_{h'} \xi_{c^{h'}} c_\xi^{h'}} \sum_{h'} v_\xi^{h'} + \lambda \tau^{s,h'} \cdot \overline{\overline{c}}_\xi^r \text{ by (92)} \\
&= \left(\lambda - v_w^h\right) c^{hr} + \lambda \tau^{s,h} c_p^{hr} + \xi_{c^h} c_{\tau^{s,h}}^h \Xi \\
&= \left(\lambda - v_w^h\right) c^{hr} + \lambda \tau^{s,h} c_p^{hr} - \lambda \tau^{\xi h} c_p^h \\
&= \left(\lambda - v_w^h\right) c^{hr} + \lambda \left(\tau^{sh} - \tau^{\xi h}\right) c_p^{hr} \\
&= \left(\lambda - v_w^h\right) c^{hr} + \lambda \left(\tau^{sh} - \tau^{\xi h}\right) \left[S^{hr} - c_w^r c^h\right] \\
&= \left[\lambda - v_w^h - \lambda \left(\tau^{s,h} - \tau^{\xi h}\right) c_w\right] c^h + \lambda \left(\tau^{s,h} - \tau^{\xi h}\right) S^{hr} \\
&= \lambda \left(\tau^{s,h} - \tau^{\xi h}\right) S^{hr}
\end{aligned} \tag{94}
$$

Hence, observing that $\tau^{s,h} - \tau^{\xi h} = 0$ at the optimum,

$$
g_{\tau^{s,h}\tau^{s,h'}} = \lambda S^{hr} \left(1_{h=h'} - \frac{d\tau^{\xi h}\left(\left\{\tau^{s,h''}\right\}\right)}{d\tau^{s,h'}}\right) \tag{95}
$$

Example with quasi-linear utility, additive externality

When $u(c,\xi) = u(c_1, ..., c_n) + \lambda c_0 + \frac{1}{H}\xi$, we have $c(p,\xi)$ independent of ξ, and $\Xi = \frac{\sum_h \left[\beta^h \frac{v_\xi^h}{v_w^h} + \lambda \tau \cdot \overline{\overline{c}}_\xi^h\right]}{1 - \sum_h \xi_{c^h} c_\xi^h} = 1$, and $\tau^{\xi h} := -\frac{1}{\lambda}\xi_{c^h}$.

So, $\frac{d\tau^{\xi,h}}{d\tau^{s,h'}} = -\frac{1}{\lambda}\xi_{c^h c^{h'}} \frac{dc^{h'}}{d\tau^{h'}} = -\frac{1}{\lambda}\xi_{c^h c^{h'}} S^{rh'}$:

$$
\frac{d\tau^{\xi,h}}{d\tau^{s,h'}} = -\frac{1}{\lambda}\xi_{c^h c^{h'}} S^{rh'}
$$

so

$$
g_{\tau^{s,h}\tau_{s,h'}} = \lambda S^{rh} 1_{h=h'} + H S^{rh} \xi_{c^h c^{h'}} S^{rh'} \tag{96}
$$

xx that should generalize to additive externality: $u(c,\xi) = u(c) + \frac{1}{H}\xi$. Then $\Xi = \frac{\sum_h \left[\beta^h \frac{v_\xi^h}{v_w^h} + \lambda \tau \cdot \overline{\overline{c}}_\xi^h\right]}{1 - \sum_h \xi_{c^h} c_\xi^h} = 1$. And $\tau^{\xi h} := -\frac{1}{\lambda}\xi_{c^h}\left(\left\{\tau^{sh}\right\}\right)$. When $\left\{\tau^{-h'}\right\}$ are held constant, varying $\tau^{h\prime}$ changing $c^{h'}$ and ξ, but doesn't change the marginal utility of the agent $-h'$, so doesn't change their consumption, so

$\frac{dc^{h''}}{d\tau^{h'}} = 0$ for $h'' \neq h'$,

$$\frac{d\tau^{\xi,h}}{d\tau^{s,h'}} = -\frac{1}{\lambda}\xi_{c^h c^{h'}}\frac{dc^{h'}}{d\tau^{h'}} = -\frac{1}{\lambda}\xi_{c^h c^{h'}}S^{rh'} + \sum_{h''} -\frac{1}{\lambda}\xi_{c^h c^{h''}}\frac{\partial c^{h''}}{dw^{h''}}\frac{dw^{h''}}{d\tau^{h'}}$$
$$= -\frac{1}{\lambda}\xi_{c^h c^{h'}}\left(S^{rh'} - c^{h'}_{w^{h'}}c + c^{h'}_{w^{h'}}\frac{dw^{h'}}{d\tau^{h'}}\right)$$
$$= -\frac{1}{\lambda}\xi_{c^h c^{h'}}\left[S^{rh'} - c^{h'}_{w^{h'}}\left(c - \frac{dw^{h'}}{d\tau^{h'}}\right)\right]$$

as $\gamma^{h,\xi} := v^h_w + \lambda\left(\tau^{sh} - \tau^{\xi h}\right)\cdot c^h_w = \lambda$ implies $dv^h_w + \lambda\left(d\tau^{sh}\right)\cdot c_w = 0$.

12.6 Mirrlees problems with extensive margin

We provide a behavioral enrichment to Saez (2002). We take his simplest framework (Proposition 1). Activity 0 is unemployment, and there are I other activities. One type i of agent chooses between working and not working: working gives utility $u^h(c_i, i)$, not working utility $u^h(c_0, 0)$, where $c_i = z_i - T(z_i)$. If the agent is rational, he solves

$$i^* = \arg\max_{i^* \in \{0,i\}} u^h(c_i, i)$$

but our behavioral agent may make a mistake. E.g., in the misperception model, he might perceive c^s_i, so that he decides according to

$$i^* = \arg\max_{i^* \in \{0,i\}} u^{h,b}(c^s_i, i).$$

In general, we will simply model the choice as some $i^*(h, \{T_j\})$. We say that an agent is "at the margin for tax i" if the agent changes activity as tax i changes $B^+_i := \{m \text{ s.t. } \partial i^*(h, \{T_j\})/\partial T_i < 0\}$ (which is the set of agent moving into active employment if the tax rate on activity i falls) and $B^-_i := \{m \text{ s.t. } \partial i^*(h, \{T_j\})/\partial T_i > 0\}$ (which is the set of agent moving out of active employment if the tax rate on activity i falls). The normal case is that B^-_i is an empty set. The derivative is in the sense of distributions, and simply indicates a change in agent's behavior.

Suppose that the government increases tax T_i on activity i by dT_i. That induces a quantity dH_j of people to switch to employment, where

$$dH_j = -H_j\eta_{ji}\frac{dT_i}{c_i - c_0}$$

We have $h_j(\{T_k\})$ =number of agents of type j who work.

Each h has a potential earnings level $j(h)$. We call

$$\tau^b_{ji} := -\sum_{\varepsilon\in\{-,+\}} \mathbb{E}\left[\varepsilon\mu^m\left(u^h(c_j,j)-u^h(c_0,0)\right) \mid j(h)=j \text{ and } h\in B_i^\varepsilon\right]$$

We have

$$\frac{\partial H^j}{\partial T_i} = -\sum_{\varepsilon\in\{-,+\}}\int \varepsilon 1_{\{j(h)=j \text{ and } h\in B_i^\varepsilon\}} d\nu(m)$$

The change in welfare from dT_i is then

$$\begin{aligned}
dL &= (1-g_i)H_i dT_i - \sum_j\sum_{\varepsilon\in\{-,+\}}\int \varepsilon\left[T_j - T_0 + \mu^m\left(u^h(c_j,j)-u^h(c_0,0)\right)\right] 1_{\{j(h)=j \text{ and } h\in B_i^\varepsilon\}} d\nu(m)\\
&= (1-g_i)H_i dT_i + \sum_j (T_j-T_0)\frac{\partial H^j}{\partial T_i} - \sum_j\sum_{\varepsilon\in\{-,+\}}\int \varepsilon\left[\mu^m\left(u^h(c_j,j)-u^h(c_0,0)\right)\right] 1_{\{j(h)=j \text{ and } h\in B_i^\varepsilon\}} d\nu(m)\\
&= (1-g_i)H_i dT_i + \sum_j \left(T_j - T_0 - \tau^b_{ji}\right)\frac{\partial H^j}{\partial T_i}\\
&= (1-g_i)H_i dT_i - \sum_j \left(T_j - T_0 - \tau^b_{ji}\right) H_j \eta_{ji}\frac{dT_i}{c_i - c_0}
\end{aligned}$$

Hence, at the optimum:

$$\sum_j \frac{T_j - T_0 - \tau^b_{ji}}{c_i - c_0}\frac{H_j}{H_i}\eta_{ji} = (1-g_i)$$

For instance, suppose that people overestimate taxes, i.e. underperceive the benefits from working: $c_i^s < c_i$, and no cross-effects. Then,

$$\tau^b_{ji} = -1_{j=i}\sum_{\varepsilon\in\{-,+\}} \mathbb{E}\left[\varepsilon\mu^m\left(u^h(c_i,i)-u^h(c_0,0)\right) \mid j(h)=j \text{ and } h\in B_i^\varepsilon\right]$$

12.7 Supply Elasticities: Mirrlees case

We now verify that the logic of section 7.1 applies to the Mirrlees case: with behavioral agent, the supply elasticities generally are featured in the optimal income tax formula. To make the point, take the case where the aggregate constraint is: $\Phi\left(\int nL(n)f(n)\,dn\right) \le g + \int c(n)f(n)\,dn$, where $\Phi(L)$ is an aggregate production function. Indeed, recall that in the Mirrlees framework an agent of productivity n is considered to supply n units of effective labor. Call $w = \Phi'(L)$ the wage rate. We can extend the analysis of section 5, with an index w (which can be thought as being normalized to $w=1$ in that section). Then, we can write the agents' utility function problem as $u^n(c,z,w) = U\left(c,\frac{z}{nw}\right)$ and the earnings supply as $z^n(q,Q,r_0,r,w) = wnL^n(q,Q,r_0,r,w)$. Other functions acquire a w term, e.g. indirect utility becomes $v^n(q,Q,r_0,r,w)$. Given a tax system, the equilibrium wage w satisfies:

$$w = \Phi' \left(\int \frac{z^n (q, Q, r_0, r, w)}{w} f(n)\, dn \right) \tag{97}$$

which defines an equilibrium wage $w(Q, r_0)$

The objective function is:

$$L(Q, r_0, w) = \int W(v(n)) f(n)\, dn + \lambda \left[\Phi \left(\int \frac{z(n)}{w} f(n)\, dn \right) - \int c(n) f(n)\, dn \right]$$

and we can define $L(Q, r_0) = L(Q, r_0, w(Q, r_0))$ when taking into account the equilibrium wage w.

Proposition 12.14 *In the Mirrlees model, suppose that the production function is imperfectly elastic. Then, the optimum tax features $L_{q_{z^*}}(Q, r_0) = 0$, with*

$$L_{Q_{z^*}}(Q, r_0) = L_{Q_{z^*}}(Q, r_0, w) + L_w(Q, r_0, w)\, w_{Q_{z^*}}(Q, r_0) \tag{98}$$

The term $L_{Q_{z^*}}(Q, r_0, w)$, with fixed wage, was calculated in Proposition 13.1 (with the normalization $w = 1$). Hence, the optimal tax formula $w_{Q_{z^*}}(Q, r_0)$ generally depends on production elasticity, and does not coincide with the one in Section 7.1.

When agents are rational, one can verify that $L_w(Q, r_0, w) = 0$ at the optimum (see the proof of Proposition 12.14). Hence, in the traditional analysis, the supply elasticity (captured by $w_{Q_{z^*}}(Q, r_0)$) doesn't appear in the optimal tax formula. This is not true any more with a behavioral model.

Proof of Proposition 12.14 The tax formula in the Proposition follows from the Chain rule. Next, we verify that when agents are rational, $L_w = 0$ at $w = w_0$. We normalize $w_0 = 1$ for simplicity. Suppose a given value of w and $R(z)$. Define $\widetilde{R}(z', w) = R(z'w)$. Then, as $z^n = \arg\max_z U^n \left(R(z), \frac{z}{nw}\right)$, i.e.

$$\frac{z}{w} = \arg\max U^n \left(\widetilde{R}\left(\frac{z}{w}, w\right), \frac{z}{nw} \right)$$

So $L(R(\cdot), w) = L\left(\widetilde{R}(\cdot, w), 1\right)$. That is, the welfare is the same as if we had a different tax system $\widetilde{R}$, and a wage $w = 1$. Thus, given we started at an optimum tax system ($R^0(\cdot) = \arg\max_{R(\cdot)} L(R(\cdot), 1)$), we have $L_{\widetilde{R}}\left(\widetilde{R}, 1\right) = 0$, hence $L_w = 0$.□

13 Proofs not included in the paper

13.1 Proofs of the Ramsey and Pigou cases

Proof of Lemma 2.1 We give two proofs of this result. The first, and most elementary one, is that this is a Corollary to Lemma 12.4, using the behavioral elasticity $\alpha_i = m_i \psi_i$.

The other proof is as follow. Because utility is quasilinear, $v(\boldsymbol{p}, \boldsymbol{p}^s, w) = w + v(\boldsymbol{p}, \boldsymbol{p}^s, 0)$, so $e(\boldsymbol{p}, \boldsymbol{p}^s, 0) = -v(\boldsymbol{p}, \boldsymbol{p}^s, 0)$. We have

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\tau}) &= u(\boldsymbol{c}) + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c} \\
&= v(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, w) + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, w) \\
&= w - e(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, 0) + (1+\Lambda)\boldsymbol{\tau} \cdot \boldsymbol{c}(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, w)
\end{aligned}$$

By Taylor expansion, around $(\boldsymbol{\tau}, \boldsymbol{\tau}^s) = (0,0)$, using Propositions 14.1 and 14.5, we have:

$$\begin{aligned}
e(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, 0) - e(\boldsymbol{p}, \boldsymbol{p}) &= [e_{\boldsymbol{p}}\boldsymbol{\tau} + e_{\boldsymbol{p}^s}\boldsymbol{\tau}^s] + \left[\frac{1}{2}\boldsymbol{\tau} e_{\boldsymbol{pp}}\boldsymbol{\tau} + \boldsymbol{\tau} e_{\boldsymbol{p},\boldsymbol{p}^s}\boldsymbol{\tau}^s + \frac{1}{2}\boldsymbol{\tau}^{s\prime} e_{\boldsymbol{p}^s\boldsymbol{p}^s}\boldsymbol{\tau}^s\right] + o\left(\|\boldsymbol{\tau}\|^2\right) \\
&= \left[\boldsymbol{c}^d\boldsymbol{\tau} + 0\cdot\boldsymbol{\tau}^s\right] + \left[0 + \boldsymbol{\tau}\boldsymbol{S}^r\boldsymbol{\tau}^s + \frac{1}{2}\boldsymbol{\tau}^{s\prime}\boldsymbol{S}^r\boldsymbol{\tau}^s\right] + o\left(\|\boldsymbol{\tau}\|^2\right) \\
&= \boldsymbol{c}^d\boldsymbol{\tau} + \boldsymbol{\tau}\boldsymbol{S}^r\boldsymbol{\tau}^s + \frac{1}{2}\boldsymbol{\tau}^{s\prime}\boldsymbol{S}^r\boldsymbol{\tau}^s + o\left(\|\boldsymbol{\tau}\|^2\right)
\end{aligned}$$

Using Proposition 14.3,

$$\begin{aligned}
\boldsymbol{\tau} \cdot \boldsymbol{c}(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, w) &= \boldsymbol{\tau} \cdot \left(\boldsymbol{c}^d + \boldsymbol{c}_{\boldsymbol{p}}\boldsymbol{\tau} + \boldsymbol{c}_{\boldsymbol{p}^s}\boldsymbol{\tau}^s\right) \\
&= \boldsymbol{\tau} \cdot \left(\boldsymbol{c}^d + 0 + \boldsymbol{S}^r\boldsymbol{\tau}^s + o(\|\boldsymbol{\tau}\|)\right) = \boldsymbol{\tau}\boldsymbol{c}^d + \boldsymbol{\tau}'\boldsymbol{S}^r\boldsymbol{\tau}^s + o\left(\|\boldsymbol{\tau}\|^2\right)
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\tau}) - \mathcal{L}(0) &= -\left[e(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, 0) - e(\boldsymbol{p}, \boldsymbol{p})\right] + (1+\Lambda)\boldsymbol{\tau} \cdot \boldsymbol{c}(\boldsymbol{p}+\boldsymbol{\tau}, \boldsymbol{p}+\boldsymbol{\tau}^s, w) \\
&= -\boldsymbol{c}^d\boldsymbol{\tau} - \boldsymbol{\tau}\boldsymbol{S}^r\boldsymbol{\tau}^s - \frac{1}{2}\boldsymbol{\tau}^{s\prime}\boldsymbol{S}^r\boldsymbol{\tau}^s + (1+\Lambda)\left(\boldsymbol{\tau}\boldsymbol{c}^d + \boldsymbol{\tau}'\boldsymbol{S}^r\cdot\boldsymbol{\tau}^s\right) + o\left(\|\boldsymbol{\tau}\|^2\right) \\
&= \Lambda\left(\boldsymbol{\tau}\boldsymbol{c}^d + \boldsymbol{\tau}'\boldsymbol{S}^r\cdot\boldsymbol{\tau}^s\right) - \frac{1}{2}\boldsymbol{\tau}^{s\prime}\boldsymbol{S}^r\boldsymbol{\tau}^s + o\left(\|\boldsymbol{\tau}\|^2\right) \\
&= \Lambda\boldsymbol{\tau}\boldsymbol{c}^d - \frac{1}{2}\boldsymbol{\tau}^{s\prime}\boldsymbol{S}^r\boldsymbol{\tau}^s + o\left(\|\boldsymbol{\tau}\|^2\right) + O\left(\|\boldsymbol{\tau}\|^2\Lambda\right)
\end{aligned}$$

Proof of Proposition 2.3 *Optimal Pigouvian tax.* At the optimum, $U^{h\prime}\left(c^{h*}\right) = \boldsymbol{p} + \xi$. If the agent perceives only $m^h\tau$, his demand is off the ideal c^{h*} (up to second order terms) as:

$$c^h = c^{h*} - \Psi\left(m^h\tau - \xi^h\right)$$

This expression is exact in the quadratic functional form about, and otherwise the leading term of a Taylor expansion of a general function, with now the interpretation $\Psi = \psi^h c^{h*}$ then. So the welfare loss is:

$$L^h = W^h - W^{h*} = \frac{1}{2} u^{h\prime\prime} \cdot \left(-\Psi \cdot \left(m\tau - \xi^h\right)\right)^2 = -\frac{1}{2}\Psi\left(m^h\tau - \xi^h\right)^2$$

and social welfare is $L = \sum_h L^h = -\frac{\Psi}{2}\sum_h \left(m^h\tau - \xi^h\right)^2$

Because $L_\tau = -\Psi \sum_h m^h \left(m^h\tau - \xi^h\right)$, the optimal tax is

$$\tau^* = \frac{\sum_h \xi_h m^h}{\sum_h m_h^2} = \frac{\mathbb{E}\left[\xi_h m^h\right]}{\mathbb{E}\left[m^{h2}\right]}.$$

Let us calculate $V = \mathbb{E}\left[\left(m^h\tau - \xi^h\right)^2\right]$ at this optimum $\tau = \tau^*$,

$$\begin{aligned}
V &= \mathbb{E}\left[m^{h2}\right]\tau^{*2} - 2\mathbb{E}\left[m^h\xi^h\right]\tau^* + \mathbb{E}\left[\xi^{h2}\right] \\
&= \mathbb{E}\left[m^{h2}\right]\frac{\mathbb{E}\left[\xi_h m^h\right]^2}{\mathbb{E}\left[m^{h2}\right]^2} - 2\mathbb{E}\left[m^h\xi_h\right]\frac{\mathbb{E}\left[\xi_h m^h\right]}{\mathbb{E}\left[m^{h2}\right]} + \mathbb{E}\left[\xi_h^2\right] = -\frac{\mathbb{E}\left[\xi_h m^h\right]^2}{\mathbb{E}\left[m^{h2}\right]} + \mathbb{E}\left[\xi_h^2\right] \\
&= \frac{\mathbb{E}\left[\xi_h^2\right]\mathbb{E}\left[m^{h2}\right] - \mathbb{E}\left[\xi_h m^h\right]^2}{\mathbb{E}\left[m^{h2}\right]}.
\end{aligned}$$

hence the welfare loss is: $L = -\frac{1}{2}H\Psi\frac{\mathbb{E}\left[\xi_h^2\right]\mathbb{E}\left[m^{h2}\right] - \left(\mathbb{E}\left[\xi_h m^h\right]\right)^2}{\mathbb{E}\left[m^{h2}\right]}$.

If there is no tax, the loss is (from equation 8):

$$L^{\text{no tax}} = -\frac{\Psi}{2}\sum_h \left(m^h \cdot 0 - \xi^h\right)^2 = -\frac{\Psi}{2}\sum_h \xi_h^2 = -\frac{1}{2}H\Psi\mathbb{E}\left[\xi_h^2\right].$$

So, $L = L^{\text{no tax}}\frac{\mathbb{E}\left[\xi_h^2\right]\mathbb{E}\left[m^{h2}\right] - \left(\mathbb{E}\left[\xi_h m^h\right]\right)^2}{\mathbb{E}\left[m^{h2}\right]\mathbb{E}\left[\xi_h^2\right]}$.

Optimal quantity mandate. Welfare is $\sum_h \left[U^h\left(c^*\right) - \left(p + \xi^h\right)c^*\right]$. The optimal quantity restriction c^* is characterized by:

$$\frac{1}{H}\sum_h U^{h\prime}\left(c^*\right) = p + \frac{1}{H}\sum_h \xi^h \tag{99}$$

The welfare loss compared to the first best, which entails $U^{h\prime}\left(c^{h*}\right) = p + \xi^h$ is

$$L^h = \frac{1}{2}U^{h\prime\prime}(c)\left(c^{h*} - c^*\right)^2 = -\frac{1}{2}\frac{1}{\Psi}\left(c^{h*} - c^*\right)^2$$

The best consumption satisfies: $L^Q_{c^*} = \frac{1}{2}\sum_h \frac{1}{\Psi}\left(c^{h*} - c^*\right) = 0$, i.e. $c^* = \mathbb{E}\left[c^{h*}\right]$

The loss is:

$$L^Q = -\frac{1}{2}\frac{H}{\Psi}\mathbb{E}\left[\left(c^{h*} - c^*\right)^2\right] = -\frac{1}{2}\frac{H}{\Psi}var\left(c^{h*}\right)$$

Proof of Proposition 3.2 We observe that a tax τ_i modifies the externality as:

$$\frac{d\xi}{d\tau_i} = \sum_h \xi_{c^h}\left[c^h_{q_i}\left(\boldsymbol{q}, w, \xi\right) + c^h_\xi \frac{d\xi}{d\tau_i}\right]$$

so $\frac{d\xi}{d\tau_i} = \frac{\sum_h \xi_{c^h} c^h_{q_i}}{1-\sum_h \xi_{c^h} c^h_\xi}$. The term $\frac{1}{1-\sum_h \xi_{c^h} c^h_\xi}$ represents the "multiplier" effect of one unit of pollution on consumption, then on more pollution. So, calling $\frac{\partial L}{\partial \tau_i}^{\text{no }\xi}$ the value of $\frac{\partial L}{\partial \tau_i}$ without the externality (that was derived in Proposition 3.1)

$$\begin{aligned}\frac{\partial L}{\partial \tau_i} - \frac{\partial L}{\partial \tau_i}^{\text{no }\xi} &= \frac{d\xi}{d\tau_i}\left\{\sum_h W_{v^h} v^h_w \frac{v^h_\xi}{v^h_w} + \lambda \sum_h \boldsymbol{\tau}\cdot c^h_\xi\left(\boldsymbol{q}, w, \xi\right)\right\} = \frac{d\xi}{d\tau_i}\sum_h\left[\beta^h \frac{v^h_\xi}{v^h_w} + \lambda\boldsymbol{\tau}\cdot c^h_\xi\right] \\ &= \frac{\sum_h \xi_{c^h} c^h_{q_i}}{1-\sum_h \xi_{c^h} c^h_\xi}\sum_h\left[\beta^h \frac{v^h_\xi}{v^h_w} + \lambda\boldsymbol{\tau}\cdot c^h_\xi\right] = \Xi\sum_h \xi_{c^h} c^h_{q_i}\end{aligned}$$

Using Proposition 3.1,

$$\begin{aligned}\frac{\partial L}{\partial \tau_i} &= \sum_h [(\lambda - \gamma^h)c^h_i + \lambda\boldsymbol{\tau}\cdot \boldsymbol{S}^{C,h}_i - \beta^h \boldsymbol{\tau}^{b,h}\cdot \boldsymbol{S}^{C,h}_i + \Xi\xi_{c^h}\cdot(-c^h_w c^h_i + \boldsymbol{S}^{C,h}_i)] \\ &= \sum_h [(\lambda - \gamma^h - \Xi\xi_{c^h}\cdot c^h_w)c^h_i + \lambda(\boldsymbol{\tau} + \frac{\Xi}{\lambda}\xi_{c^h})\cdot \boldsymbol{S}^{C,h}_i - \beta^h\boldsymbol{\tau}^{b,h}\cdot \boldsymbol{S}^{C,h}_i].\end{aligned}$$

Proof of Proposition 3.4 We have, from Proposition 12.1

$$\boldsymbol{\tau}^{b,h} = u^{s,h}_{\boldsymbol{C}}\left(\boldsymbol{C}^h\right) - u^h_{\boldsymbol{C}}\left(\boldsymbol{C}^h\right) + \boldsymbol{p} - \boldsymbol{p}^{s,h} + \boldsymbol{\tau} - \boldsymbol{\tau}^{s,h} = \boldsymbol{\tau}^{I,h} + \boldsymbol{\tau} - \boldsymbol{\tau}^{s,h} = \boldsymbol{\tau}^{I,h} + \left(I - \boldsymbol{M}^h\right)\boldsymbol{\tau}$$

hence

$$\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h} - \frac{\gamma^{\xi,h}}{\lambda}\boldsymbol{\tau}^{b,h} = \boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h} - \frac{\gamma^{\xi,h}}{\lambda}\left(\boldsymbol{\tau}^{I,h} + \left(I - \boldsymbol{M}^h\right)\boldsymbol{\tau}\right) = [I - \left(I - \boldsymbol{M}^h\right)\frac{\gamma^{\xi,h}}{\lambda}]\boldsymbol{\tau} - \boldsymbol{\tau}^{X,h}$$

Hence, Proposition 3.2 implies:

$$\sum_h (1 - \frac{\gamma^{\xi,h}}{\lambda})\boldsymbol{c}^h = -\sum_h \left(\boldsymbol{S}^{H,h}\right)'\left(\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi h} - \tilde{\boldsymbol{\tau}}^{b,\xi,h}\right) = -\sum_h \boldsymbol{M}^{h\prime}\boldsymbol{S}^{h,r}[[I - \left(I - \boldsymbol{M}^h\right)\frac{\gamma^{\xi,h}}{\lambda}]\boldsymbol{\tau} - \boldsymbol{\tau}^{X,h}]. \tag{100}$$

□

Proof of Proposition 4.1 We derived: $\boldsymbol{\tau} = \mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r}\boldsymbol{M}^{h}\right]^{-1}\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r}\right](\xi_*, 0)'$. We have $\left(\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r}\boldsymbol{M}^{h}\right]\right)_{ij} = \boldsymbol{S}^{r}_{ij}\mathbb{E}\left[m_i^h m_j^h\right]$ and $\left(\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r}\right]\right)_{ij} = \mathbb{E}\left[m_i^h\right]S^r_{ij}$. Matrix inversion gives:

$$\tau_2 = \frac{S^r_{11}S^r_{12}\mathbb{E}\left[m_1\right]\left(\mathbb{E}\left[m_1^2\right]\mathbb{E}\left[m_2\right] - \mathbb{E}\left[m_1 m_2\right]\mathbb{E}\left[m_1\right]\right)}{\det\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r}\boldsymbol{M}^{h}\right]}\xi_*$$

Because $\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r}\boldsymbol{M}^{h}\right]$ is a dimension 2×2 and has negative roots (there is a good 0, so that $\boldsymbol{S}^r$ is the block matrix excluding good 0, and has only negative root), $\det\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^{r}\boldsymbol{M}^{h}\right] > 0$. The condition in the Proposition is that $\mathbb{E}\left[m_1^2\right]\mathbb{E}\left[m_2\right] - \mathbb{E}\left[m_1 m_2\right]\mathbb{E}\left[m_1\right] > 0$. Hence, $sign\left(\tau_2\right) = -sign\left(S_{12}\right)$.

Proof of Proposition 4.5 We have apply our tax formulas (23):

$$\begin{aligned}
\frac{\partial L}{\partial \chi} &= -\sum_h \iota^h\gamma^h c^h - \sum_h \left[\left(\lambda - \gamma^h\left(1 - m^h\right)\right)\tau - \lambda\tau^{X,h} + \gamma^h\chi\eta^h\right]\Psi\tau_\chi^{\chi,h} \\
&= -\sum_h \iota^h\gamma^h c^h - \Psi\sum_h \left[\left(\lambda - \gamma^h\left(1 - m^h\right)\right)\tau - \lambda\tau^{X,h} + \gamma^h\chi\eta^h\right]\eta^h && (101) \\
&= \sum_h \left[-\iota^h\gamma^h c^h - \Psi x^h\eta^h\right] && (102)
\end{aligned}$$

where

$$x^h := \left(\lambda - \gamma^h\left(1 - m^h\right)\right)\tau - \lambda\tau^{X,h} + \gamma^h\chi\eta^h$$

Likewise,

$$\frac{\partial L}{\partial \tau} = \sum_h \left[\left(\lambda - \gamma_h\right)c^h - \Psi x^h m^h\right]$$

The problem is $\max_{\chi,\tau} L\left(\tau,\chi\right)$ s.t. $\chi \geq 0, \tau \geq 0$. The Lagrangian is:

$$L^*\left(\tau,\chi\right) := L\left(\tau,\chi\right) + \pi\chi + \pi'\tau$$

where π, π' are Lagrangian multipliers.

We observe that when there is no intervention ($\tau = \chi = 0$), then $x^h < 0$.

$$\begin{aligned}
\frac{L_\chi}{\eta^h} &= -\frac{\iota^h\gamma^h}{\eta^h}c^h - \Psi x^h \\
\frac{L_\tau}{m^h} &= \frac{\lambda - \gamma^h}{m^h}c^h - \Psi x^h
\end{aligned}$$

so

$$\frac{L_\tau}{m^h} - \frac{L_\chi}{\eta^h} = \left[\frac{\lambda - \gamma^h}{m^h} + \frac{\iota^h\gamma^h}{\eta^h}\right]c^h =: \Delta$$

If the optimum features $\chi > 0, \tau = 0$, then $L_\tau = -\pi' < 0 = L_\chi$, which implies $\Delta < 0$.

If the optimum features $\chi = 0, \tau > 0$, then $L_\chi = -\pi < 0 = L_\tau$, which implies $\Delta > 0$.

If the optimum features $\chi = \tau = 0$, then $L_\chi = -\pi < 0$, $L_\tau = -\pi' < 0$, so $\Psi x^h > \max\left(-\iota^h \gamma^h c^h, \left(\lambda - \gamma^h\right) c^h\right)$. This implies in particular that $\lambda < \gamma^h$.

We note that if the problem had with no inequality constraints, and just one type of agent, then an interior solution features: $\lambda - \gamma^h = -\iota^h \gamma^h$. there is a large subsidy in place, (to help the agent), and the excess consumption is corrected via the nudge. That is, the policy is to "Subsidize the poor, and nudge them away from the good at the same time". This results is a bit knife-edge.

Proof of Proposition 4.6 It easy to see that

$$\frac{\partial^2 L}{\partial A^h \partial \tau_z} = (1-\gamma)\left(A^h\right)^{-\sigma}\left(\bar{z} - z^h\right)\left[z^h + \tau_z\left(\bar{z} - z^h\right)\right]^{-\sigma}.$$

If $\sigma > 1$, then the sign of $\frac{\partial^2 L}{\partial A^h \partial \tau_z}$ is that of $\bar{z} - z^h$. As a result, larger behavioral biases (reductions in A^h) for the poor (agents with $z^h < \bar{z}$) lead to more redistribution (higher taxes τ_z). Conversely if $\sigma < 1$, then the sign of $\frac{\partial^2 L}{\partial A^h \partial \tau_z}$ is that of $\bar{z} - z^h$. As a result, larger behavioral biases for the poor lead to less redistribution. Finally if $\sigma = 1$ then $\frac{\partial^2 L}{\partial A^h \partial \tau_z} = 0$. As a result redistribution is independent of behavioral biases. □

Proof of Proposition 12.7 We use the extended utility function $v^h(\boldsymbol{q}, \boldsymbol{\omega})$ and demand function $\boldsymbol{c}^h(\boldsymbol{q}, \boldsymbol{\omega})$. We use the Roy's identify with mental accounts, Proposition 12.8:

$$\begin{aligned}
\frac{\partial L}{\partial \tau_i} &= \sum_h [W_{v^h} v^h_{\omega^{k(i)}} \frac{v^h_{q_i}}{v^h_{\omega^{k(i)}}} + \sum_k W_{v^h} v^h_{\omega^k} \omega^k_{q_i} + \lambda c^h_i + \lambda \boldsymbol{\tau} \cdot [\boldsymbol{c}^h_{q_i} + \sum_k \boldsymbol{c}^h_{\omega^k} \omega^k_{q_i}]], \\
&= \sum_h [-\beta^{k(i),h}\left(c^h_i + \boldsymbol{\tau}^{b,k(i)} \cdot \boldsymbol{S}^{C,k(i),h}_i\right) + \sum_k \beta^{k,h} \omega^k_{q_i} + \lambda c^h_i + \lambda \boldsymbol{\tau} \cdot \left(\left(\boldsymbol{S}^{C,k(i),h}_i - \boldsymbol{c}^{k(i),h}_{\omega^{k(i)}} c_i\right) + \sum_k \boldsymbol{c}^h_{\omega^k} \omega^k_{q_i}\right)], \text{ u} \\
&= \sum_h [\left(\lambda - \beta^{k(i),h} - \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}^{k(i),h}_{\omega^{k(i)}}\right) c^h_i + \sum_k \left[\beta^{k,h} + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}^h_{\omega^k}\right] \omega^k_{q_i} + \lambda\left(\boldsymbol{\tau} - \tilde{\boldsymbol{\tau}}^{b,k(i),h}\right) \cdot \boldsymbol{S}^{C,k(i),h}_i], \\
&= \sum_h [\left(\lambda - \gamma^{k(i),h}\right) c^h_i + \sum_k \gamma^{k,h} \omega^{k,h}_{q_i} + \lambda\left(\boldsymbol{\tau} - \tilde{\boldsymbol{\tau}}^{b,k(i),h}\right) \cdot \boldsymbol{S}^{C,k(i),h}_i], \\
\frac{\partial L}{\partial \tau_i} &= \sum_h [\left(\lambda - \gamma^{k(i),h}\right) c^h_i + \lambda\left(\boldsymbol{\tau} - \tilde{\boldsymbol{\tau}}^{b,k(i),h}\right) \cdot \boldsymbol{S}^{C,k(i),h}_i + \sum_k \gamma^{k,h} \omega^{k,h}_{q_i}].
\end{aligned}$$

□

Proof of Proposition 4.7

$$0 = q_i \frac{\partial L}{\partial \tau_i} = \sum_h [\left(\lambda - \gamma^{k,h}\right) q_i c_i^h + \lambda \boldsymbol{\tau} \cdot \boldsymbol{S}_i^{C,k,h} q_i]$$
$$= \sum_h [\left(\lambda - \gamma^{k,h}\right) q_i c_i^h + \lambda \sum_j \tau_j S_{ji}^{C,k,h} q_i]$$

Summing over i in the account gives:

$$0 = \sum_h [\left(\lambda - \gamma^{k,h}\right) \sum_i q_i c_i^h + \lambda \sum_j \tau_j \sum_i S_{ji}^{C,k,h} q_i]$$

By the traditional Slutsky relation with account $\sum_i S_{ji}^{C,k,h} q_i = 0$ for all j, h, so

$$0 = \sum_h [\left(\lambda - \gamma^{k,h}\right) \sum_i q_i c_i^h] = 0$$

With just one type of agent h, this gives $\lambda - \gamma^k = 0$. This implies, for all i :

$$\sum_j \tau_j S_{ji}^{C,k} q_i = 0$$

This first order condition is verified by $\tau_j = \tau$ for all j. For a generic Slutsky matrix, the only solution of $x \cdot S^C = 0$ is $x = tq$ for some real t (we do not have a proof of this, but this is highly likely). This implies that $\tau_j = \tau$ for some τ. $\square$

Proof of Proposition 7.1 We compute the derivatives of the Lagrangian:

$$\frac{\partial L}{\partial \tau_i^\kappa} = \sum_h \left[W_{v^h} \left(v_{\tau_i^\kappa}^h + v_{q^p}^h \cdot p_{\tau_i^\kappa}\right) - \mu p \cdot \left(c_{\tau_i^\kappa}^h + c_{q^p}^h \cdot p_{\tau_i^\kappa}\right)\right].$$

To calculate this, let us make an analogy with our basic Ramsey model with fixed prices. We expressed it $L = W + \lambda \sum_h \left(\boldsymbol{\tau} \cdot c^h - w\right)$, and it can be re-expressed:

$$L = W + \lambda \sum_h \left(\boldsymbol{\tau} \cdot c^h - w\right) = W + \lambda \sum_h \left[(p + \tau) \cdot c^h - w - p \cdot c^h\right]$$
$$= W - \lambda \sum_h p \cdot c^h$$

as $q \cdot c^h - w = 0$. So

$$L_{\tau_i^\kappa}^{\text{fixed price}} = \frac{\partial \left(W - \lambda \sum_h p \cdot c^h\right)}{\partial \tau_i^\kappa} = \sum_h \left[W_{v^h} v_{\tau_i^\kappa}^h - \mu p \cdot c_{\tau_i^\kappa}^h\right]$$
$$= \sum_h [\left(\lambda - \gamma^h\right) c_i^h + \lambda(\bar{\boldsymbol{\tau}} - \widetilde{\boldsymbol{\tau}}^{b,h}) \cdot S_i^{H,\kappa,h}] \tag{103}$$

We then have

$$L_{\tau_i^\kappa} = L_{\tau_i^\kappa}^{\text{fixed price}} + \sum_j L_{\tau_j^p} \varepsilon_{ji}^\kappa = L_{\tau_i^\kappa}^{\text{fixed price}} + L_{\tau^p} \cdot \varepsilon_i^\kappa. \tag{104}$$

□

Proof of Proposition 7.3 *Case of an inattentive consumer.* Call $q_a = p_a + \tau_a$. Equilibrium requires $q_a = q_b = 1$. Competitive pricing in good 1 requires that firms choose inputs according to: $\max_{l_{ia}, l_{ib}} p_i \left(\frac{l_{ia}}{\alpha_i}\right)^{\alpha_i} \left(\frac{l_{ib}}{1-\alpha_i}\right)^{1-\alpha_i} - (1+\tau_{ia}) l_{ia} - l_{ib}$ with $\tau_{0a} = 0$. Hence, the equilibrium price is $p_i = (1+\tau_{ia})^{\alpha_i}$, and input use features $(1+\tau_{ia}) l_{ia} = \alpha_i p_i y_i$, so $l_{ia} = \alpha_i (1+\tau_{ia})^{\alpha_i - 1} y_i$ and $l_{ib} = (1-\alpha_i)(1+\tau_{ia})^{\alpha_i} y_i$

The planning problem is $\max_{\tau_{1a}} L(p_1)$ with $p_1 = (1+\tau_{1a})^{\alpha_1}$, so that:

$$L(p_1) = c_0 + U^s(c_1(p_1)) - \xi_* c_1(p_1) - \sum_{i=0}^{1} (l_{ia}(p_1) + l_{ib}(p_1))$$
$$= U^s(c_1(p_1)) - \left(\xi_* + \alpha_1 (1+\tau_{1a})^{\alpha_1 - 1} + (1-\alpha_1)(1+\tau_{1a})^{\alpha_1}\right) c_1(p_1) \text{ as } c_0 - (l_{0a}(p_1) + l_{0b}(p_1))$$
$$= U^s(c_1(p_1)) - \left(\xi_* + \alpha_1 p_1^{1-\frac{1}{\alpha_1}} + (1-\alpha_1) p_1\right) c_1(p_1)$$

with $c_1(p_1) = U'^{-1}(p_1)$.

Hence, as $U^{s\prime}(c_1(p_1)) = p_1$,

$$L_{p_1} = c_1'(p_1) \left[p_1 - \left(\xi_* + \alpha_1 p_1^{1-\frac{1}{\alpha_1}} + (1-\alpha_1) p_1\right)\right] - \left((\alpha_1 - 1) p_1^{-\frac{1}{\alpha_1}} + (1-\alpha_1)\right) c_1(p_1)$$

When there is production efficiency, $p_1 = 1$ and,

$$L_{p_1 | \tau_{1a}=0} = c_1'(p_1) \left[U^{s\prime}(c_1(p_1)) - (\xi_* + 1)\right]$$
$$= -\xi_* c_1'(p_1) > 0$$

Hence, production efficiency is not an optimum. Starting from it, it is optimal to increase the tax τ_{1a} to discourage the production of good 1, increase its price, and discourage its consumption.

Case of an attentive consumer. It is enough to do a Pigouvian tax $\tau_1^c = \xi_*$, and restore production efficiency ($\tau_{1a} = 1$). Then, we achieve the first best. □

Proof of Proposition 7.4 Suppose that $\phi = \phi^s$. Let $e(\phi, q)$ be the expenditure function associated with $\phi(c_1, ..., c_n)$. Since ϕ is homogeneous of degree 1, we have $e(\phi, q) = \phi e(1, q)$. Consider a non homogeneous tax system with associated prices q. Tax revenues are

$$\sum_h \phi^h \sum_{i=1}^n (q_i - p_i) e_{q_i}(1, q)$$

Now consider a reformed uniform tax system with associated prices $\hat{q}_i = x p_i$ for some scalar x, which delivers the same c_0 and the same ϕ^h for all h. We just need to solve in x the following equation

$$e(1, q) = e(1, xp).$$

The reformed tax system leaves the experienced utility of all agents identical (this step crucially uses $\phi = \phi^s$). We claim that the reformed tax system also raises more revenues. This concludes the proof that the optimal tax system must be uniform. Laroque ("Indirect taxation is superfluous under separability and taste homogeneity: A simple proof," *Economics Letters* 2005) presents related arguments). This amounts to showing that

$$\sum_{i=1}^n (q_i - p_i) e_{q_i}(1, q) < \sum_{i=1}^n (\hat{q}_i - p_i) e_{q_i}(1, \hat{q}),$$

or using $\sum_{i=1}^n q_i e_{q_i}(1, q) = e(1, q) = e(1, \hat{q}) = \sum_{i=1}^n \hat{q}_i e_{q_i}(1, \hat{q})$,this amounts to showing that

$$0 < \sum_{i=1}^n p_i [e_{q_i}(1, q) - e_{q_i}(1, \hat{q})],$$

or equivalently since $\hat{q}_i = x p_i$, to showing that

$$0 < \sum_{i=1}^n \hat{q}_i [e_{q_i}(1, q) - e_{q_i}(1, \hat{q})],$$

which holds by a straight revealed preference argument.

□

13.2 Intermediary results for the Mirrlees problem

13.2.1 Basic Effects

Impact of a change in taxes on earnings and individual utility We first study the impact of a small change δq_{z^*} of the marginal retention rate at z^* and how it affects labor supply at z (e.g. via misperceptions). We simultaneously study the impact of a lump-sum (independent of

z) virtual income change δK. It will prove conceptually and notationally useful to define:

$$\overline{\overline{\zeta}}^{c}_{Q_{z^*}}(z) = \zeta^c_{Q_{z^*}}(z) + \zeta^c(z)\,\delta_z(z^*) \tag{105}$$

where δ_z is a Dirac distribution at point z. Hence, as $\zeta^c_{Q_{z^*}}(z)$ was a potentially smooth function of z^*, $\overline{\overline{\zeta}}^{c}_{Q_{z^*}}(z)$ is a generalized function of z^*, in the sense of the theory of distributions. From now on, we mostly use our notation convention of dropping the dependency on z.

Lemma 13.1 (Impact of changes in taxes on behavior and welfare) *Suppose that there is a change $(\delta q_{z^*})_{z^*\geq 0}$ to marginal retention rate schedule and a lump sum increase in revenue δK. The impact on earnings and agent's welfare is:*

$$\delta z = \frac{\eta \delta K + z\int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*}{q - \zeta^c z R''}, \tag{106}$$

$$\frac{\delta v}{v_r} = \delta K - z\frac{\tau^b}{q}\left(\zeta^c R''\delta z + \int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*\right). \tag{107}$$

In these equations, the integrals involving $z\int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*$ should be understood in the sense of the theory of distributions as $z\zeta^c(z)\,\delta q_z + \int_{z^*=0}^\infty z\zeta^c_{Q_{z^*}}(z)\,\delta q_{z^*}dz^*$ (reintroducing in these equations the dependency on z), leading to

$$\delta z = \frac{\eta(z)\,\delta K + z\zeta^c(z)\,\delta q_z + \int_0^\infty z\zeta^c_{Q_{z^*}}(z)\,\delta q_{z^*}dz^*}{q(z) - \zeta^c(z)\,zR''(z)},$$

$$\frac{\delta v}{v_r} = \delta K - z\frac{\tau^b}{q}\zeta^c(z)\,R''(z)\,\delta z - z\frac{\tau^b(z)}{q(z)}\zeta^c(z)\,\delta q_Z - \int_{z^*=0}^\infty z\frac{\tau^b(z)}{q(z)}\zeta^c_{Q_{z^*}}(z)\,\delta q_{z^*}dz^*.$$

To interpret the economics of (106), start with an increase in income δK. It has, first, an impact on labor supply: it creates a direct change in earnings supply equal to $\frac{\eta}{q}\delta K$. The additional term $\zeta^c zR''$ in the denominator of (106) is more subtle and arises from the fact that as the agent adjusts his labor supply, he experiences a different marginal tax rate (which changes as $R''\delta z$), leading to an additional change in income $\frac{\zeta^c}{q}zR''\delta z$. The final expression solves for δz as a fixed point. The term $z\int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*$ reflects the impact of a change in the marginal tax rate on earnings. The difference with Saez (2001) is that it is non-zero even when the change in the tax schedule occurs at $z^* \neq z$. This is because when agents have behavioral biases, a change of the marginal rate at z^* potentially affects the perceived tax at z.

In (107), the term δK is a mechanical income effect and is the only term present in the traditional model of Saez (2001). The term $-z\frac{\tau^b}{q}\left(\zeta^c R''\delta z + \int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*\right)$ represent the welfare impact arising from changes in behavior due to the failure of the envelope theorem because of misoptimization, respectively, because movements in labor supply change the marginal tax rate ($-z\frac{\tau^b}{q}\zeta^c R''\delta z$) along the initial schedule and because of changes in the tax schedule itself ($-z\frac{\tau^b}{q}\int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*$).

Impact of a change in taxes on social welfare We next study the impact of the above changes on welfare. Following Saez (2001), we call $h(z)$ the density of agents with earnings z at the optimum, and $H(z) = \int_0^z h(z')\,dz'$. We also define the virtual density $h^*(z) = \frac{q(z)}{q(z)-\zeta^c z R''(z)} h(z)$, which can also be written as $\frac{1-T'(z)}{1-T'(z)+\zeta^c z T''(z)} h(z)$.

Lemma 13.2 *Under the conditions of the Lemma 13.1, the change in the government objective function associated with the agent is*

$$\delta L(z) = (\gamma(z) - 1)\,\delta K + \frac{T'(z) - \widetilde{\tau}^b(z)}{1 - T'(z)} \frac{h^*(z)}{h(z)} z \int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}(z)\,\delta q_{z^*} dz^*, \tag{108}$$

where $\gamma(z)$ is the marginal social utility of income:

$$\gamma(z) = g(z) + \eta(z) \frac{\widetilde{\tau}^b(z)}{1 - T'(z)} + \eta(z) \frac{T'(z) - \widetilde{\tau}^b(z)}{1 - T'(z)} \frac{h^*(z)}{h(z)}. \tag{109}$$

This definition of the social marginal utility of income $\gamma(z)$ is similar to the one we encountered in the Ramsey problem. It encompasses the direct impact of one extra dollar on the agent's welfare (the $g(z)$ term) and the impact coming from a change in labor supply on tax revenues ($\frac{T'(z)}{1-T'(z)}\eta(z)\frac{h^*(z)}{h(z)}$). Compared to Saez (2001), it features a new term arising from the failure of the envelope theorem, $\eta\frac{\widetilde{\tau}^b(z)}{1-T'(z)}\left(1 - \frac{h^*(z)}{h(z)}\right)$.

The effect on the government objective function (108) is much like in the many-person Ramsey of Proposition 3.1. The term $(\gamma(z) - 1)\,\delta K$ is a mechanical effect, abstracting from changes in behavior. As the government gives (back) δK to agent, the impact on revenues is $-\delta K$, while the impact on the agent is valued as $\gamma(z)\,\delta K$. Next, there is a substitution effect $\frac{T'(z)}{1-T'(z)}\frac{h^*}{h} z \int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}(z)\,\delta q_{z^*} dz^*$: as the agent changes his labor supply, there is a change in tax revenues proportional to

$$\frac{T'(z)}{1 - T'(z)} \int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}(z)\,\delta q_{z^*} dz^*.$$

Third, there is a misoptimization term, $\frac{-\widetilde{\tau}^b(z)}{1-T'(z)}\frac{h^*(z)}{h(z)} z \int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}(z)\,\delta q_{z^*} dz^*$.

We also note the following first order condition for the intercept of the tax schedule, r_0.

Lemma 13.3 *At the optimum,*

$$\int_0^\infty \left(1 - \gamma(z) - \frac{T'(z) - \widetilde{\tau}^b(z)}{1 - T'(z)} \frac{h^*(z)}{h(z)} z \zeta^c_{r_0}(z)\right) h(z)\,dz = 0 \tag{110}$$

We next state the impact of a marginal change in the tax rate, $\frac{\partial L}{\partial \tau_{z^*}} \equiv -\frac{\partial L}{\partial q_{z^*}}$.

Proposition 13.1 (Impact of a local change on the marginal tax rate on the government objective function) *We have*

$$\frac{\partial L}{\partial \tau_{z^*}} = \int_{z^*}^{\infty} (1-\gamma(z))\, h(z)\, dz - \zeta^c(z^*) \frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1-T'(z^*)} z^* h^*(z^*) - \int_0^{\infty} \zeta^c_{Q_{z^*}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1-T'(z)} z h^*(z)\, dz \tag{111}$$

This equation involves an equality between two generalized functions of z^*. This is the income tax equivalent of the formula in Proposition 3.1 for the many-person Ramsey. The three terms in (111) correspond to the, by now familiar, mechanical ($\int_{z^*}^{\infty} (1-\gamma(z))\, h(z)\, dz$), substitution ($-\zeta^c(z^*) \frac{T'(z^*)}{1-T'(z^*)} z^* h^*(z^*)$), and misoptimization ($\zeta^c(z^*) \frac{\widetilde{\tau}^b(z^*)}{1-T'(z^*)} z^* h^*(z^*) - \int_0^{\infty} \zeta^c_{Q_{z^*}}(z) \frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)} z h^*(z)\, dz$) effects. The first two terms are exactly as in Saez (2001), and the third one is new as it is present only with behavioral agents. We will describe its meaning shortly. We also note that formula (111) can be written in a more compact way as:

$$\frac{\partial L}{\partial \tau_{z^*}} = \int_{z^*}^{\infty} (1-\gamma(z))\, h(z)\, dz - \int_0^{\infty} \overline{\overline{\zeta}}^c_{Q_{z^*}}(z) \frac{T'(z) - \widetilde{\tau}^b(z)}{1-T'(z)} z h^*(z)\, dz. \tag{112}$$

13.3 Proofs for the Mirrlees case

Notations and Derivation of relation (36) We take the material from section 11.1. The extended good is the two-dimensional $\boldsymbol{c} = (c, z)$, the (generalized) price vector is $\boldsymbol{q} = (1, q, \boldsymbol{Q}, r_0)$. The budget function is $B(\boldsymbol{c}, \boldsymbol{q}) = q_1 c_1 - q_2 c_2 = c - qz$, so that the budget constraint is $B(\boldsymbol{c}, \boldsymbol{q}) \le r$. Note that the Saez r is also the w in the rest of the paper (as the budget constraint is generally expressed as $B(\boldsymbol{c}, \boldsymbol{q}) \le r$); we still found useful to stick here to the Saez notations; so in the derivations of the Mirrlees case, we will use r and w interchangeably, depending on what the context calls for.

Applying definition (55) gives

$$\boldsymbol{\tau}^b = (1, -q) - \frac{(u_c, u_z)}{v_r} \tag{113}$$

We know that $c_{Q_{z^*}} = q z_{Q^*_z}$ (which comes from differentiating $c = qz + r$ w.r.t. Q^*_z), so

$$\boldsymbol{S}^C_{Q_{z*}} = \left(c_{Q_{z^*}}, z_{Q^*_z}\right)' = (q, 1)'\, z_{Q^*_z}$$

Proposition 11.1 implies:

$$\begin{aligned} \frac{v_{Q_{z*}}(\boldsymbol{q}, r)}{v_{\boldsymbol{r}}(\boldsymbol{q}, r)} &= -\boldsymbol{\tau}^b(\boldsymbol{q}, r) \cdot \boldsymbol{S}^C_{Q_{z*}}(\boldsymbol{q}, r) = -\boldsymbol{\tau}^b(\boldsymbol{q}, r)\, (q, 1)'\, z_{Q^*_z} = -\tau^b z_{Q^*_z} \\ &= -\tau^b \frac{z}{q} \zeta^c_{Q_{z^*}} \end{aligned}$$

as we defined

$$\tau^b = \boldsymbol{\tau}^b(\boldsymbol{q}, r) \cdot (q, 1) = -\frac{q u_c + u_z}{v_r}. \tag{114}$$

Likewise, $c - qz = r$ implies (taking the derivative w.r.t. q): $c_q - qz_q - z = 0$ and (taking the derivative w.r.t. r) $c_r - qz_r = 1$, so

$$\begin{aligned}
\boldsymbol{S}_q^C(\boldsymbol{q}, r) &= \boldsymbol{c}_q - \boldsymbol{c}_r z = (c_q - c_r z, z_q - z_r z) \\
&= (qz_q + z - (qz_r + 1) z, z_q - z_r z) = (q, 1)(z_q - z_r z) \\
&= (q, 1) z \frac{\zeta^c}{q}
\end{aligned} \tag{115}$$

Proposition 11.1 implies:

$$\begin{aligned}
\frac{v_q(\boldsymbol{q}, r)}{v_r(\boldsymbol{q}, r)} &= z - \boldsymbol{\tau}^b(\boldsymbol{q}, r) \cdot \boldsymbol{S}_q^C(\boldsymbol{q}, r) = z - \boldsymbol{\tau}^b(\boldsymbol{q}, r) \cdot (q, 1) z \frac{\zeta^c}{q} \\
&= z - \frac{z}{q} \tau^b \zeta^c.
\end{aligned}$$

□

Proof of Elasticity relations (37) in the Mirrlees framework: Concrete values of the general model in the misperception case Now consider the model with misperception. As above, the extended good is $\boldsymbol{c} = (c, z)$, and the (generalized) price $\boldsymbol{q} = (1, q, \boldsymbol{Q}, r_0)$, and the budget function is $B(\boldsymbol{c}, \boldsymbol{q}) = c_1 q_1 - c_2 q_2 = c - qz$, so that

$$B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{q}) = (1, -q) \tag{116}$$

We use (61)

$$\begin{aligned}
\boldsymbol{\tau}^b &= B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{q}) - \frac{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{q}^s)}{B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{q}^s) \cdot \boldsymbol{c}_r(\boldsymbol{q}, r)} \\
&= (1, -q) - \frac{(1, -q^s)}{(1, -q^s) \cdot (qz_r + 1, z_r)} \quad \text{as } c(q, r) = qz(q, r) + r \text{ gives } c_r = qz_r + 1. \\
&= (1, -q) - \frac{(1, -q^s)}{1 + (q - q^s) z_r} \\
\boldsymbol{\tau}^b &= (1, -q) - \frac{(1, -q^s)}{1 + (q - q^s) \frac{\eta}{q}}
\end{aligned} \tag{117}$$

Next, recall (114),

$$
\begin{aligned}
\tau^b &= \boldsymbol{\tau}^b\left(\boldsymbol{q}, r\right) \cdot (q, 1) \\
&= \left[(1, -q) - \frac{(1, -q^s)}{1 + (q - q^s)\frac{\eta}{q}}\right] \cdot (q, 1) = \frac{q^s - q}{1 + (q - q^s)\frac{\eta}{q}} \\
&= \frac{\tau - \tau^s}{1 - (\tau - \tau^s)\frac{\eta}{q}} \qquad (118)
\end{aligned}
$$

using $q = 1 - \tau$, $q^s = 1 - \tau^s$. Thus we have proven (38).

Next, we calculate ζ^c. We call $\boldsymbol{e} := (0, 1)$ the vector singling earnings on the vector $\boldsymbol{c} = (c, z)$. We apply (60) with $p_j = q$, the price of earnings. We have:

$$
\begin{aligned}
\boldsymbol{e} \cdot \boldsymbol{S}_j^H &= \boldsymbol{e} \cdot \boldsymbol{S}^r\left(\boldsymbol{p}, r\right) \cdot \boldsymbol{p}_{p_j}^s\left(\boldsymbol{p}, r\right) = z_q^r \frac{\partial q^s}{\partial q} = \frac{z}{q}\zeta^{c,r} m_{zz} \\
\boldsymbol{e} \cdot \boldsymbol{S}_j^H &= \frac{z}{q}\zeta^{c,r} m_{zz}. \qquad (119)
\end{aligned}
$$

Next, using the notation D_j of Proposition 11.1,

$$
\begin{aligned}
D_j &= -\boldsymbol{\tau}^b \cdot \boldsymbol{S}_j^H \text{ by } (54) \\
&= \left[B_{\boldsymbol{c}}\left(\boldsymbol{p}, \boldsymbol{c}\right) - B_{\boldsymbol{c}}\left(\boldsymbol{p}^s, \boldsymbol{c}\right)\right] \cdot \boldsymbol{S}_j^H \text{ by } (62) \\
&= \left[(1, q) - (1, q^s)\right] \cdot \boldsymbol{S}_j^H \text{ by } (116) \\
&= (q - q^s)\, \boldsymbol{e} \cdot \boldsymbol{S}_j^H \text{ as } \boldsymbol{e} = (0, 1) \\
&= (q - q^s)\frac{z}{q}\zeta^{c,r} m_{zz} \text{ by } (119)
\end{aligned}
$$

We record:

$$
D_j = (q - q^s)\frac{z}{q}\zeta^{c,r} m_{zz} \qquad (120)
$$

Next, we apply (58): $\boldsymbol{S}_j^C = \boldsymbol{S}_j^H + \boldsymbol{c}_r D_j$, which implies:

$$
\begin{aligned}
\boldsymbol{e} \cdot \boldsymbol{S}_j^C &= \boldsymbol{e} \cdot \boldsymbol{S}_j^H + \boldsymbol{e} \cdot \boldsymbol{c}_r D_j \\
&= \frac{z}{q}\zeta^{c,r} m_{zz} + \frac{\eta}{q} D_j \text{ as } \boldsymbol{e} \cdot \boldsymbol{c}_r = z_r = \frac{\eta}{q} \\
&= \frac{z}{q}\zeta^{c,r} m_{zz}\left(1 + \frac{\eta}{q}(q - q^s)\right) \\
&= \frac{z}{q}\left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{c,r} m_{zz} \text{ as } q = 1 - \tau, q^s = 1 - \tau^s
\end{aligned}
$$

Now, as $\zeta^c = \frac{q}{z}\boldsymbol{e} \cdot \boldsymbol{S}_j^C$ is the compensated earnings elasticity (see e.g. (115)):

$$\zeta^c = \frac{q}{z}\boldsymbol{e} \cdot \boldsymbol{S}_j^C$$

$$\zeta^c = \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{c,r}m_{zz} \tag{121}$$

Exactly the same reasoning (using then $p_j = q_{z^*}$, and $\boldsymbol{e} \cdot \boldsymbol{S}_j^H = \frac{z}{q}\zeta^{c,r}m_{zz*}$) shows

$$\zeta^c_{Q_{z^*}} = \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{c,r}m_{zz^*} \tag{122}$$

Hence, we have proven (37).

Proof of (39): Decision vs Experienced utility model The agent's optimization gives $qu^s_c + u^s_z = 0$. Equation (114) gives:

$$\tau^b = -\frac{qu_c + u_z}{v_r} = \frac{u_c\frac{u^s_z}{u^s_c} - u_z}{v_r}.$$

Dirac / Double bar Notation for the proofs in the Mirrlees framework We define:

$$\overline{\overline{\zeta}}^c_{Q_{z^*}}(z) = \zeta^c_{Q_{z^*}}(z) + \zeta^c(z)\,\delta_z(z^*)$$

Informally, this definition means that $\overline{\overline{\zeta}}^c_{Q_{z^*}}$ is like $\zeta^c_{Q_{z^*}}(z)$, but with an extra Dirac term when $z = z^*$.

Proof of Lemma 13.1 We have

$$z = z\left(q\left(z\right), \boldsymbol{Q}, r\left(z\right)\right)$$
$$r\left(z\right) = R\left(z\right) - zq\left(z\right)$$

so

$$\delta r = r'\left(z\right)\delta z + \delta r_{|\,\text{constant } z} = -zq'\left(z\right)\delta z + \left(\delta K - z\delta q_z\right)$$
$$= -zR''\delta z + \delta K - z\delta q_z$$

$$\begin{aligned}
\delta z &= z_q(q'(z)\,\delta z + \delta q_z) + \int_0^\infty z_{Q_{z^*}}\delta q_{z^*}dz^* + z_r\delta r \\
&= \frac{z}{q}\zeta^u q'(z)\,\delta z + \frac{z}{q}\zeta^u\delta q_z + \frac{z}{q}\int_0^\infty \zeta^c_{Q_{z^*}}\delta q_{z^*}dz^* + \frac{\eta}{q}\left(-zR''\delta z + \delta K - z\delta q_z\right) \\
&= \frac{z}{q}\zeta^u R''\delta z + \frac{z}{q}\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^* + \frac{\eta}{q}\left(-zR''\delta z + \delta K\right)
\end{aligned}$$

so

$$\delta z = \frac{z\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^* + \eta\delta K}{q + (\eta - \zeta^u)\,zR''} = \frac{z\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^* + \eta\delta K}{q - \zeta^c zR''}$$

For welfare $v(q, \boldsymbol{Q}, r_0, r)$, we have:

$$\begin{aligned}
\frac{\delta v}{v_r} &= \frac{v_q}{v_r}(q'(z)\,\delta z + \delta q_z) + \int_0^\infty \frac{v_{Q_{z^*}}}{v_r}\delta q_{z^*}dz^* + \delta r, \\
&= z\left(1 - \frac{\tau^b\zeta^c}{q}\right)(R''dz + \delta q_z) + z\int_0^\infty \left(-\frac{\tau^b\zeta^c_{Q_{z^*}}}{q}\right)\delta q_{z^*}dz^* - zR''\delta z + \delta K - z\delta q_z, \\
\frac{\delta v}{v_r} &= z\left(-\frac{\tau^b\zeta^c}{q}\right)R''\delta z + z\left(\int_0^\infty \left(-\frac{\tau^b\zeta^c_{Q_{z^*}}}{q}\right)\delta q_{z^*}dz^* + \left(-\frac{\tau^b\zeta^c}{q}\right)\delta q_z\right) + \delta K, \\
&= \delta K - z\frac{\tau^b}{q}\zeta^c R''(z)\,\delta z - \frac{\tau^b}{q}z\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^*, \\
&= \delta K - z\frac{\tau^b}{q}\left(\zeta^c R''(z)\,\delta z + \int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^*\right).
\end{aligned}$$

□

Proof of Lemma 13.2 Observe that

$$\frac{h^*(z)}{q} = \frac{h(z)}{q - \zeta^c zR''(z)}$$

so that $q - \zeta^c zR''(z) = \frac{qh}{h^*}$ and

$$zR'' = q\frac{h^* - h}{\zeta^c h^*}. \tag{123}$$

We have

$$\begin{aligned}
\delta T &= \delta\left(z\left(1 - q(z)\right) - r\right), \\
&= (1 - q_z)\,\delta z - zq'(z)\,\delta z - z\delta q_z - \delta r, \\
&= (1 - q_z)\,\delta z - zR''\delta z - z\delta q_z - (-zR''\delta z + \delta K - z\delta q_z), \\
&= T'(z)\,\delta z - \delta K.
\end{aligned}$$

We also have

$$\delta L = \delta T + g(z)\frac{\delta v}{v_r},$$
$$= T'(z)\,\delta z + g(z)\frac{\delta v}{v_r} - \delta K.$$

Using Lemma 13.1, we can rewrite this as

$$\delta L = T'(z)\frac{z\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^* + \eta\delta K}{q-\zeta^c zR''} + g(z)\left(\delta K - z\frac{\tau^b}{q}\left(\zeta^c R''(z)\,\delta z + \int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^*\right)\right) - \delta K.$$

Using equation (123) and Lemma 13.1, we can rewrite this as

$$\begin{aligned}\delta L &= \left[-1+g(z)+\eta\frac{T'(z)}{q}\frac{h^*}{h}+g(z)\left(-\frac{\tau^b\zeta^c}{q}\right)\eta\frac{h^*-h}{\zeta^c h}\right]\delta K,\\ &+\left(-\frac{g(z)\tau^b}{q}+\frac{T'(z)}{q}\frac{h^*}{h}+g(z)\left(-\frac{\tau^b\zeta^c}{q}\right)\frac{h^*-h}{\zeta^c h}\right)z\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^*,\\ &= (\gamma(z)-1)\,\delta K + \frac{h^*}{h}z\frac{T'(z)-\widetilde{\tau}^b(z)}{q}\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^*,\end{aligned}$$

where

$$\gamma(z) = g(z) + \eta\frac{\widetilde{\tau}^b(z)}{q} + \frac{T'(z)-\widetilde{\tau}^b(z)}{q}\eta\frac{h^*(z)}{h(z)}.$$

□

Proof of Proposition 13.1 We use the following notations:

$$\begin{aligned}\overline{\overline{F}}(z_*) &:= \int_0^\infty\left[-\overline{\overline{\zeta}}^c_{Q_{z^*}}\frac{\widetilde{\tau}^b(z)}{q(z)}+\zeta^c_{Q_{z^*}}\frac{T'(z)}{q}\right]zh^*(z)\,dz,\\ J(z^*) &:= \zeta^c z^*\frac{T'(z^*)}{q}h^*(z^*).\\ \overline{\overline{\overline{F}}}(z_*) &= \int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\frac{T'(z)-\widetilde{\tau}^b(z)}{q(z)}zh^*(z)\,dz = \overline{\overline{F}}(z_*)+J(z_*),\end{aligned}$$

We consider a change δq_{z^*} at z^*. This leads to a lump-sum change $\delta K = 1_{z>z^*}\delta q_{z^*}$. Hence, Lemma 13.2 gives the change in the government objective function

$$\delta L(z) = (\gamma(z)-1)\,1_{z>z^*}\delta q_{z^*} + \frac{h^*}{h}z\frac{T'(z)-\widetilde{\tau}^b(z)}{q}\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}\delta q_{z^*}dz^*.$$

The total change is

$$\begin{aligned}
\delta L &= \int_0^\infty \delta L(z)\, h(z)\, dz, \\
\frac{\delta L}{\delta q_{z^*}} &= \int_{z^*}^\infty (\gamma(z) - 1)\, h(z)\, dz + \int_0^\infty \left[\frac{T'(z) - \widetilde{\tau}^b(z)}{q} \overline{\overline{\zeta}}^c_{Q_{z^*}}\right] \frac{h^*}{h} z h(z)\, dz, \\
\frac{\partial L}{\partial \tau_{z^*}} &= -\overline{\overline{\overline{F}}}(z_*) + \int_{z^*}^\infty (1 - \gamma(z))\, h(z)\, dz. \qquad (124)
\end{aligned}$$

We also have

$$\frac{\partial L}{\partial \tau_{z^*}} \equiv -\frac{\partial L}{\partial Q_{z^*}} = -\overline{\overline{\overline{F}}}(z^*) + \int_{z^*}^\infty (1 - \gamma(z))\, h(z)\, dz. \qquad (125)$$

□

Proof of Lemma 13.3 Using Lemma 13.2, applied to a change δr_0 to all agents, and slightly generalizing, we find

$$\delta L(z) = (\gamma(z) - 1)\, \delta r_0 + \frac{T'(z) - \widetilde{\tau}^b(z)}{1 - T'(z)} \frac{h^*(z)}{h(z)} z \zeta^c_{r_0}(z)\, \delta r_0$$

and $\delta L = \int \delta L(z)\, h(z)\, dz$ should be 0.□

Proof of Proposition 5.1 Let us now solve for the optimal J, which ensures $\frac{\partial L}{\partial Q_{z^*}} = 0$. We can write

$$-\frac{\partial L}{\partial \tau_{z^*}} = J(z^*) - \int_{z^*}^\infty a(z)\, dz - b(z^*) + \int_{z^*}^\infty J(z)\, \rho(z)\, dz. \qquad (126)$$

We use the notations

$$\begin{aligned}
\rho(z) &:= \frac{\eta}{\zeta^c} \frac{1}{z}, \qquad (127) \\
a(z) &:= (1 - g(z))\, h(z) - \rho g(z)\, z \left(-\tau^b(z) \frac{\zeta^c}{q(z)}\right) (h^*(z) - h(z)), \qquad (128) \\
b(z) &:= -\overline{\overline{F}}(z_*). \qquad (129)
\end{aligned}$$

$a(z)$ is the effect of giving \$1 to agent z (that's the $(1 - g(z))\, h(z)$ term), corrected from distortions from the non-linearity of the income tax.

$\overline{\overline{F}}(z^*)$ is the part impact on the government's objective function of increase δq_{z^*}, coming from the distortions from perceptions

$$\overline{\overline{F}}(z^*) := \int_0^\infty \left[-\overline{\overline{\zeta}}^c_{Q_{z^*}} \frac{\widetilde{\tau}^b(z)}{q(z)} + \zeta^c_{Q_{z^*}} \frac{T'(z)}{q}\right] zh^*(z)\,dz = F(z^*) + zg(z)\left(-\tau^b \frac{\zeta^c}{q(z)}\right) h^*(z),$$
$$F(z^*) := \int_0^\infty \left[\zeta^c_{Q_{z^*}} \frac{T'(z) - \tilde{\tau}^b(z)}{q}\right] zh^*(z)\,dz.$$

We note that

$$a = (1-\gamma)h + \eta \frac{T'(z)}{q}\frac{h^*}{h}h,$$
$$a = (1-\gamma)h + \rho J. \tag{130}$$

We also have

$$J(z^*) = \int_{z^*}^\infty a(z)\,dz + b(z^*) - \int_{z^*}^\infty J(z)\rho(z)\,dz,$$
$$\dot{J} = -a + \dot{b} + J\rho,$$
$$\frac{d}{dz}\left[J(z)e^{-\int_0^z \rho(s)ds}\right] = e^{-\int_0^z \rho(s)ds}\left(-a(z) + \dot{b}(z)\right),$$
$$J(z)e^{-\int_0^z \rho(s)ds} = C + \int_z^\infty e^{-\int_0^{z'} \rho(s)ds}\left(a(z') - \dot{b}(z')\right)dz',$$
$$J(z) = \int_z^\infty e^{-\int_z^{z'} \rho(s)ds}\left(a(z') - \dot{b}(z')\right)dz'. \tag{131}$$

Integrating by parts, we get

$$\int_z^\infty e^{-\int_z^{z'} \rho(s)ds}\dot{b}(z')\,dz' = \left[e^{-\int_z^{z'} \rho(s)ds} b(z')\right]_z^\infty + \int_z^\infty e^{-\int_z^{z'} \rho(s)ds}\rho(z')b(z')\,dz'$$
$$= -b(z) + \int_z^\infty e^{-\int_z^{z'} \rho(s)ds}\rho(z')b(z')\,dz', \tag{132}$$
$$J(z) = b(z) + \int_z^\infty e^{-\int_z^{z'} \rho(s)ds}\left(a(z') - \rho(z')b(z')\right)dz'. \tag{133}$$

We can rewrite this as

$$J(z^*) = -\overline{\overline{F}}(z^*) + \int_{z^*}^\infty e^{-\int_{z^*}^z \rho(s)ds}\left(a(z) + \rho\overline{\overline{F}}(z)\right)dz \tag{134}$$
$$= -z^*g(z^*)\left(-\tau^b(z^*)\frac{\zeta^c(z^*)}{q(z^*)}\right)h^*(z^*) - F(z^*)$$
$$+ \int_{z^*}^\infty e^{-\int_{z^*}^z \rho(s)ds}\left[(1-g(z))h(z) + g(z)\rho(z)z\left(-\tau^b(z)\frac{\zeta^c(z)}{q(z)}\right)h(z) + \rho(z)F(z)\right]dz.$$

Using

$$J(z^*) = \zeta^c(z^*)\, z^* h^*(z^*)\, \frac{T'(z^*)}{1 - T'(z^*)}$$

and rearranging gives

$$\zeta^c(z^*)\, z^* h^*(z^*)\, \frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} + F(z_*) - \int_{z^*}^{\infty} e^{-\int_{z^*}^{z} \rho(s) ds} \rho(z)\, F(z)\, dz$$
$$= \int_{z^*}^{\infty} e^{-\int_{z^*}^{z} \rho(s) ds} \left[(1 - g(z))\, h(z) + g(z)\, \rho(z)\, z \left(-\tau^b(z)\, \frac{\zeta^c(z)}{q(z)} \right) h(z) \right] dz,$$

which can be rewritten to get the announced formula

$$\frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} + \frac{1}{\zeta^c} \frac{1}{z^* h^*(z^*)} \int_{z=0}^{\infty} \left(\zeta^c_{Q_{z^*}}(z) - \int_{z'=z^*}^{\infty} e^{-\int_{z^*}^{z'} \rho(s) ds} \rho(z')\, \zeta^c_{Q_{z'}}(z)\, dz' \right) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz,$$
$$= \frac{1}{\zeta^c} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} e^{-\int_{z^*}^{z} \rho(s) ds} \left(1 - g(z) - \eta \frac{\tilde{\tau}^b(z)}{q(z)} \right) \frac{h(z)}{1 - H(z^*)} dz.$$

□

Proof of Proposition 5.2 We use (13.1):

$$\frac{\partial L}{\partial \tau_{z^*}} = \int_{z^*}^{\infty} (1 - \gamma(z))\, h(z)\, dz - \zeta^c(z^*) \frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} z^* h^*(z^*) - \int_0^{\infty} \zeta^c_{Q_{z^*}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz$$
$$= \int_{z^*}^{\infty} \left(1 - g(z) - \eta(z)\, \frac{T'(z)}{1 - T'(z)} \right) h(z)\, dz - \zeta^c(z^*) \frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} z^* h^*(z^*)$$
$$- \int_0^{\infty} \zeta^c_{Q_{z^*}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz \text{ using (40)}$$
$$= (1 - H(z_*)) \left(1 - \overline{g} - \overline{\eta} \frac{\overline{\tau}}{1 - \overline{\tau}} \right) - \overline{\zeta}^c \frac{\overline{\tau} - \overline{g\tau^b}}{1 - \overline{\tau}} z^* h^*(z^*) - \int_0^{\infty} \zeta^c_{Q_{z^*}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz$$

Recall that

$$\int_{z_0}^{\infty} (1 - H(z_*))\, dz_* = \left[(1 - H(z_*))\, (z - z_*) \right]_{z_0}^{\infty} - \int_{z_0}^{\infty} h(z_*)\, (z - z_*)\, dz$$
$$= \mathbb{E}\left[(z - z_0)\, 1_{z \geq z_0} \right]$$
$$= \frac{1}{\pi} \mathbb{E}\left[z 1_{z \geq z_0} \right]$$

Given the constraint that $T'(z) = \overline{\tau}$ for $z > z_0$, the FOC on $\overline{\tau}$ is: $0 = \int_{z_0}^{\infty} \frac{\partial L}{\partial \tau_{z^*}} dz_*$, i.e.

$$
\begin{aligned}
0 &= \int_{z_0}^{\infty} \frac{\partial L}{\partial \tau_{z^*}} dz_* = \left(1 - \overline{g} - \overline{\eta}\frac{\overline{\tau}}{1-\overline{\tau}}\right) \int_{z_0}^{\infty} (1 - H(z_*))\, dz_* - \overline{\zeta}^c \frac{\overline{\tau} - \overline{g\tau}^b}{1-\overline{\tau}} \mathbb{E}\left[z 1_{z \geq z_0}\right] \\
&- \int_0^{\infty} \left(\int_{z_0}^{\infty} \zeta^c_{Q_{z^*}}(z) dz_*\right) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz \\
&= \left(1 - \overline{g} - \overline{\eta}\frac{\overline{\tau}}{1-\overline{\tau}}\right) \frac{1}{\pi} \mathbb{E}\left[z 1_{z \geq z_0}\right] - \overline{\zeta}^c \frac{\overline{\tau} - \overline{g\tau}^b}{1-\overline{\tau}} \mathbb{E}\left[z 1_{z \geq z_0}\right] - \int_0^{\infty} \zeta^c_{\overline{q}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz \\
0 &= \left(1 - \overline{g} - \overline{\eta}\frac{\overline{\tau}}{1-\overline{\tau}}\right) - \overline{\zeta}^c \pi \frac{\overline{\tau} - \overline{g\tau}^b}{1-\overline{\tau}} - \frac{\pi}{\mathbb{E}\left[z 1_{z \geq z_0}\right]} \int_0^{\infty} \zeta^c_{\overline{q}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz
\end{aligned}
$$

Hence, we have:

$$
\frac{\overline{\tau}}{1-\overline{\tau}} = \frac{1 - \overline{g} - \beta + \zeta^c \pi \overline{g} \frac{\overline{\tau}^b}{1-\overline{\tau}}}{\zeta^c \pi + \overline{\eta}} \tag{135}
$$

with

$$
\begin{aligned}
\beta &:= \frac{\pi}{\mathbb{E}\left[z 1_{z \geq z_0}\right]} \int_0^{\infty} \zeta^c_{\overline{q}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} z h^*(z)\, dz \\
&= \mathbb{E}^z \left[\frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)}\right] \mathbb{E}^* \left[\zeta^c_{\overline{q}}(z)\right] \frac{\pi \mathbb{E}[z]}{\mathbb{E}\left[z 1_{z \geq z_0}\right]}
\end{aligned}
$$

where $\mathbb{E}^z[\phi(z)] = \frac{\int \phi(z) h(z) z dz}{\int \phi(z) z dz}$ and $\mathbb{E}^*\left[\zeta^c_{\overline{q}}(z)\right] = \frac{\mathbb{E}^*\left[\zeta^c_{\overline{q}}(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} h^*(z) dz\right]}{\mathbb{E}^*\left[\frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} h^*(z) dz\right]}$ is a weighted average too.

We can rewrite the equation as : $\frac{\overline{\tau}}{1-\overline{\tau}} = \frac{a + \frac{b}{1-\overline{\tau}}}{c}$, which gives $\overline{\tau} = \frac{a+b}{a+c}$, so that:

$$
\overline{\tau} = \frac{1 - \overline{g} - \beta + \zeta^c \pi \overline{g} \tau^b}{1 - \overline{g} - \beta + \zeta^c \pi + \overline{\eta}}
$$

14 Appendix: Basic behavioral consumer theory with linear budget constraints

14.1 Traditional theory: Recap

The objects in the traditional theory are $e(\boldsymbol{p}, u)$, $v(\boldsymbol{p}, w)$, $\boldsymbol{h}(\boldsymbol{p}, u) = \arg\min_{\boldsymbol{c}} \boldsymbol{p} \cdot \boldsymbol{c}$ s.t. $u(\boldsymbol{c}) = u$. Let us prove the traditional relations – a warm up for the proof in the behavioral case.

Roy's identity is proven as follows: $v(\boldsymbol{p}, w) = \max_{\boldsymbol{c}} u(\boldsymbol{c}) + \lambda(w - \boldsymbol{p} \cdot \boldsymbol{c})$, so $v_{p_j} = -\lambda c_j, v_w = \lambda$, so:

$$v_{p_j} + v_w c^j = 0 \tag{136}$$

Shepard's lemma is proven as follows: The envelope theorem gives $e_p(\boldsymbol{p}, u) = \boldsymbol{h}(\boldsymbol{p}, u)$, i.e.

$$e_{p_i}(\boldsymbol{p}, u) = h^i \tag{137}$$

and differentiating once more gives:

$$e_{p_i p_j} = h^i_{p_j}(\boldsymbol{p}, u) := S_{ij} \tag{138}$$

We have $\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w))$, which implies

$$\begin{aligned}\boldsymbol{c}_{p_j} &= \boldsymbol{h}_{p_j} + \boldsymbol{h}_u v_{p_j} \\ \boldsymbol{c}_w &= \boldsymbol{h}_u v_w\end{aligned}$$

and because of Roy, we have Slutsky's relation:

$$\boldsymbol{c}_{p_j} + \boldsymbol{c}_w c_j = \boldsymbol{h}_{p_j} =: \boldsymbol{S}_j$$

i.e.

$$c^i_{p_j} + c^i_w c_j = h^i_{p_j} = S_{ij} \tag{139}$$

14.2 Behavioral version with perceived prices

The sparse max demand

$$\operatorname*{smax}_{\boldsymbol{c}|\boldsymbol{p}^s} u(\boldsymbol{c}) \text{ s.t. } \boldsymbol{p} \cdot \boldsymbol{c} \leq w$$

of a behavioral agent perceiving prices $\boldsymbol{p}^s$ (while true prices are $\boldsymbol{p}$ and the true budget is w) is:

$$\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w'(\boldsymbol{p}, \boldsymbol{p}^s, w)) \tag{140}$$

where perceived budget w' satisfies:

$$\boldsymbol{p} \cdot \boldsymbol{c}^r(\boldsymbol{p}^s, w'(\boldsymbol{p}, \boldsymbol{p}^s, w)) = w \tag{141}$$

We call $\boldsymbol{c}^r(\boldsymbol{p}^s, w')$ the rational Marshallian demand under prices $\boldsymbol{p}^s$ and budget w'.

We define $v(\boldsymbol{p}, \boldsymbol{p}^s, w) = u(\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w))$. The expenditure function is $e(\boldsymbol{p}, \boldsymbol{p}^s, u) = \min_w w$ s.t. $v(\boldsymbol{p}, \boldsymbol{p}^s, w) \geq u$. We define the Hicksian demand $\boldsymbol{h}(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}) = \arg\operatorname{smax}_{\boldsymbol{c}|\boldsymbol{p}^s} -\boldsymbol{p} \cdot \boldsymbol{c}$ s.t. $u(\boldsymbol{c}) = \overline{u}$ with perception $\boldsymbol{p}^s$ by the agent, which gives $\boldsymbol{h}(\boldsymbol{p}, \boldsymbol{p}^s, u) = \boldsymbol{h}^r(\boldsymbol{p}^s, u)$. So $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) = \boldsymbol{h}^r(\boldsymbol{p}^s, v(\boldsymbol{p}, \boldsymbol{p}^s, w))$.

Here we derive Shepard, Roy etc. for this behavioral model. This generalizes Gabaix (2014), which derives similar relations under the assumption that $\boldsymbol{p}^s = \boldsymbol{M}p + (1 - \boldsymbol{M})\, p^d$.

We call

$$\boldsymbol{S}_j^r = \boldsymbol{h}^r_{p_j^s}\left(\boldsymbol{p}^s, u\right)$$

the rational Slutsky matrix, and $\boldsymbol{S}_j^r = \left(S_{ij}^r\right)_{i=1\ldots n}$ the vector of Slutsky sensitivities with respect to price p_j.

Proposition 14.1 (Generalized Shepard's lemma) *Given the function* $e\left(\boldsymbol{p}, \boldsymbol{p}^s, u\right) = \boldsymbol{p} \cdot \boldsymbol{h}^r\left(\boldsymbol{p}^s, u\right)$, *we have:*

$$\begin{aligned} e_{p_j} &= c_j \\ e_{p_j^s} &= \left(\boldsymbol{p} - \boldsymbol{p}^s\right) \cdot \boldsymbol{S}_j^r, \end{aligned}$$

Proof. We have: $e\left(\boldsymbol{p}, \boldsymbol{p}^s, u\right) = \boldsymbol{p} \cdot \boldsymbol{h}^r\left(\boldsymbol{p}^s, u\right)$, so

$$\begin{aligned} e_{p_j} &= \boldsymbol{h}_j^r = c_j \\ e_{p_j^s} &= \boldsymbol{p} \cdot \boldsymbol{h}^r_{p_j^s}\left(\boldsymbol{p}^s, u\right) = \left(\boldsymbol{p} - \boldsymbol{p}^s\right) \cdot \boldsymbol{h}^r_{p_j^s}\left(\boldsymbol{p}^s, u\right), \end{aligned}$$

Indeed, we have $\boldsymbol{p}^s \cdot \boldsymbol{h}^r\left(\boldsymbol{p}^s, u\right) = 0$. To prove this, observe that

$$\begin{aligned} \boldsymbol{q} \cdot \boldsymbol{h}^r_{q_j}\left(\boldsymbol{q}, u\right) &= \sum_i q_i h^i_{q_j} = \sum_i q_i h^j_{q_i} \text{ by symmetry} \\ &= 0 \text{ as } h^j\left(q, u\right) \text{ is homogeneous of degree 0.} \end{aligned}$$

□

Proposition 14.2 *(*Generalized Roy's identity). *Given the function* $v\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right)$, *we have:*

$$\begin{aligned} \frac{v_{p_j}}{v_w} &= -c_j \\ \frac{v_{p_j^s}}{v_w} &= \left(\boldsymbol{p}^s - \boldsymbol{p}\right) \cdot \boldsymbol{S}_j^r := D_j^s \end{aligned}$$

i.e. $\frac{v_{p_j^s}}{v_w} = \sum_i \left(p_i^s - \boldsymbol{p}_i\right) \boldsymbol{S}_{ij}^r$.

To gain intuition for the term in $v^s_{p_j^s}$, observe that:

$$v^s_{p^s} \cdot \delta p^s \geq 0 \text{ with } \delta p^s = 0.01\left(p - p^s\right)$$

This is, the agent is better off if his perceived price goes towards the true price.

Proof of Proposition 14.2

For a number $\overline{u}$, we have the identity $\overline{u} = v(\boldsymbol{p}, \boldsymbol{p}^s, e(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}))$ for all $\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}$. Deriving w.r.t. p_j gives:

$$0 = v_{p_j} + v_w e_{p_j} = v_{p_j} + v_w c_j$$

by the behavioral Shepard's lemma (Proposition 14.1).

Deriving w.r.t. p_j^s gives:

$$0 = v_{p_j^s} + v_w e_{p_j^s} = v_{p_j^s} + v_w (\boldsymbol{p} - \boldsymbol{p}^s) \cdot \boldsymbol{S}_j^r$$

again by the behavioral Shepard's lemma (Proposition 14.1).□

Proposition 14.3 *(Marshallian demand) Given the consumption function $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w)$, we have:*

$$\begin{aligned}
\boldsymbol{c}_{p_j} &= -\boldsymbol{c}_w c_j && (142)\\
\boldsymbol{c}_{p_j^s} &= \boldsymbol{S}_j^r + \boldsymbol{c}_w D_j^s =: S_j^s && (143)\\
&= \boldsymbol{S}_j^r + \boldsymbol{c}_w \left[(\boldsymbol{p}^s - \boldsymbol{p}) \cdot \boldsymbol{S}_j^r \right] && (144)\\
&= \left(1 + \boldsymbol{c}_w (\boldsymbol{p}^s - \boldsymbol{p})' \right) \boldsymbol{S}_j^r && (145)
\end{aligned}$$

i.e. $c_{p_j}^i = -c_w^i c_j$ *and* $c_{p_j^s}^i = S_{ij}^r + c_w^i D_j^s$. *In addition,* $\boldsymbol{c}_w = \frac{\boldsymbol{1}}{\boldsymbol{p} \cdot \boldsymbol{c}_{w'}^r (\boldsymbol{p}^s, w')} \boldsymbol{c}_{w'}^r = \frac{v_w}{\lambda} \boldsymbol{c}_{w'}^r$.

The new term is $c_w^i D_j^s$. To interpret it, consider again what happens if if the agent's perceived price goes towards the true price. $dp^s = \chi (p - p^s)$, $\chi > 0$. Then,

$$\begin{aligned}
d\boldsymbol{c} &= \boldsymbol{c}_{p^s} dp^s = S dp^s + \boldsymbol{c}_w dE \\
dE &= [(\boldsymbol{p}^s - \boldsymbol{p}) \cdot \boldsymbol{S}^r \cdot dp^s] \geq 0
\end{aligned}$$

The extra term dE is positive: it's as if the agent became richer. That creates in income effect, and shift his consumption $\boldsymbol{c}_w dE$. We can summarize: *"If the agent's perceived price goes towards the true price, the agent is better off, and the consumer consumes as if she was richer"*.

Proof. We have $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) = \boldsymbol{h}^r (\boldsymbol{p}^s, v(\boldsymbol{p}, \boldsymbol{p}^s, w))$, which implies

$$\boldsymbol{c}_w = \boldsymbol{h}_u^r v_w, \qquad \boldsymbol{c}_{p_j} = \boldsymbol{h}_u^r v_{p_j} \tag{146}$$

Because of Roy ($v_{p_j} + c_j v_w = 0$), we have: $\boldsymbol{c}_{p_j} + \boldsymbol{c}_w c_j = 0$.

Also, $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) = \boldsymbol{h}(\boldsymbol{p}^s, v(\boldsymbol{p}, \boldsymbol{p}^s, w))$ gives:

$$\begin{aligned}
\boldsymbol{c}_{p_j^s}(\boldsymbol{p}, \boldsymbol{p}^s, w) &= \boldsymbol{h}_{p_j^s} + \boldsymbol{h}_u v_{p_j^s} \\
&= \boldsymbol{S}_j^r + \boldsymbol{c}_w \frac{v_{p_j^s}}{v_w} \text{ using } \boldsymbol{c}_w = \boldsymbol{h}_u^r v_w \\
&= \boldsymbol{S}_j^r + \boldsymbol{c}_w D_j^s \text{ using Proposition 14.2}
\end{aligned}$$

We have (141): so $\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}\frac{\partial w'}{\partial w}=1$, and $\frac{\partial w'}{\partial w}(\boldsymbol{p},\boldsymbol{p}^s,w)=\frac{1}{\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}$. So $\boldsymbol{c}(\boldsymbol{p},\boldsymbol{p}^s,w)=\boldsymbol{c}^r(\boldsymbol{p}^s,w'(\boldsymbol{p},\boldsymbol{p}^s,w))$, we have $\boldsymbol{c}^s_w=\boldsymbol{c}^r_{w'}\frac{\partial w'}{\partial w}$.

□

In the traditional model $v(\boldsymbol{p},w)=\max_{\boldsymbol{c}} u(\boldsymbol{c})+\lambda(w-\boldsymbol{p}\cdot\boldsymbol{c})$ implies $v_w=\lambda$. There is a deviation here, as indicated below.

Proposition 14.4 *(Envelope theorem, modified) Call λ the Lagrange multiplier such that $u'(\boldsymbol{c})=\lambda p^s$. We have:*

$$\frac{v_w}{\lambda}=p^s\cdot\boldsymbol{c}_w(\boldsymbol{p},\boldsymbol{p}^s,w)=1+(p^s-p)\cdot\boldsymbol{c}_w(\boldsymbol{p},\boldsymbol{p}^s,w)$$

Proof. We have:

$$\boldsymbol{p}\cdot\boldsymbol{c}^r(\boldsymbol{p}^s,w'(\boldsymbol{p},\boldsymbol{p}^s,w))=w$$

so $\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}\frac{\partial w'}{\partial w}=1$, and

$$\frac{\partial w'}{\partial w}=\frac{1}{\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}$$

Also, given $\boldsymbol{c}(\boldsymbol{p},\boldsymbol{p}^s,w)=\boldsymbol{c}^r(\boldsymbol{p}^s,w'(\boldsymbol{p},\boldsymbol{p}^s,w))$

$$\boldsymbol{c}_w=\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')\frac{\partial w'}{\partial w}=\frac{\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}{\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}.$$

Next, given $v(\boldsymbol{p},\boldsymbol{p}^s,w)=u(\boldsymbol{c}^s(\boldsymbol{p},\boldsymbol{p}^s,w))$ we have:

$$\begin{aligned}
v_w &= u'(\boldsymbol{c}^s)\cdot\boldsymbol{c}_w=\lambda p^s\cdot\boldsymbol{c}_w=\lambda p^s\cdot\left(\frac{\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}{\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}\right)=\lambda\frac{p^s\cdot\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}{\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}\\
v_w &= \frac{\boldsymbol{\lambda}}{\boldsymbol{p}\cdot\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s,w')}
\end{aligned}$$

□

We can check that things are consistent: with $u_{\boldsymbol{c}}=\lambda\boldsymbol{p}^s$,

$$v_{p^s_j}=u_{\boldsymbol{c}}\boldsymbol{c}_{p^s_j}=\lambda\boldsymbol{p}^s(1+\boldsymbol{c}_w p')\boldsymbol{S}^r_j=\lambda\boldsymbol{p}^s\boldsymbol{c}_w p'\boldsymbol{S}^r_j=v_w p'\boldsymbol{S}^r_j=v_w D^s_j.$$

Proposition 14.5 *(Expenditure function – second derivatives) Given $e^s(\boldsymbol{p},\boldsymbol{p}^s,u)=\boldsymbol{p}\cdot\boldsymbol{h}(\boldsymbol{p}^s,u)$, we have*

$$\begin{aligned}
e^s_{p_ip_j} &= 0\\
e^s_{p_ip^s_j} &= S^r_{ij}\\
e^s_{p^s_ip^s_j} &= -S^r_{ij}+(\boldsymbol{p}-\boldsymbol{p}^s)\cdot\boldsymbol{h}_{p^s_ip^s_j}=-S^r_{ij}+\sum_k(p_k-p^s_k)\,h^k_{p^s_ip^s_j}
\end{aligned}$$

The first derivatives of the expenditure functions were calculated in Proposition 14.1.

Proof. Given $e^s(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}) = \boldsymbol{p}\cdot\boldsymbol{h}(\boldsymbol{p}^s, \overline{u})$, we saw earlier in Proposition 14.1.

$$\begin{aligned} e^s_{p_j} &= h^j(\boldsymbol{p}^s, \overline{u}) \\ e^s_{p^s_j} &= (\boldsymbol{p} - \boldsymbol{p}^s)\cdot \boldsymbol{h}_{p^s_j}(\boldsymbol{p}^s, \overline{u}) = \boldsymbol{p}\cdot \boldsymbol{h}_{p^s_j}(\boldsymbol{p}^s, \overline{u}) \end{aligned}$$

Differentiating more,

$$e^s_{p_i p_j} = 0 \tag{147}$$

$$e^s_{p_i p^s_j} = h^i_{p^s_j}(\boldsymbol{p}^s, u) = S^r_{ij} \tag{148}$$

and as $e^s_{p^s_j} = (\boldsymbol{p} - \boldsymbol{p}^s)\cdot \boldsymbol{h}^r_{p^s_j}(\boldsymbol{p}^s, u)$,

$$e^s_{p^s_i p^s_j} = -h^i_{p^s_j} + \sum_k (p_k - p^s_k)\, h^k_{p^s_i p^s_j}$$

□

15 Complements on basic consumer theory with nonlinear budget constraints

We give here complements to Section 11.

15.1 Rational agent

Primal is: $v(\boldsymbol{p}, w) = \max_{\boldsymbol{c}} u(\boldsymbol{c})$ s.t. $B(\boldsymbol{p}, \boldsymbol{c}) \le w$ and demand $\boldsymbol{c}(\boldsymbol{p}, w)$. We can also define $e(\boldsymbol{p}, \overline{u}) = \arg\min_c B(\boldsymbol{p}, \boldsymbol{c})$ s.t. $u(c) \ge \overline{u}$ and Hicksian demand $\boldsymbol{h}(\boldsymbol{p}, \overline{u})$. We next derive the traditional consumer relation with that non-linear budget constraint.

Shepard's lemma: The envelope theorem gives

$$\begin{aligned} e_{p_i}(\boldsymbol{p}, \overline{u}) &= B_{p_i}(\boldsymbol{h}(\boldsymbol{p}, \overline{u}), \boldsymbol{p}) \\ e_{p_i \boldsymbol{p}_j} &= B_{p_i \boldsymbol{p}_j}(\boldsymbol{h}(\boldsymbol{p}, \overline{u}), \boldsymbol{p}) + B_{\boldsymbol{p}_i c}\cdot \boldsymbol{h}_{\boldsymbol{p}_j}(\boldsymbol{p}, \overline{u}) \end{aligned}$$

(the last term is to be read: $B_{\boldsymbol{p}_i c}\cdot \boldsymbol{h}_{\boldsymbol{p}_j} = \sum_k B_{\boldsymbol{p}_i c_k}\cdot \boldsymbol{h}^{c_k}_{\boldsymbol{p}_j}$). We note that $B_{\boldsymbol{p}_i c}\cdot \boldsymbol{h}_{\boldsymbol{p}_j}$ is symmetric.

Roy's identity: $\overline{u} = v(\boldsymbol{p}, e(\boldsymbol{p}, \overline{u}))$, so $0 = v_{p_i} + v_w e_{p_i}$, i.e.

$$v_{p_i} = -v_w B_{p_i} \tag{149}$$

Given $\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w))$, we have $\boldsymbol{c}_w = \boldsymbol{h}_u v_w$, $\boldsymbol{c}_{p_i} = \boldsymbol{h}_{p_i} + \boldsymbol{h}_u v_{p_i} = \boldsymbol{h}_{p_i} + \boldsymbol{c}_w \frac{v_{p_i}}{v_w}$ and because of Roy, $\boldsymbol{c}_{p_i} = \boldsymbol{h}_{p_i} - \boldsymbol{c}_w B_{p_i}$, i.e. the Slutsky relation for nonlinear budget constraints:

$$\boldsymbol{h}_{p_i} = \boldsymbol{c}_{p_i} + \boldsymbol{c}_w B_{p_i} \tag{150}$$

Finally, given $B\left(\boldsymbol{c}\left(\boldsymbol{p}, w\right), \boldsymbol{p}\right) = w$,

$$B_{\boldsymbol{c}} \boldsymbol{c}_{p_i} = -B_{p_i}, B_{\boldsymbol{c}} \boldsymbol{c}_w = 1 \tag{151}$$

Premultiplying (150) by $B_{\boldsymbol{c}}$ gives: $B_{\boldsymbol{c}} \boldsymbol{h}_{p_i} = B_{\boldsymbol{c}} \boldsymbol{c}_{p_i} + B_{\boldsymbol{c}} \boldsymbol{c}_w B_{\boldsymbol{p}_i} = -B_{p_i} + B_{p_i} = 0$,

$$B_{\boldsymbol{c}} \boldsymbol{h}_{p_i} = 0 \tag{152}$$

All in all, traditional consumer theory holds, replacing p by $B_{\boldsymbol{c}}$.

15.2 Misperceiving Agent

We now study

$$v\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right) = \underset{\boldsymbol{c}|\boldsymbol{p}^s}{\operatorname{smax}}\, u\left(\boldsymbol{c}\right) \text{ s.t. } B\left(\boldsymbol{c}, \boldsymbol{p}\right) \leq w \tag{153}$$

We call $\boldsymbol{c}^s\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right)$ the demand function. Recall that's it's characterized by $u_{\boldsymbol{c}}\left(\boldsymbol{c}\right) = \lambda B_{\boldsymbol{c}}\left(\boldsymbol{c}, \boldsymbol{p}^s\right)$ for some λ, and $\boldsymbol{c}\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right) = \boldsymbol{c}^r\left(\boldsymbol{p}^s, w'\right)$ for a w' that ensures

$$B\left(\boldsymbol{c}^s\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right), \boldsymbol{p}\right) = w.$$

Given $B\left(\boldsymbol{c}\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right), \boldsymbol{p}\right) = w$, we have (taking the derivatives w.r.t. $\boldsymbol{p}, \boldsymbol{p}^s, w$):

$$\begin{aligned} B_{\boldsymbol{c}} \boldsymbol{c}_{\boldsymbol{p}} &= -B_{\boldsymbol{p}} \\ B_{\boldsymbol{c}} \boldsymbol{c}_{\boldsymbol{p}^s} &= 0 \\ B_{\boldsymbol{c}} \boldsymbol{c}_w &= 1 \end{aligned}$$

where $B_{\boldsymbol{c}} := B_{\boldsymbol{c}}\left(\boldsymbol{c}, \boldsymbol{p}^s\right)$, $B_{\boldsymbol{p}} := B_{\boldsymbol{p}}\left(\boldsymbol{c}, \boldsymbol{p}^s\right)$.

Likewise, $B\left(\boldsymbol{c}^r\left(\boldsymbol{p}^s, w'\right), \boldsymbol{p}^s\right) = w'$ gives

$$\begin{aligned} B^s_{\boldsymbol{c}} \boldsymbol{c}^r_{\boldsymbol{p}^s} &= -B^s_{\boldsymbol{p}^s} \\ B_{\boldsymbol{c}} \boldsymbol{c}^r_{w'} &= 1 \end{aligned}$$

Define the rational Hicksian action

$$\boldsymbol{h}^r\left(\boldsymbol{p}^s, \overline{u}\right) = \arg\min_{\boldsymbol{c}} B\left(\boldsymbol{c}, \boldsymbol{p}^s\right) \text{ s.t. } u\left(\boldsymbol{c}\right) \geq \overline{u} \tag{154}$$

and the corresponding Slutsky matrix, and the perceived $\boldsymbol{p}^s$

$$\boldsymbol{S}_j^r = \boldsymbol{h}_{\boldsymbol{p}_j^s}^r \left(\boldsymbol{p}^s, \overline{u}\right)_{|\overline{u}=v(\boldsymbol{p},\boldsymbol{p}^s,w)} \tag{155}$$

Define the dual expenditure function

$$e\left(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}\right) = \underset{\boldsymbol{c}|\boldsymbol{p}^s}{\text{smin}}\, B\left(\boldsymbol{c}, \boldsymbol{p}\right) \text{ s.t. } u\left(\boldsymbol{c}\right) \geq \overline{u} \tag{156}$$

We have the simpler representation:

$$e\left(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}\right) = B\left(\boldsymbol{h}^r\left(\boldsymbol{p}^s, \overline{u}\right), \boldsymbol{p}\right) \tag{157}$$

Proposition 15.1 (Shepard's lemma, nonlinear and behavioral). *Given the expenditure function* $e\left(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}\right) = B\left(\boldsymbol{h}^r\left(\boldsymbol{p}^s, \overline{u}\right), \boldsymbol{p}\right)$, *we have*

$$\begin{aligned} e_{\boldsymbol{p}_j} &= B_{\boldsymbol{p}_j} \\ e_{\boldsymbol{p}_j^s} &= \left(B_{\boldsymbol{c}} - B_{\boldsymbol{c}}^s\right) \cdot \boldsymbol{S}_j^r \end{aligned}$$

Proof. $e\left(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}\right) = B\left(\boldsymbol{h}^r\left(\boldsymbol{p}^s, \overline{u}\right), \boldsymbol{p}\right)$ gives: $e_{\boldsymbol{p}_j} = B_{\boldsymbol{p}_j}$, and

$$e_{\boldsymbol{p}_j^s} = B_{\boldsymbol{c}} \boldsymbol{h}_{\boldsymbol{p}_j^s}^r = \left(B_{\boldsymbol{c}} - B_{\boldsymbol{c}}^s\right) \boldsymbol{h}_{\boldsymbol{p}_j^s}^r$$

as (152) gives $B_{\boldsymbol{c}}^s \boldsymbol{h}_{\boldsymbol{p}_j^s}^r = 0$. $\square$

Proposition 15.2 (Generalized Roy's identity). *Given the function* $v\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right)$, *we have:*

$$\begin{aligned} \frac{v_{\boldsymbol{p}_j}}{v_w} &= -B_{\boldsymbol{p}_j} \\ \frac{v_{\boldsymbol{p}_j^s}}{v_w} &= \left(B_{\boldsymbol{c}}^s - B_{\boldsymbol{c}}\right) \cdot \boldsymbol{S}_j^r = D_j^s \end{aligned}$$

i.e. $\frac{v^s_{p^s_j}}{v^s_w} = \sum_i \left(\boldsymbol{B}_i^s - \boldsymbol{B}_i\right) \boldsymbol{S}_{ij}^r$.

Proof of Proposition 15.2

For a number $\overline{u}$, we have the identity $\overline{u} = v\left(\boldsymbol{p}, \boldsymbol{p}^s, e\left(\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}\right)\right)$ for all $\boldsymbol{p}, \boldsymbol{p}^s, \overline{u}$. Deriving w.r.t. $\boldsymbol{p}_j$ gives:

$$0 = v_{\boldsymbol{p}_j} + v_w e_{\boldsymbol{p}_j} = v_{p_j} + v_w B_{\boldsymbol{p}_j}$$

by the behavioral Shepard's lemma (Proposition 15.1).

Deriving w.r.t. $\boldsymbol{p}_j^s$ gives:

$$0 = v_{\boldsymbol{p}_j^s} + v_w e_{\boldsymbol{p}_j^s} = v_{\boldsymbol{p}_j^s} + v_w \left(B_{\boldsymbol{c}} - B_{\boldsymbol{c}}^s\right) \cdot \boldsymbol{S}_j^r$$

again by the behavioral Shepard's lemma (Proposition 15.1).□

Proposition 15.3 *(Marshallian demand) Given the Marshallian action* $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w)$*, we have:*

$$\boldsymbol{c}_{\boldsymbol{p}_j} = -\boldsymbol{c}_w B_{\boldsymbol{p}_j} \tag{158}$$
$$\boldsymbol{c}_{\boldsymbol{p}_j^s} = \boldsymbol{S}_j^r + \boldsymbol{c}_w D_j^s \tag{159}$$

i.e. $\boldsymbol{c}_{\boldsymbol{p}_j}^i = -B_{\boldsymbol{p}_j}^i \cdot \boldsymbol{c}_w$ *and* $c_{\boldsymbol{p}_j^s}^i = \boldsymbol{S}_j^{i,r} + c_w^i D_j^s$.

Proof. We have $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) = \boldsymbol{h}(\boldsymbol{p}^s, v(\boldsymbol{p}, \boldsymbol{p}^s, w))$, which implies

$$\boldsymbol{c}_w = \boldsymbol{h}_u v_w, \qquad \boldsymbol{c}_{\boldsymbol{p}_j} = \boldsymbol{h}_u v_{\boldsymbol{p}_j} \tag{160}$$

Because of Roy ($v_{\boldsymbol{p}_j} = -B_{\boldsymbol{p}_j} v_w$), we have: $\boldsymbol{c}_w B_{\boldsymbol{p}_j} = \boldsymbol{h}_u v_w B_{\boldsymbol{p}_j} = -\boldsymbol{h}_u v_{\boldsymbol{p}_j} = -\boldsymbol{c}_{\boldsymbol{p}_j}$, hence $\boldsymbol{c}_{\boldsymbol{p}_j} = -\boldsymbol{c}_w B_{\boldsymbol{p}_j}$.

Also, $\boldsymbol{c}_{p_j^s}(\boldsymbol{p}, \boldsymbol{p}^s, w) = \boldsymbol{h}(\boldsymbol{p}^s, v(\boldsymbol{p}, \boldsymbol{p}^s, w))$ gives:

$$\begin{aligned}
\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) &= \boldsymbol{h}_{\boldsymbol{p}_j^s} + \boldsymbol{h}_u v_{\boldsymbol{p}_j^s} \\
&= \boldsymbol{S}_j^r + \boldsymbol{c}_w \frac{v_{p_j^s}}{v_w} \text{ using } \boldsymbol{c}_w = \boldsymbol{h}_u v_w \\
&= \boldsymbol{S}_j^r + \boldsymbol{c}_w \left[(B_{\boldsymbol{c}}^s - B_{\boldsymbol{c}}) \cdot \boldsymbol{S}_j^r \right] \text{ using Proposition 15.2}
\end{aligned}$$

□

In the traditional model $v(\boldsymbol{p}, w) = \max_{\boldsymbol{c}} u(\boldsymbol{c}) + \lambda (w - B(\boldsymbol{c}, \boldsymbol{p}))$ implies $v_w = \lambda$. There is a deviation here, as indicated below.

Proposition 15.4 *(Envelope theorem, modified) Call* λ *the Lagrange multiplier such that* $u_{\boldsymbol{c}}(\boldsymbol{c}) = \lambda B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s)$*. We have:*

$$\frac{v_w}{\lambda} = B_{\boldsymbol{c}}^s \boldsymbol{c}_w = 1 + (B_{\boldsymbol{c}}^s - B_{\boldsymbol{c}}) \boldsymbol{c}_w \tag{161}$$

where $B_{\boldsymbol{c}}^s := B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p}^s)$ *and* $B_{\boldsymbol{c}} = B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{p})$.

Proof. We have: $v(\boldsymbol{p}, \boldsymbol{p}^s, w) = u(\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w))$, so

$$\begin{aligned}
\frac{v_w}{\lambda} &= \frac{1}{\lambda} u_{\boldsymbol{c}} \boldsymbol{c}_w = B_{\boldsymbol{c}}^s \boldsymbol{c}_w \\
&= B_{\boldsymbol{c}}^s \boldsymbol{c}_w + 1 - B_{\boldsymbol{c}} \boldsymbol{c}_w \text{ using } B_{\boldsymbol{c}} \boldsymbol{c}_w = 1 \text{ from } B(c, p) = w \\
&= 1 + (B_{\boldsymbol{c}}^s - B_{\boldsymbol{c}}) \boldsymbol{c}_w
\end{aligned}$$

□