# Online Appendix for "Optimal Taxation with Behavioral Agents"

Emmanuel Farhi and Xavier Gabaix

May 2018

Section 9 contains additional results on the paper, in particular on setups with heterogeneous agents and endogenous attention. Section 10 gives much more detail on the Mirrlees model. Section 11 contains proofs not included in the main paper. Sections 12 and 13 gives complements to consumer theory, with linear and nonlinear budget constraints respectively.

# 9 Additional Results

## 9.1 Complements on optimal tax with heterogeneous agents

### 9.1.1 Calibration: Optimal Ramsey tax with heterogeneous agents

Here we provide details to the calibration done in Section 3.1

With heterogeneous agents, the misperception is distributed as a 2-point distribution with the following properties:

$$m_i^h = \begin{cases} 1 & \text{with probability } p \\ a & \text{with probability } 1-p \end{cases}$$

with $a \in [0,1]$, and

$$\begin{aligned} \mathbb{E}[m_i^h] =& p \times 1 + (1-p) \times a = 0.25 \\ \mathbb{E}[(m_i^h)^2] =& p \times 1 + (1-p) \times a^2 = 0.25^2 + 0.13. \end{aligned}$$

These equations are satisfied at $p = .1877$ and $a = .0767$. We then take equation (13), with

$$\begin{aligned} S_i^h =& -\frac{c_i^h \psi_i}{q_i^h} \\ q_i^h =& p_i + m_i^h \tau_i \\ c_i^h =& (q_i^h)^{-\psi_i}. \end{aligned}$$

This yields:

$$\frac{\tau_i^*}{p_i} = \frac{1}{p_i} \frac{\left(\sum_h \pi_h c_i^h\right)\left(1-\frac{\gamma}{\lambda}\right)}{\psi_i \sum_h \pi_h \left[m_i^h \frac{c_i^h}{q_i^h}\left(1-\left(1-m_i^h\right)\frac{\gamma}{\lambda}\right)\right]},$$

where $\pi_h \in \{p, (1-p)\}$ is the fraction of agents of each type. Assume values $1-\frac{\gamma}{\lambda} = \Lambda = 1.25\%$ and $\psi_i = 1$. Then, under the case with heterogeneity ($p = .1877$ and $a = .0767$), we have $\frac{\tau_i^*}{p_i} = 0.0729$, or 7.29%. Under homogeneity with the same average misperception $m_i^h = .25$ for all agents, $\frac{\tau_i^*}{p_i} = 20.3\%$, for a ratio $.203/.0729 = 2.78$. When the taxes are fully salient, so $m_i^h = 1$ for all agents, then the optimal tax is 1.27%, giving a ratio $.0127/.0729 = .174$.

### 9.1.2 Optimal taxes with default tax perceptions and heterogeneous agents

**Nonzero default tax** It is sometimes important to introduce a distinction between the misperception of marginal tax changes and the misperception of the average level of taxes. To capture this possibility, we assume that perceived taxes are given by $\tau_i^s = m_i \tau_i + (1-m_i)\,\tau_i^d$, where $\tau_i^d$ is a default tax (which could be for instance the averate tax in the economy, or an averate of past tax rates). This change introduces a new additive term $-\frac{\tau_i^d}{p_i}\frac{(1-m_i)(1-\Lambda)}{m_i+(1-m_i)\Lambda}$ in the optimal tax formula (17) to correct for this new form of misperception, in the case of isoelastic utility.

To take a concrete example, suppose that we start from an equilibrium where taxes are optimal and default taxes are equal to true taxes. Imagine that there is a reduction in the need for public funds $\Lambda$, but that default taxes $\tau_i^d$ remain high at the pre-change level. Then lowering taxes induces agents to over-perceive the average level taxes, and creates a force for the government to lower taxes even further to correct this new bias.

**Heterogeneous attention and default taxes** Agent $h$ has utility $u^h(\mathbf{c}) = c_0^h + \sum_i U^h(c_i^h)$, with quadratic utility $U^h(c_i^h) = \frac{a^h c_i^h - \frac{1}{2}(c_i^h)^2}{\Psi^h}$. Agents are heterogeneous in attention $m_i^h$ and default taxes $\tau_i^{d,h}$. In particular, agent $h$ perceives tax as $\tau_i^{s,h} = m_i^h \tau_i + (1-m_i^h)\tau_i^{d,h}$. Each agent has the same social welfare weight $\gamma$. The demand for good $i$ is $c_i^h(\tau_i) = a^h - \Psi^h(p_i + \tau_i^{s,h}(\tau_i))$.

The Ramsey planning problem is

$$\max_{\tau} \quad L(\tau)$$

where

$$L(\tau) = \sum_{h=1}^{H} \gamma \sum_{i=1}^{n} \left(U^h(c_i^h(\tau_i)) - (p_i+\tau_i)c_i^h(\tau_i) + \lambda \tau_i c_i^h(\tau_i)\right)$$

First-order condition

$$\frac{\partial L}{\partial \tau_i} = \gamma \sum_{h=1}^{H} \left(U_{c_i^h}^h \frac{\partial c_i^h}{\partial \tau_i} - (p_i+\tau_i)\frac{\partial c_i^h}{\partial \tau_i} - c_i^h(\tau_i) + \frac{\lambda}{\gamma} c_i^h(\tau_i) + \frac{\lambda}{\gamma}\tau_i \frac{\partial c_i^h}{\partial \tau_i}\right) = 0$$

Let $\Lambda' \equiv \lambda/\gamma - 1$, and note that $\partial c_i^h/\partial \tau_i = -\Psi^h m_i^h$, we can rewrite the FOC as:

$$\begin{aligned}
\frac{\partial L}{\partial \tau_i} &= \gamma \sum_{h=1}^{H} \left( \left( \frac{a^h - c_i^h(\tau_i)}{\Psi^h} - p_i + \Lambda' \tau_i \right) (-\Psi^h m_i^h) + \Lambda' c_i^h(\tau_i) \right) \\
&= \gamma \sum_{h=1}^{H} \left( \left( \tau_i^{s,h} + \Lambda' \tau_i \right) (-\Psi^h m_i^h) + \Lambda' [a^h - \Psi^h (p_i + \tau_i^{s,h}(\tau_i))] \right) \\
&= \gamma \sum_{h=1}^{H} \left( -\Psi^h m_i^h (m_i^h + \Lambda') \tau_i - \Psi^h m_i^h (1 - m_i^h) \tau_i^{d,h} + \Lambda' [a^h - \Psi^h p_i] - \Lambda' \Psi^h m_i^h \tau_i - \Lambda' \Psi^h (1 - m_i^h) \tau_i^{d,h} \right) \\
&= \gamma \sum_{h=1}^{H} \left( -\Psi^h m_i^h (m_i^h + 2\Lambda') \tau_i - \Psi^h (1 - m_i^h)(m_i^h + \Lambda') \tau_i^{d,h} + \Lambda' [a^h - \Psi^h p_i] \right) \\
&= -\gamma H \left( \mathbb{E}[\Psi^h m_i^h (m_i^h + 2\Lambda')] \tau_i + \mathbb{E}[\Psi^h (1 - m_i^h)(m_i^h + \Lambda') \tau_i^{d,h}] - \Lambda' \mathbb{E}[a^h - \Psi^h p_i] \right) = 0
\end{aligned}$$

We can solve explicitly for optimal Ramsey tax in this case:

$$\tau_i = \frac{\Lambda' \mathbb{E}[a^h - \Psi^h p_i] - \mathbb{E}[\Psi^h (1 - m_i^h)(m_i^h + \Lambda') \tau_i^{d,h}]}{\mathbb{E}[\Psi^h m_i^h (m_i^h + 2\Lambda')]} \tag{47}$$

### 9.1.3 Pigouvian Nudges with heterogeneous agents

We start with the Pigouvian example of Section 3.2. There is only one taxed good $n = 1$. We use the specialization of the general model developed in Section 2.7. We assume no redistribution or revenue-raising motives ($\gamma^h = \beta^h = \lambda$).

We model the nudge as a psychological tax, as in Section 2.4. Agent $h$'s demand is given by $\arg\max_c U(c) - \left(p + \eta^h \chi\right) c$, where $\chi$ is the nudge and $\eta^h$ is the agent's nudgeability. We use quadratic utilities, exactly as in the Pigouvian taxes of Section 3.2. The demand of a consumer can then be expressed as $c^h(\tau, \chi) = a^h - \Psi\left(p + \eta^h \chi\right)$, where $\eta^h$ is the nudgeability of agent $h$. We apply the optimal nudge formula (11).

When the nudge is the only instrument, the optimal nudge is

$$\chi = \frac{\mathbb{E}\left[\xi^h \eta^h\right]}{\mathbb{E}\left[\eta^{h2}\right]} = \frac{\mathbb{E}\left[\xi^h\right] \mathbb{E}\left[\eta^h\right] + cov\left(\xi^h, \eta^h\right)}{\mathbb{E}\left[\eta^h\right]^2 + var\left[\eta^h\right]}, \tag{48}$$

where again $\mathbb{E}$ denotes the average over agents $h$.[74]

Heterogeneities in nudgeability determine how well targeted the nudge is to the internality/externality. The optimal nudge is stronger when it is well-targeted, in the sense that nudgeable agents are also those with high internality/externality (higher $cov\left(\xi^h, \eta^h\right)$). The optimal nudge is weaker when

[74] The intermediate steps are as follows. Using $\boldsymbol{c}_\chi^h = -\Psi \eta^h$, $\tau = 0, \tau^{b,h} = \tau^{X,h} - \chi \eta^h$, we get $\frac{\partial L}{\partial \chi}(\boldsymbol{\tau}, \chi) = \sum_h \left[\lambda \boldsymbol{\tau} - \lambda \boldsymbol{\tau}^{\xi,h} - \beta^h \boldsymbol{\tau}^{b,h}\right] \cdot \boldsymbol{c}_\chi^h = \lambda \sum_h \left[0 - \tau^{X,h} + \chi \eta^h\right] \Psi \eta^h$.

there is more heterogeneity in nudgeability (higher $var\left[\eta^{h^2}\right]$).[75]

### 9.1.4 Nudges vs. Taxes with Redistributional Concerns

**Jointly optimal nudges and taxes** We normalize $p = 1$. Agent $h$ has utility $u^h(\mathbf{c}) = c_0^h + \frac{a^h c^h - \frac{1}{2}(c^h)^2}{\Psi}$, so that $\beta^h = \gamma^h$ and $c^h\left(\tau, \tau^{\chi,h}\right) = a^h - \Psi\left(m^h\tau + \tau^{\chi,h}\right)$. We investigate the optimal joint policy using both nudges and taxes on goods $c$. One can show that

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

As a result, if $\gamma^h = \lambda$ so that there are no revenue raising or redistributive motives, then taxes and nudges are substitutes. Taxes and nudges are complements if and only if $\mathbb{E}\left[\left(\lambda - \gamma^h\left(1 - m^h\right)\right)\eta^h\right] \leq 0$. Nudges and taxes can be complement if social marginal utility of income $\gamma^h$ and nudgeability $\eta^h$ are positively correlated. Loosely speaking, if poor agents (with a high $\gamma^h$) are highly nudgeable, then taxes and nudges can become complements, because in that case, nudges reduces the consumption of poor nudged agents, thereby improving the redistributive incidence of the tax. We next state the exact values of taxes and nudges, in the case $\gamma^h = \lambda$.

**Proposition 9.1** *Assume $\gamma^h = \lambda$. Then jointly optimal nudges and taxes are given by the following formulas*

$$\tau = \frac{\mathbb{E}\left[(\eta^h)^2\right]\mathbb{E}\left[\tau^{X,h}m^h\right] - \mathbb{E}\left[\eta^h m^h\right]\mathbb{E}\left[\tau^{X,h}\eta^h\right]}{\mathbb{E}\left[(\eta^h)^2\right]\mathbb{E}\left[(m^h)^2\right] - \left(\mathbb{E}\left[\eta^h m^h\right]\right)^2},$$
$$\chi = \frac{\mathbb{E}\left[\tau^{X,h}\eta^h\right]\mathbb{E}\left[(m^h)^2\right] - \mathbb{E}\left[\tau^{X,h}m^h\right]\mathbb{E}\left[\eta^h m^h\right]}{\mathbb{E}\left[(\eta^h)^2\right]\mathbb{E}\left[(m^h)^2\right] - \left(\mathbb{E}\left[\eta^h m^h\right]\right)^2}.$$

The more powerful the nudge is for high-internality agents (the higher is $\mathbb{E}\left[\tau^{X,h}\eta^h\right]$, keeping all other moments constant), the more optimal policy relies on the nudge and the less it relies on the tax (the higher is $\chi$, the lower is $\tau$). Symmetrically, if the better perceived is the tax by high-internality people (the higher is $\mathbb{E}\left[\tau^{X,h}m^h\right]$), the more optimal policy relies on the tax and the less it relies on the nudge.

The more heterogeneity there is in the perception of taxes (the higher is $\mathbb{E}\left[(m^h)^2\right]$, holding all other moments constant), the less targeted the tax is to the internality/externality, and, as a result, the lower is the optimal tax $\tau$, and under certain conditions, the higher the optimal nudge $\chi$.[76] Similarly, the more heterogeneity there is in nudgeability (the higher is $\mathbb{E}\left[(\eta^h)^2\right]$, holding all other moments constant), then lower is the optimal nudge $\chi$, and, under similar conditions, the higher is the optimal tax $\tau$.

[75] Some recent studies study the demographic covariates of nudgeability (Chetty et al. (2014), Beshears et al. (2016)), and it would be good to measure the covariance between nudgeability and internality.

[76] The condition is $\mathbb{E}\left[\tau^{X,h}m^h\right]\mathbb{E}\left[\eta^{h^2}\right] \geq \mathbb{E}\left[\tau^{X,h}\eta^h\right]\mathbb{E}\left[\eta^h m^h\right]$. It is verified if $\eta^h, m^h, \tau^{X,h}$ are independent.

In the general case, we assume heterogenous welfare weights with $\mathbb{E}\left[\gamma^h\right] = \lambda$.

**Proposition 9.2** *The optimal tax and nudge satisfy*

$$\tau = \frac{\mathbb{E}\left[\gamma^h \eta^{h2}\right] \mathbb{E}\left[\lambda \tau^{X,h} m^h - \sigma_{\gamma,c/\Psi}\right] - \mathbb{E}\left[\gamma^h \eta^h m^h\right] \mathbb{E}\left[\lambda \tau^{X,h} \eta^h\right]}{\mathbb{E}\left[\gamma^h \eta^{h2}\right] \mathbb{E}\left[\gamma^h m^{h2} - \sigma_{\gamma,m}\right] - \mathbb{E}\left[\gamma^h \eta^h m^h\right] \mathbb{E}\left[\gamma^h \eta^h m^h - \sigma_{\gamma,\eta}\right]}$$

$$\chi = \frac{\mathbb{E}\left[\lambda \tau^{X,h} \eta^h\right] \mathbb{E}\left[\gamma^h m^{h2} - \sigma_{\gamma,m}\right] - \mathbb{E}\left[\lambda \tau^{X,h} m^h - \sigma_{\gamma,c/\Psi}\right] \mathbb{E}\left[\gamma^h \eta^h m^h - \sigma_{\gamma,\eta}\right]}{\mathbb{E}\left[\gamma^h \eta^{h2}\right] \mathbb{E}\left[\gamma^h m^{h2} - \sigma_{\gamma,m}\right] - \mathbb{E}\left[\gamma^h \eta^h m^h\right] \mathbb{E}\left[\gamma^h \eta^h m^h - \sigma_{\gamma,\eta}\right]}.$$

**Proof**. As in Section 2.7, we call $\xi^h = \tau^{X,h} = \frac{\gamma^{\xi,h}}{\lambda} \tau^{I,h} + \tau^{\xi,h}$ as the sum of the internality plus externality. Individual $h$ creates an externality plus internality. We have successively,

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

Proposition 2.2 gives optimal tax, using $S^{H,h} = -\Psi m^h$:

$$\begin{aligned}
\frac{\partial L}{\partial \tau} &= \mathbb{E} \sum_h \left(\lambda - \gamma_h\right) c^h - \lambda \left(\tau - \tau^{\xi,h} - \widetilde{\tau}^{\xi,h}\right) \Psi m^h \\
&= \mathbb{E} \sum_h \left(\lambda - \gamma_h\right) c^h - \left[\left(\lambda - \gamma^h \left(1 - m^h\right)\right) \tau - \lambda \tau^{X,h} + \gamma^h \tau^{\chi,h}\right] \Psi m^h. \qquad (49)
\end{aligned}$$

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

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

Proposition 2.3 gives optimal nudge:

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

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

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

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:

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

Solving for the two unknowns $\tau$ and $\chi$ gives the formulas.

**Nudges vs. taxes** We now ask how to choose, if one must, between nudges and taxes. We could analyze this question using the model outlined just above, comparing the relative merits of nudges and taxes in terms of internality targeting and redistributive incidence. Instead, we choose to investigate this question in the context of a model with no heterogeneity, but where the nudges are potentially aversive.

We augment the example of Section 3.4 with aversive nudges. We use quadratic utility functions. We use the nudge as a tax model developed in Section 2.4. We again normalize $p_i = 1$.

For concreteness, we interpret the harmful good (good 1) as cigarettes. We extend the model to account for the possibility that the nudge may directly create an aversive reaction (perhaps via a disgusting image of a cancerous lung), which we capture as a separable utility cost $\iota^h \chi c_i$ so that experienced utility is now

$$u^h(\boldsymbol{c}, \chi) = u^h(\boldsymbol{c}) - \iota^h \chi c_i,$$

where $\iota^h \chi c_i$ is the nudge aversion term. And we assume that there is no heterogeneity across agents.

The next proposition formalizes how nudge aversion changes the relative attractiveness of nudges vs. taxes. The planner must choose between two instruments to discourage cigarette consumption: a weakly positive tax ($\tau \geq 0$) or an aversive nudge ($\chi \geq 0$).

**Proposition 9.3** ("Nudge the poor, tax the rich") *Consider a good with a "bad" internality (e.g. cigarettes). Suppose that at most one of two instruments (nudges and nonnegative taxes) can be used to correct this internality. And suppose that there is no heterogeneity across agents. Then an optimal tax is superior to an optimal nudge if and only if*

$$\frac{\lambda - \gamma^h}{m^h} > \frac{-\iota^h \gamma^h}{\eta^h}. \tag{50}$$

This proposition captures a new interesting trade-off between taxes and nudges. Both taxes and nudges correct internalities. But taxes also raise revenues on the agents consuming the good under consideration, which is desirable if $\lambda > \gamma^h$ but undesirable if $\lambda < \gamma^h$. Nudges do not raise revenues, and instead directly reduce utility.

When $\lambda > \gamma^h$, taxes dominate nudges as taxes have desirable side effects by raising revenues while nudges have adverse side effects by reducing utility. But when $\lambda < \gamma^h$ taxes and nudges both have undesirable side effects. Taxes dominate nudges when the desire to redistribute income towards agents consuming the good associated with the internality is weak ($\gamma^h - \lambda$ is low), and when these agents are attentive to the tax ($m^h$ is high). Nudges dominate taxes when nudge aversion is low ($\iota^h$ is low) and when agents are easily nudged ($\eta^h$ is high).

**Proof**. We apply our tax formulas (11):

$$\begin{aligned}
\frac{\partial L}{\partial \chi} &= -\sum_h \iota^h \gamma^h c^h - \sum_h \left[ \left( \lambda - \gamma^h \left( 1 - m^h \right) \right) \tau - \lambda \tau^{X,h} + \gamma^h \chi \eta^h \right] \Psi \tau_\chi^{\chi,h} \\
&= -\sum_h \iota^h \gamma^h c^h - \Psi \sum_h \left[ \left( \lambda - \gamma^h \left( 1 - m^h \right) \right) \tau - \lambda \tau^{X,h} + \gamma^h \chi \eta^h \right] \eta^h \qquad (51) \\
&= \sum_h \left[ -\iota^h \gamma^h c^h - \Psi x^h \eta^h \right], \qquad (52)
\end{aligned}$$

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

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

where $\pi, \pi'$ are Lagrangian multipliers.

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

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

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 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 result is a bit knife-edge.

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

### 9.1.5 Discouragement formula

In the traditional model without behavioral biases we can use the symmetry of the Slutsky matrix $\boldsymbol{S}^{r,h}$ to write $\boldsymbol{\tau} \cdot \boldsymbol{S}_i^{r,h} = \sum_j \tau_j S_{ji}^{r,h}$ as $\boldsymbol{\tau} \cdot \boldsymbol{S}_i^{r,h} = \sum_j \tau_j S_{ij}^{r,h}$. We can then rewrite the optimal tax formula of Proposition 2.1 in "discouragement" form as

$$\frac{-\sum_{h,j} \tau_j S_{ij}^{r,h}}{c_i} = 1 - \frac{\bar{\gamma}}{\lambda} - cov(\frac{\gamma^h}{\lambda}, \frac{H c_i^h}{c_i}), \tag{53}$$

The left-hand side is the discouragement index of good $i$, which loosely captures how much the consumption of good $i$ is discouraged by the taxes $\tau_j$ on all the different commodities $j$. The right-hand side indicates that in the absence of distributive concerns (homogenous $\gamma^h = \gamma$), all goods should be

uniformly discouraged in proportion to the relative intensity $1-\frac{\bar{\gamma}}{\lambda}$ of the raising revenue objective. With redistributive concerns (heterogenous $\gamma^h$), goods that are disproportionately consumed by agents that society tries to redistribute towards (agents with a high $\gamma^h$) should be discouraged less.

## 9.2 Complements on Endogenous Attention: Attention as a good

### 9.2.1 Interpreting attention as a good

To capture attention and its costs, we propose the following reinterpretation of the general framework. We imagine that we have the decomposition $\boldsymbol{c}=(\boldsymbol{C},\boldsymbol{m})$, where $\boldsymbol{C}$ is the vector of traditional goods (champagne, leisure), and $\boldsymbol{m}$ is the vector of attention (e.g. $m_i$ is attention to good $i$). We call $I^{\boldsymbol{C}}$ (respectively $I^m$) the set of indices corresponding to traditional goods (respectively attention). Then, all the analyses and propositions apply without modification.

This flexible modeling strategy allows us to capture many potential interesting features of attention. The framework allows (but does not require) attention to be chosen and react endogenously to incentives in a general way (optimally or not). It also allows (but does not require) attention to be produced, purchased and taxed.

We find it most natural to consider the case where attention is not produced, cannot be purchased, and cannot be taxed. This case can be captured in the model by imposing that $p_i=\tau_i=0$ for $i\in I^m$.

It is useful to consider two benchmarks. The first benchmark is "no attention cost in welfare," where attention is endogenous (given by a function $\boldsymbol{m}(\boldsymbol{q},w)$) but its cost is assumed not to directly affect welfare so that $u(\boldsymbol{C},\boldsymbol{m})=U(\boldsymbol{C})$. For instance, as a decision vs. experienced utility generalization of the example of the previous paragraph, we could have $\boldsymbol{m}(\boldsymbol{q},w)=\arg\max_{\boldsymbol{m}} u^s(\boldsymbol{C}(\boldsymbol{q},w,\boldsymbol{m}),\boldsymbol{m})$, where $u^s(\boldsymbol{C},\boldsymbol{m})=U(\boldsymbol{C})-g(\boldsymbol{m})$, but still $u(\boldsymbol{C},\boldsymbol{m})=U(\boldsymbol{C})$. In that view, people use decisions heuristics that can respond to incentives, but the cost of those decision heuristics is not counted in the utility function. In this benchmark, we have $\tau_i^b=0$ for $i\in I^m$.

The second benchmark is "attention cost in welfare". For simplicity, we outline this case under the extra assumption, which is easy to relax, that attention is allocated optimally. We suppose that there is a primitive choice function $\boldsymbol{C}(\boldsymbol{q},w,\boldsymbol{m})$ for traditional goods that depends on attention $\boldsymbol{m}=(m_1,...,m_A)$ so that $\boldsymbol{c}(\boldsymbol{q},w,\boldsymbol{m})=(\boldsymbol{C}(\boldsymbol{q},w,\boldsymbol{m}),\boldsymbol{m})$.[77] Attention $\boldsymbol{m}=\boldsymbol{m}(\boldsymbol{q},w)$ is then chosen to maximize $u(\boldsymbol{C}(\boldsymbol{q},w,\boldsymbol{m}),\boldsymbol{m})$. This generates a function $\boldsymbol{c}(\boldsymbol{q},w)=(\boldsymbol{C}(\boldsymbol{q},w,\boldsymbol{m}(\boldsymbol{q},w)),\boldsymbol{m}(\boldsymbol{q},w))$. In that benchmark, attention costs are incorporated in welfare.[78] For instance we might consider a

[77] For instance, in a misperception model, attention operates by changing the perceived price $\boldsymbol{q}^s(\boldsymbol{q},w,\boldsymbol{m})$ which in turn changes consumption as $\boldsymbol{C}(\boldsymbol{q},w,\boldsymbol{m})=\boldsymbol{C}^s(\boldsymbol{q},\boldsymbol{q}^s(\boldsymbol{q},w,\boldsymbol{m}),w)$.

[78] The first order condition characterizing the optimal allocation of attention can be written as $\boldsymbol{\tau}^b\cdot\boldsymbol{c}_{m_j}(\boldsymbol{q},w,\boldsymbol{m})=0$ for all $j\in\{1,...,A\}$. This condition can be re-expressed more conveniently by introducing the following notation: we call $k(i)$ the index $k\in I^m$ corresponding to dimension $i\in\{1,...,A\}$ of attention. We then get $\sum_{i\in I^C}\tau_i^b\boldsymbol{C}_{m_j}(\boldsymbol{q},w,\boldsymbol{m})+\tau_{k(j)}^b=0$ for all $j\in\{1,...,A\}$.

separable utility function $u(\boldsymbol{C},\boldsymbol{m}) = U(\boldsymbol{C}) - g(\boldsymbol{m})$ for some cost function $g(\boldsymbol{m})$. A non-separable $u$ might capture that attention is affected by consumption (e.g., of coffee) and attention affects consumption (by needing aspirin).

The tax formula (8) has a term $(\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}}^{b,h})\cdot \boldsymbol{S}_i^{C,h} = \sum_{k\in I^m \bigcup I^c}(\tau_k - \widetilde{\tau}_k^{b,h})\boldsymbol{S}_{ki}^{C,h}$, a sum that includes the "attention" goods $k \in I^m$. As attention is assumed to have zero tax, we have $\tau_k = 0$ for $k \in I^m$. The term $\widetilde{\tau}_k^{b,h}$, which accounts for potential misoptimization in the allocation of attention, requires no special treatment. However, two polar special cases are worth considering that simplify the calculations. First, consider the "no attention cost in welfare" case. In this case we saw that $\widetilde{\tau}_k^{b,h} = 0$ for $k \in I^m$. Together with $\tau_k = 0$ for $k \in I^m$, this implies that $(\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}}^{b,h})\cdot \boldsymbol{S}_i^{C,h} = \sum_{k\in I^C}(\tau_k - \widetilde{\tau}_k^{b,h})S_{ki}^{C,h}$ is the sum restricted to commodities. Second, consider the "optimally allocated attention" case. Then (see Proposition 9.5) $(\boldsymbol{\tau} - \widetilde{\boldsymbol{\tau}}^{b,h})\cdot \boldsymbol{S}_i^{C,h} = \sum_{k\in I^C}(\tau_k S_{ki}^{C,h} - \widetilde{\tau}_k^{b,h} S_{ki|m}^{C,h})$, where $\boldsymbol{S}_{i|\boldsymbol{m}}^{C,h}$ is a Slutsky matrix holding attention constant, which is in general different from $\boldsymbol{S}_i^{C,h}$. For tax revenues, the full Slutsky matrix, including changes in attention, matters (the term $\tau_k S_{ki}^{C,h}$). However, for welfare, when attention is assumed to be optimally allocated, it is the Slutsky matrix holding attention constant that matters (the term $\widetilde{\tau}_k^{b,h} S_{ki|\boldsymbol{m}}^{C,h}$). This is a version of the envelope theorem.

**Characterizing optimal allocation of attention** Suppose that we have a constraint: $\boldsymbol{c} = \boldsymbol{c}(\boldsymbol{p}, w, \theta)$ for some parameter $\theta$. 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.[79]

**Proposition 9.4** (Characterizing optimal allocation of attention) *The first order condition for the optimal allocation of parameter $\theta$ (i.e., $\theta(\boldsymbol{p}, w) = \arg\max_\theta u(\boldsymbol{c}(\boldsymbol{p}, w, \theta))$) is:*

$$\boldsymbol{\tau}^b \cdot \boldsymbol{c}_\theta(\boldsymbol{p}, w, \theta) = 0. \tag{54}$$

**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 $u_{\boldsymbol{c}} \cdot \boldsymbol{c}_\theta = 0$. □

**Proposition 9.5** (Value of $D_j$ when attention is optimal). *When attention is of the form $\boldsymbol{c}(\boldsymbol{p}, w, \theta) =$*

[79] 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.

$(\boldsymbol{C}(\boldsymbol{p}, w, \boldsymbol{m}(\theta)), \boldsymbol{m}(\theta))$, *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 behavioral 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

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

□

**"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 that attention $\boldsymbol{m}$ just moves with prices, but as an automatic process whose "cost" is not counted: that is, $u(\boldsymbol{C}, \boldsymbol{m}) = u(\boldsymbol{C})$ and attention has 0 price, $\boldsymbol{p}_{\boldsymbol{m}} = 0$. 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 9.6** (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*

$$-D_j = \left(\boldsymbol{\tau}^b_C, 0\right) \cdot \boldsymbol{S}^H_j(\boldsymbol{p}, w) = \sum_{i=1}^{n} \tau^b_{C_i} S^H_{ij} = \left(\boldsymbol{\tau}^b_{\boldsymbol{C}}, 0\right) \cdot \boldsymbol{S}^C_j(\boldsymbol{p}, w) = \left(\boldsymbol{\tau}^b_{\boldsymbol{C}}, 0\right) \cdot \boldsymbol{c}_j(\boldsymbol{p}, w).$$

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

**Proof**
We have $\boldsymbol{\tau}^b = \left(\boldsymbol{\tau}^b_{\boldsymbol{c}}, \boldsymbol{\tau}^b_{\boldsymbol{m}}\right) = \left(\boldsymbol{\tau}^b_{\boldsymbol{c}}, 0\right)$ as $u_{\boldsymbol{m}} = 0$. So, $-D_j = \boldsymbol{\tau}^b \cdot \boldsymbol{S}^H_j(\boldsymbol{p}, w) = \boldsymbol{\tau}^b_{\boldsymbol{C}} \cdot \boldsymbol{S}^H_{C,j}$. □

**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 9.5 with $\boldsymbol{m}(\theta) = \theta$. This gives: $-D_j = \boldsymbol{\tau}^b_{\boldsymbol{C}} \cdot \boldsymbol{S}^{H,\boldsymbol{m}}_{\boldsymbol{C}j}$ with

$$\boldsymbol{S}^H_{\boldsymbol{C}j|\boldsymbol{m}} = \boldsymbol{S}^r \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w, \boldsymbol{m}), \tag{55}$$

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

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

$$-D_j = \boldsymbol{\tau}^b_{\boldsymbol{C}} \cdot \boldsymbol{S}^H_{\boldsymbol{C}j}.$$

*When attention is not necessarily optimal*, we also have (from (44)), using again decomposition $\boldsymbol{\tau}^b = \left(\boldsymbol{\tau}^b_{\boldsymbol{c}}, \boldsymbol{\tau}^b_{\boldsymbol{m}}\right)$ :

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

where $\boldsymbol{S}^H_{\boldsymbol{C}j} = \boldsymbol{S}^r \cdot \boldsymbol{p}^s_{p_j}(\boldsymbol{p}, w)$, where now the total derivative matters, including the variable attention.

### 9.2.2 Attention as a good: examples

In this subsection we normalize the pre-tax price to 1.

**Optimal taxes with endogenous attention: the case of small taxes** Given attention $m(\tau)$, the perceived tax is $\tau^s(\tau) = \tau m(\tau)$, and demand is $c(\tau) = y(1 - \psi m(\tau)\tau)$. We assume that attention comes from an optimal cost-benefit analysis:

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

The first term represents the private costs of misunderstanding taxes, $-\frac{1}{2}\psi y(\tau - \tau^s)^2$, while the term $-g(m)$ is the psychic cost of attention, $g(m)$ (see Gabaix (2014)). The planner's problem is $\max_\tau L(\tau)$ with

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

where $A = 1$ in the "optimally allocated attention" case and $A = 0$ in the "no attention cost in welfare" case. In the "fixed attention" case, $m(\tau)$ is fixed with $m'(\tau) = 0$, and $g(m) = 0$. The optimal tax satisfies

$$L'(\tau) = -\psi y m(\tau)\tau(m(\tau) + \tau m'(\tau)) - Ag'(m(\tau))m'(\tau) + \Lambda y = 0.$$

In the "optimally allocated attention" case, we use the agent's first order condition $g'(m(\tau)) =$

$\psi y \tau^2 (1 - m(\tau))$ and $A = 1$, and the optimal tax is

$$\tau^{m,*} = \frac{\Lambda/\psi}{m(\tau)^2 + \tau m'(\tau)}. \tag{56}$$

In the "no attention cost in welfare case," $A = 0$, the optimal tax is

$$\tau^{m,0} = \frac{\Lambda/\psi}{m(\tau)^2 + \tau m(\tau) m'(\tau)}. \tag{57}$$

When attention is fixed, the optimal tax is

$$\tau^{m,F} = \frac{\Lambda/\psi}{m(\tau)^2}. \tag{58}$$

**Proposition 9.7** *In the interior region where attention has an increasing cost ($\tau m(\tau) m'(\tau) > 0$), the optimal tax is lowest when attention is chosen optimally and its cost is taken into account in welfare; intermediate in the "no attention cost in welfare" case; and largest with fixed attention—$\tau^{m,*} < \tau^{m,0} < \tau^{m,F}$.*

When attention's cost is taken into account, the planner chooses lower taxes $\tau^{m,*} < \tau^{m,0}$ to minimize both consumption distortions and attention costs. [80] Plainly, the tax is higher when attention is variable than when attention is fixed—this is basically because demand is more elastic then $(-\frac{p}{c}\frac{\partial c}{\partial \tau} = \psi(m(\tau) + \tau m'(\tau)))$.

For more illustrations, see section 9.2.3 for completely worked out linear-quadratic and isoelastic examples.

### 9.2.3 Worked out examples of endogenous attention

**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 $q_1^s(p_1, m) = p_1^d + m\tau_1$, where $\tau_1$ 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 $U'(c_1) = q^s$, so $c_1^r(p^s) = a - \Psi q^s$. The perceived tax is:

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

and demand is

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

[80] The example allows to appreciate the Slutsky matrix with or without constant attention. The Slutsky matrix with constant $m$ has $S^C_{11|m} = \frac{\partial c(1+\tau, m)}{\partial \tau} = -\psi c m$, while the Slutsky matrix with variable $m$ has $S^C_{11} = \frac{dc(1+\tau, m(\tau))}{d\tau} = -\psi c(m + \tau m'(\tau_1))$.

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

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

whose first order condition is:

$$g'\left(m\right) = \Psi\left(1-m\right)\tau_1^2. \tag{59}$$

The Slutsky matrix with constant $m$ has:

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

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

$$\begin{aligned} S^H_{11} &= \frac{dc_1\left(p_1^d + m\tau_1, m\left(\tau_1\right)\right)}{d\tau_1} = -\Psi\left(m + \tau m'\left(\tau_1\right)\right) \\ S^H_{21} &= \frac{\partial m}{\partial\tau_1} = m'\left(\tau_1\right). \end{aligned}$$

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

$$\begin{aligned} \tau - \tau^b &= \left(0, \tau_1 m, -Ag'\left(m\right)\right) \\ S_1^H &= \left(0, -\Psi\left(m + \tau_1 m'\left(\tau_1\right)\right), m'\left(\tau_1\right)\right). \end{aligned}$$

Applying Proposition 2.1 gives:

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

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

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

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, $A = 0$

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

so

$$\tau_1^s = \frac{-c_1\Lambda}{S^H_{11}} = \frac{\Lambda}{\left(m + \tau_1 m'\left(\tau_1\right)\right)\psi_1}, \qquad \tau_1 = \frac{\Lambda}{\left(m^2 + \tau_1 m m'\left(\tau_1\right)\right)\psi_1}.$$

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

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

**An isoelastic example** We now work out completely an isolastic example. Take decision utility have $u^s(c_0, c_1, m) = c_0 + U(c_1) - g(m)$, with $U(c) = \frac{c^{1-1/\psi} - 1}{1 - 1/\psi}$ and attention technology $q_1^s(p_1, m) = p_1^d + m\tau_1$, where $\tau_1$ 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 $p_0 = 1$. The perceived tax is:

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

and demand is

$$c_1 = (p_1 + m(\tau_1)\tau_1)^{-\psi}.$$

The Slutsky matrix with constant $m$ has:

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

where (to leverage the calculations already done for the quadratic utility case) we define:

$$\Psi = \psi \frac{c_1}{q_1^s},$$

with $q_1^s = p_1 + m_1(\tau_1)$, while with variable attention $m(\tau)$, we have:

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

Then, we have: $\tau^b = (0, q - q^s, Ag'(m)) = (0, \tau_1(1-m), Ag'(m))$, and given $\tau = (0, \tau_1, 0)$, and $\tilde{\tau}^b = \left(1 - \frac{\beta}{\lambda}\right)\tau = (1-\Lambda)\tau$ so

$$\begin{aligned}\tau - \tilde{\tau}^b = \tau - (1-\Lambda)\tau^b &= (0, \tau_1(1-(1-m)(1-\Lambda)), -(1-\Lambda)Ag'(m)) \\ &= (0, \tau_1(m + \Lambda(1-m)), -(1-\Lambda)Ag'(m)) \\ S_1^H &= (0, -\Psi(m + \tau_1 m'(\tau_1)), m'(\tau_1)).\end{aligned}$$

Applying Proposition 2.1 gives (with $\lambda = 1$,$\gamma = 1 - \Lambda$)

$$\begin{aligned}\frac{\partial L(\boldsymbol{\tau}, w)}{\partial \tau_1} &= (\lambda - \gamma)c_1 + \lambda\left(\tau - \tilde{\tau}^b\right) \cdot S_1^H \\ &= \Lambda c_1 - \Psi\tau_1(m + \Lambda(1-m))(m + \tau_1 m'(\tau_1)) - Ag'(m)m'(\tau_1).\end{aligned}$$

Define

$$\psi_1(c_1) = \frac{\Psi}{c_1} = \frac{\psi}{q_1^s}.$$

First, when $m_1$ is exogenous, we verify our formula (17):

$$0 = \Lambda - \frac{\psi}{q_1^s}\tau_1(m + \Lambda(1-m))m,$$

i.e. $\frac{\tau_1}{q_1^s} = \frac{\Lambda}{\psi_1(m+\Lambda(1-m))m}$, which is equivalent to (17).

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

$$\begin{aligned}\frac{\partial L(\boldsymbol{\tau}, w)}{\partial \tau_1} &= \Lambda c_1 - \Psi\tau_1(m + \Lambda(1-m))(m + \tau_1 m'(\tau_1)) \\ &= \Lambda c_1 - \frac{\psi}{1 + m\tau} c\tau(m + \Lambda(1-m))(m + \tau m'(\tau)),\end{aligned}$$

so

$$\tau^0 = \frac{\Lambda/\psi}{m^2 + \tau m m'(\tau) + \Lambda\left((1-m)(m + \tau m') - \frac{m}{\psi}\right)}. \tag{60}$$

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 $\boldsymbol{S}_{j|m}^H (\boldsymbol{p}, w, m) = -\Psi m$.

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

$$\frac{\partial L(\boldsymbol{\tau}, w)}{\partial \tau_1} = \Lambda c_1 - \frac{\psi c_1}{1 + m\tau} \tau (m(m + \Lambda(1-m)) + \tau m'(\tau)),$$

so $\frac{\partial L(\boldsymbol{\tau}, w)}{\partial \tau_1} = 0$ gives

$$\tau^1 = \frac{\Lambda/\psi}{m^2 + \tau m'(\tau) + \Lambda\left((1-m)m - \frac{m}{\psi}\right)}. \tag{61}$$

We can compare this to the following re-rewrite of the optimal tax in the no-attention in welfare case:

$$\tau^0 = \frac{\Lambda/\psi}{m^2 + \tau_1 m'(\tau_1) + \Lambda\left((1-m)m - \frac{m}{\psi}\right) - \tau(1-m)m' + \Lambda(1-m)\tau m'} \tag{62}$$

$$= \frac{\Lambda/\psi}{m^2 + \tau_1 m'(\tau_1) + \Lambda\left((1-m)m - \frac{m}{\psi}\right) - (1-\Lambda)\tau(1-m)m'}. \tag{63}$$

**Proposition 9.8** *The optimal tax is lower in the "attention in welfare" case than in the "no attention in welfare" case.*

**Proof**. Suppose the opposite, i.e. $\tau^1(m_1) \geq \tau^0(m_0)$.

We observe that, for all $m$, (i) $\tau^0(m) \geq \tau^1(m)$ (ii) $\tau^1(m)$ is decreasing in $m$, and (iii)$m(\tau)$ is weakly increasing in $\tau$,

Then

$$\tau^1(m_1) \geq \tau^0(m_0) \geq \tau^1(m_0),$$

hence, as $\tau^1$is decreasing in $m$, we have $m_1 < m_0$. As $m(\tau)$is increasing, this implies $\tau_1 < \tau_0$. We've reached a contradiction.

There's a function $m(\tau)$; it's inverse is $\tau(m)$ we define

$$\tau^1(m) = \frac{\Lambda/\psi}{m^2 + \tau m'(\tau(m)) + \Lambda\left((1-m)m - \frac{m}{\psi}\right)}.$$

For (ii) a sufficient condition is $\psi \geq 1$ and $\tau(m) m'(\tau)$ weakly increasing in $\tau$: then we have

$$m^2 + \tau_1 m'(\tau_1) + \Lambda\left((1-m)m - \frac{m}{\psi}\right) = m^2(1-\Lambda) + \tau m' + m\Lambda\left(1 - \frac{1}{\psi}\right)$$

increasing in $m$.

For (iii), the problem is

$$\max_m u(c(p+m\tau)) - (p+\tau)c(p+m\tau) - g(m)$$

$$g'(m) = (u'(c) - q)c'(q^s)\tau = (q^s - q)c'(q^s)\tau = -c'(q^s)(1-m)\tau^2.$$

In the isoelastic case,

$$f(m,\tau) = \psi(1+m\tau)^{-\psi-1}(1-m)\tau^2 - g'(m).$$

We have

$$f(m(\tau),\tau) = 0.$$

**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 9.1** *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

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

$$\tau' = \tau\sqrt{\frac{\psi y}{\kappa}}$$
$$\Lambda' = \Lambda\sqrt{\frac{y}{\kappa\psi}} = \frac{\Lambda y}{\kappa}\frac{\tau}{\tau'}.$$

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.□

**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 9.9** *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 $\tau$ 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

$$m(\tau) = 1_{\tau > \tau_*}, \qquad \tau_* = \frac{\kappa}{\sqrt{\psi}}. \tag{64}$$

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

**Proposition 9.10** *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 $\Lambda$ 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

$\Lambda > \sqrt{2} + 1$.□

## 9.3 Other extensions

### 9.3.1 Cross-Effects of Attention

We again normalize $p_i = 1$. 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 2.7, 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 (13), in the limit of small $\Lambda$, as

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

This is a generalization of Proposition 3.1, 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 9.11** (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 $\rho < 0$.[81] When $m_2$ falls, the optimal tax on good 2 increases by the effects in Proposition 3.1, and optimal taxes on substitute goods also increase.[82]

### 9.3.2 Tax instruments with differential saliences

We elaborate on a remark we made at the end of section 3.5. As an extreme example, consider again the basic Ramsey example outlined above, and assume that the two tax systems with salience $m$ and $m'$ can be used jointly. Consider the case where there is only one agent and only one (taxed) good. With $m' > m$, we get

$$0 = (\lambda - \gamma)\, c + \left[\lambda \tau + \gamma(\bar{\tau}^s - \bar{\tau})\right] m \boldsymbol{S}^r, \qquad 0 = (\lambda - \gamma)\, c + \left[\lambda \tau + \gamma(\bar{\tau}^s - \bar{\tau})\right] m' \boldsymbol{S}^r,$$

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

[82] 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 $m_2, \psi_2$), good 2, he wants to subsidize good 1 if it is a strong complement of good 2.

where $\bar{\tau}^s$ is the total perceived tax arising from the joint perception of the two tax instruments. This requires $\lambda = \gamma$ and with $\bar{\tau}^s = 0$. In other words, the solution is the first best. This is because a planner can replicate a lump sum tax by combining a tax $\tau$ with low salience $m$ and a tax $-\tau\frac{m}{m'}$ with high salience $m' > m$, generating tax revenues $\tau\frac{m'-m}{m'}$ per unit of consumption of the taxed good with no associated distortion. This is an extreme result, already derived by Goldin (2015). In general, with more than one agent and heterogeneities in the misperceptions of the two taxes, the first best might not be achievable.

### 9.3.3 A different budget adjustment rule

The specific formulation of misperception that we have used in this section assumes that the budget adjustments required when agents misperceive taxes are all absorbed by the consumption a good (good 0) with a constant marginal utility. This renders these adjustments relatively painless.

We now explore a variant which increases their costs. We assume that the budget adjustments are concentrated on a "shock absorber" good with a sharply decreasing marginal utility. This increases the distortionary costs of non-salient taxes and reduces optimal taxes in a way that we characterize precisely below. Of course, it is difficult to know a priori which good is the "shock absorber" good (or set of goods) – this is one more place where more empirical evidence is needed to address a behavioral enrichment of the traditional model. It could be some luxury goods (e.g. some restaurant meals), or perhaps more pessimistically investments that can be postponed, e.g. health investments. Our purpose here is only to show how this possibility matters for the results.

When perceived prices $q_j^s$ are different from the true prices $q_j$, some adjustment is needed for the budget constraint. Let us study a different rule, where a certain good $n$ ("the last good", imagining a temporal order) bears the brunt of the budget adjustment (it's a "shock absorber"). This leads to

$$c^{i,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) = c^{i,r}\left(\boldsymbol{q}^s, w\right) \text{ for } i \neq n \tag{65}$$

$$c^{n,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) = \frac{1}{q_n}\left(w - \sum_{i \neq n} q_i c^{i,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right)\right). \tag{66}$$

This is: for all goods but the last one, the consumer only pays attention to perceived prices. Only for the last one does she see the budget constraint.[83] We shall see in the next proposition that we can also write

$$c^{n,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) = c^{n,r}\left(\boldsymbol{q}^s, w\right) - \frac{1}{q_n}\left(\boldsymbol{q} - \boldsymbol{q}^s\right) \cdot \boldsymbol{c}^r\left(\boldsymbol{q}^s, w\right), \tag{67}$$

i.e. actual consumption of good $n$ is planned consumption $c^{n,r}\left(\boldsymbol{q}^s, w\right)$ minus the adjustment for the surprise$\left(\boldsymbol{q} - \boldsymbol{q}^s\right) \cdot \boldsymbol{c}^r\left(\boldsymbol{q}^s, w\right)$ in the actual cost of the goods $i < n$ that have been purchased before

[83] Chetty et al. (2009) consider such a rule in a 2-good context. Gabaix (2017b) considers such a rule when doing dynamic programming, and the last good is "next period wealth".

good $n$.

For completeness, we record the Slutsky matrix properties of that rules. (Here we consider the income-compensated matrix $\boldsymbol{S}^C$).

**Proposition 9.12** (With the "last good adjusting for the budget" rule) *Consider the model above, with attention $m_j$ to price $j$. Evaluating at $\boldsymbol{q}^s = \boldsymbol{q}$, the marginal propensity to consume out of wealth isn't changed:*

$$\partial_w c_i^s \left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) = \partial_w c_i^r \left(\boldsymbol{q}, w\right). \tag{68}$$

*However, the Slutsky matrix $S_{ij}^s$ is changed as follows:*

$$S_{ij}^s = S_{ij}^r m_j + \left(\partial_w c_i^r - \frac{1}{q_n} 1_{i=n}\right)\left(1 - m_j\right) c^j, \tag{69}$$

*where $S_{ij}^r$ is the rational Slutsky matrix.*

**Proof**. The term $\partial_w c_i^s$ is trivial, as it's evaluate at $\boldsymbol{q}^s = \boldsymbol{q}$. We move on to the $S_{ij}$. First, take $i \neq n$. Then, $c^{i,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) = c^{i,r}\left(\boldsymbol{q}^s, w\right)$, hence:

$$\begin{aligned} S_{ij}^s &= \partial_{q_j} c^{i,r}\left(\boldsymbol{q}^s, w\right) + c_w^i c^j \\ &= c_{q_j}^{i,r}\left(\boldsymbol{q}, w\right) m_j + c_w^i c^j = \left(S_{ij}^r - c_w^i c^j\right) m_j + c_w^i c^j \\ &= S_{ij}^r + c_w^i c^j \left(1 - m_j\right), \end{aligned}$$

which gives the announced result.

For good $n$, we rewrite:

$$\begin{aligned} q_n c^{n,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) &= w - \sum_{i \neq n} q_i c^{i,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) \\ &= \boldsymbol{q}^s \cdot \boldsymbol{c}^r\left(\boldsymbol{q}^s, w\right) - \left(\boldsymbol{q} \cdot \boldsymbol{c}^r\left(\boldsymbol{q}^s, w\right) - q_n c^{n,r}\left(\boldsymbol{q}^s, w\right)\right) \\ &= \left(\boldsymbol{q}^s - \boldsymbol{q}\right) \cdot \boldsymbol{c}^r\left(\boldsymbol{q}^s, w\right) + q_n c^{n,r}\left(\boldsymbol{q}^s, w\right), \end{aligned}$$

i.e. another useful expression:

$$c^{n,s}\left(\boldsymbol{q}, \boldsymbol{q}^s, w\right) = c^{n,r}\left(\boldsymbol{q}^s, w\right) + \frac{1}{q_n}\left(\boldsymbol{q}^s - \boldsymbol{q}\right) \cdot \boldsymbol{c}^r\left(\boldsymbol{q}^s, w\right).$$

Its interpretation is that the consumption of the last good is the planned consumption (the first term, $c^{n,r}\left(\boldsymbol{q}^s, w\right)$), plus an adjustment for the "surprise" difference between planned and actual expenditure (the last term).

Now, differentiate w.r.t. $q_j$:

$$S_{nj} = \left(\partial_{q_j} c^{n,r}\left(\boldsymbol{q}^s, w\right) + c_w^n c^j\right) + \partial_{q_j}\left(\frac{1}{q_n}\left(\boldsymbol{q}^s - \boldsymbol{q}\right) \cdot \boldsymbol{c}^r\left(\boldsymbol{q}^s, w\right)\right).$$

By the earlier calculation of $S_{ij}$, the first term is $S^r_{ij} + c^i_w c^j (1 - m_j)$ with $i = n$, by the earlier result, and the last term is (as we evaluate at $\boldsymbol{q}^s = \boldsymbol{q}$)

$$\partial_{q_j} \left( \frac{1}{q_n} \sum_i (q_i^s - q_i) c^i (\boldsymbol{q}^s, w) \right) = \frac{1}{q_n} (m_j - 1) c^j (\boldsymbol{q}^s, w).$$

This gives the announced result.□

**A simple particular case** We next present a simple particular case. Utility is separable, $u(\mathbf{c}) = \sum_{i=0}^n u_i(c_i)$ with $u_0'(c_0) = 1$, $u_i'(c_i) = c_i^{-1/\psi_i}$ for $i = 1, ..., n-1$ and the "shock absorber" good $n$ has constant marginal utility of $u_n'(c_n) = 1 - \nu < 1$ if $c_n \geq 1$ and $1 + \mu > 1$ if $c_n < 1$.[84] We call $\mu > 0$ the marginal distortionary cost of budget adjustment. Goods 0 and $n$ cost \$1, and they are untaxed.

The agent chooses his consumption of goods $c_0, ..., c_{n-1}$ based on the perceived prices $q_i^s = 1 + m_i \tau_i$ and the rest of his money is spent on the last good. Specifically, the demands are a follows. For goods $i = 1, ..., n-1$, $c_i = (q_i^s)^{-\psi_i}$ (as the consumer solves $u'(c_i) = q_i^s$). The demand for good 0 is $c_0 = w - \sum_{i=1}^{n-1} (q_i^s)^{1-\psi_i} - 1$, as the consumer plans to consume $c_i = (q_i^s)^{-\psi_i}$ for all good $i = 1, ..., n-1$, and 1 of good $n$. Once goods 0 through $n-1$ have been purchased, the remaining disposable income for good $n$ is $c_n = w - \sum_{i=0}^{n-1} q_i c_i$.

Then (as derived shortly) he optimal tax on good $i < n$ is as in (17), replacing $\Lambda$ by

$$\Lambda_i = \frac{\Lambda - (1 - \Lambda)(1 - m_i)\mu}{1 - (1 - \Lambda)(1 - m_i)\mu}. \tag{70}$$

A direct consequence is that the optimal tax $\tau_i$ is lower than in the baseline case and is decreasing in $\mu$, particularly for less salient taxes with a small $m_i$. Indeed, the measure of the social marginal cost of public funds $\Lambda_i$ is decreasing in the marginal distortionary cost of budget adjustment $\mu$ (recall $\Lambda < 1$), coincides with its baseline value of $\Lambda$ when $\mu = 0$, and is lower than $\Lambda$ for all $\mu > 0$. Furthermore $\mu$ enters the formula through the $\mu(1 - m_i)$ so that these effects are particularly pronounced when attention $m_i$ is low.

We next proceed in greater detail. We take a particular case, which is particularly tractable. There are $n - 2$ goods, and good $n$ is the "shock absorber" good. The price of goods 0 and $n$ is normalized to 1. There's no tax on goods 0 and $n$, for simplicity.

Utility is:

$$u(c_0, ..., c_n) = c_0 + \sum_{i=1}^n u^i (c_i).$$

[84] The level of $\nu$ is unimportant provided it is between 0 and 1.

Good 0 has marginal utility of 1, which absorbs income effects, so $v_w = 1$. Hence, $\boldsymbol{\tau}^b = \boldsymbol{q} - \frac{u_{\boldsymbol{c}}}{v_w}$ is:

$$\begin{aligned}
\tau_0^b &= 0 \\
\tau_i^b &= q_i - q_i^s \text{ for } i = 1, ..., n-1 \\
\tau_n^b &= 1 - u_n'(c_n)
\end{aligned}$$

for $c_i = c_i^r(q_i^s)$ for $1 \leq i < n$ and $c_n = c_n^r + \sum_{i<n}(q_i^s - q_i)\, c_i$ (from (67)).

**Derivation of the optimal tax and $\Lambda_i = \frac{\Lambda-(1-\Lambda)(1-m_i)\mu}{1-(1-\Lambda)(1-m_i)\mu}$** We first provide an intuitive proof. We again normalize $p_i = 1$ and use normalized $\lambda = \frac{\lambda}{\gamma}$ to denote the relative benefit of raising revenue compared to welfare weight. The distortion on good $n$ is, from (67)

$$c_n - c_n^* = -\sum_{i<n}(1 - m_i)\,\tau_i c_i,$$

and the distortion on good 0 is $-(c_n - c_n^*)$. Hence in the objective function we have

$$\begin{aligned}
L &= W + \sum_{i<n}\lambda\tau_i c_i - \mu\sum_{i<n}(1 - m_i)\,\tau_i c_i \\
&= W + \sum_{i<n}(\lambda - \mu(1 - m_i))\,\tau_i c_i,
\end{aligned}$$

where $W$ = utility distortion all goods except 0 and $n$. Hence, we just replace $\lambda$ by $\lambda' = \lambda - \mu(1 - m_i)$.

Remember that we write $\lambda = \frac{1}{1-\Lambda}$. Hence, this corresponds to

$$\Lambda' = 1 - \frac{1}{\lambda'} = 1 - \frac{1}{\frac{1}{1-\Lambda} - (1 - m_i)\,\mu} = 1 - \frac{1-\Lambda}{1 - (1-\Lambda)(1 - m_i)\,\mu} = \frac{\Lambda - (1-\Lambda)(1 - m_i)\,\mu}{1 - (1-\Lambda)(1 - m_i)\,\mu}.$$

We also provide a more computational proof, which we found also instructive. Take an $i = 1, ..., n-1$. We have $\tau_i^b = (1 - m_i)\,\tau_i$ and $\tau_n^b = 1 - u'(c_n) = -\mu$. We have $S_{ii} = -\frac{\psi_i c_i}{q_i^s} m_i$, while

$$S_{ni} = -(1 - m_i)\,c_i\left(1 - \psi_i\frac{\tau_i}{q_i^s}m_i\right).$$

Plugging this into the general Ramsey optimal tax formula (8) gives:

$$\begin{aligned}
0 &= (\lambda - \gamma)\,c_i + \lambda(\boldsymbol{\tau} - \widetilde{\boldsymbol{\tau}}^b)\cdot \boldsymbol{S}_i^C \\
&= (\lambda - 1)\,c_i - \lambda\left(\tau_i - \frac{1}{\lambda}(1 - m_i)\,\tau_i\right)\frac{\psi_i c_i}{q_i^s}m_i - \lambda\left(0 + \frac{\mu}{\lambda}\right)(1 - m_i)\,c_i\left(1 - \psi_i\frac{\tau_i}{q_i^s}m_i\right) \\
&= (\lambda - 1 - \mu(1 - m_i))\,c_i - \frac{\psi_i c_i}{q_i^s}m_i\tau_i\left(\lambda - (1 - m_i) - \mu(1 - m_i)\right),
\end{aligned}$$

which is the expression with $\mu = 0$, if we replace $\lambda$ by $\lambda' = \lambda - (1 - m_i)\,\mu$.

**Analysis of small taxes** We analyze the case of small taxes. Compared to the first best, distortions are:

$$\begin{aligned} c_i - c_i^* &= -\psi_i c_i^* m_i \tau_i \\ c_n - c_n^* &= -\sum_{i=1}^{n-1} c_i \left(1 - m_i\right) \tau_i, \end{aligned}$$

Then losses are:

$$\begin{aligned} L &= -\sum_{i=1}^{n-1} \left( \frac{1}{2} \psi_i m_i^2 c_i^* \tau_i^2 + \mu c_i^* \left(1 - m_i\right) \tau_i \right) + \lambda \sum_i \tau_i y_i \\ &= -\sum_{i=1}^{n-1} \left( \frac{1}{2} \psi_i m_i^2 c_i^* \tau_i^2 \right) + \sum_i \left(\lambda - \mu \left(1 - m_i\right)\right) \tau_i c_i, \end{aligned}$$

so, optimal tax on good $i$ is 0 iff $\mu \left(1 - m_i\right) < \lambda$.

We can also investigate the case in which the utility associated with good $n$, the shock absorber, is isoelastic. In this case, a utility loss equal to $L^D$ is such that:

$$-2L^D = \sum_i \frac{1}{\psi_i c_i^{*2}} \left(c_i - c_i^*\right)^2.$$

Hence we have the following generalization of the objective function in the simple Ramsey case with small taxes (we normalized prices to $p_i = 1$, i.e. $c_i^* = 1$, for all $1 \leq i < n$)

$$L = -\frac{1}{2} \sum_{i=1}^{n-1} \tau_i^2 \psi_i m_i^2 c_i^* - \frac{1}{2} \frac{1}{\psi_n c_n^*} \left( \sum_{i=1}^{n-1} c_i^* \left(1 - m_i\right) \tau_i \right)^2 + \lambda \sum_i \tau_i c_i^*. \tag{71}$$

In particular, now the distortion is not just $\psi_i m_i^2$ as before, but there is another term, multiplied by $\frac{1}{\psi_n}$. Hence, attention is beneficial only if if risk aversion ($\frac{1}{\psi_n}$) for the shock absorber good is small enough (in the baseline model it is 0). The optimal tax is

$$\tau_i = \frac{\Lambda_i}{m_i^2 \psi_i},$$

with

$$\Lambda_i = \Lambda - \mu \left(1 - m_i\right),$$

and

$$\mu = \frac{1}{\psi_n c_n^*} \sum_{i=1}^{n-1} c_i \left(1 - m_i\right) \tau_i,$$

which is the marginal distortion on good $n$.

**Impact on Pigouvian taxes** We revisit our simple model of Section 3.2, with an externality on good 1. We have $\lambda = 1$, so that the government's objective function is:

$$L = U(c_1) - (p + \xi) c_1 + u_2 (c_2^* - (1 - m) c_1 \tau) + (1 - m) c_1 \tau.$$

i.e. utility from good 1, utility from good 2 (which absorbs the shock $(1 - m) c_1 \tau$), and consumption of good 0 is increased by the lump-sum rebate, which accounts for the last term. The consumer chooses $c_1$ according to $U'(c_1) = p + m\tau$.

We take utility $U(c_1) = Qc_1 - \frac{c_1^2}{2\Psi}$, so that demand is $c_1 = \Psi(Q - p - m\tau)$ . We keep $u_2'(c_2) = 1 + \mu$. We have:

$$\begin{aligned} L'(\tau) &= [p + m\tau - (p + \xi)](-\Psi m) + [-(1 - m)(1 + \mu) + (1 - m)] \frac{d}{d\tau}(c_1 \tau) \\ &= -(m\tau - \xi)\Psi m - (1 - m)\mu(c_1 - \Psi m\tau) \\ &= -(m\tau - \xi)\Psi m - (1 - m)\mu(\Psi(Q - p - 2m\tau)), \end{aligned}$$

which leads to:

$$\begin{aligned} \tau &= \frac{\frac{\xi}{m} - \mu\left(\frac{1-m}{m}\right)(Q - p)}{1 - 2\mu\left(\frac{1-m}{m}\right)} \\ &= \frac{\frac{\xi}{m} - \mu\left(\frac{1-m}{m}\right)\frac{c_1^0}{\Psi}}{1 - 2\mu\left(\frac{1-m}{m}\right)}, \end{aligned}$$

where $c_1^0 = \Psi(Q - p)$ is the consumption of good 1 if there is no tax.

Hence, the government doesn't tax the good if: $\xi\Psi < \mu(1 - m) c_1^0$, i.e. if the externality is too small.

### 9.3.4 Quadratic losses from imperfect tax instruments

We introduce the Lagrangian that allows for agent-specific lump-sum transfers $w^h$ and taxes $\tau^h, \tau^{s,h}$ (we normalize $p_i = 1$)

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

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

$$g\left(\left\{\tau^{s,h}\right\}\right) = \max_{\left\{w^h\right\}} L\left(\left\{\tau^{s,h}\right\}, \left\{\tau^{s,h}\right\}, \left\{w^h\right\}\right), \tag{72}$$

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 9.13** *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*

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

**Proof**. 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 $w^{**}\left(y\right)$ the optimal redistribution given a tax system $y$. So, $w^{*} = w^{**}\left(y^{*}\right)$.

$$\begin{aligned}
L^{\text{tot}} &= L\left(w,y\right) - L\left(w^{*},y^{*}\right)\\
&= \left[L\left(w,y\right) - L\left(w^{**}\left(y\right),y\right)\right] + \left[L\left(w^{**}\left(y\right),y\right) - L\left(w^{**}\left(y^{*}\right),y^{*}\right)\right]\\
&= \frac{1}{2}\left(w - w^{**}\left(y\right)\right)\cdot L_{ww}\left(w^{**}\left(y\right),y\right)\cdot\left(w - w^{**}\left(y\right)\right) + \frac{1}{2}\left(y-y^{*}\right)g_{yy}\left(y-y^{*}\right) \text{ by Lemma 11.3}\\
&= L^{\text{distribution}} + L^{\text{distortion}}\\
L^{\text{distribution}} &= \frac{1}{2}\left(w - w^{**}\left(y\right)\right)\cdot L_{ww}\left(w^{**}\left(y\right),y\right)\cdot\left(w - w^{**}\left(y\right)\right)\\
L^{\text{distortion}} &= \frac{1}{2}\left(y-y^{*}\right)g_{yy}\left(y-y^{*}\right).
\end{aligned}$$

*Redistribution terms*

From Lemma 11.2, the expression of the loss involves $L_{w^{h}}\left(w,\tau\right) = \gamma^{\xi,h} - \lambda$, the social marginal

utility. Applying that Lemma 11.2 gives a loss:

$$L^{\text{distribution}} = \frac{1}{2}\sum_{h,h'}\left(\gamma^{\xi,h}-\overline{\gamma}\right)\left(L_{ww}\left(w,\tau\right)^{-1}\right)_{h,h'}\left(\gamma^{\xi,h}-\overline{\gamma}\right). \tag{73}$$

*Tax distortion terms*

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

$$g\left(\tau^s\right) = \max_{w_1,...,w_n} L\left(w,\tau^s\right) = L\left(w^*\left(\tau^s\right),\tau^s\right)$$
$$g_{\tau^s\tau^s} = L_{\tau^s\tau^s} - L_{\tau^s w}L_{ww}^{-1}L_{\tau^s w} \text{ by Lemma 11.3.}$$

□

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

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

**Lemma 9.2** *We have*

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

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

$$g_{\tau^{s,h}\tau^{s,h'}} = \lambda S^{r,h}1_{h=h'} + S^{r,h}\xi_{c^hc^{h'}}S^{r,h'}. \tag{74}$$

**Proof**. We observe that a tax $\tau^{s,h}$ 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

$$\frac{d\xi\left(\left\{w^h\right\},\left\{\tau^{s,h}\right\}\right)}{d\tau^{s,h}} = \frac{\xi_{c^h}c^h_{\tau^{s,h}}}{1-\sum_{h'}\xi_{c^{h'}}c^{h'}_{\xi}}. \tag{75}$$

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

$$\frac{d\xi}{dw^h} = \frac{\xi_{c^h}c^h_{w^h}}{1-\sum_{h'}\xi_{c^{h'}}c^{h'}_{\xi}}. \tag{76}$$

We note that the FOC of (72) in $w^h$ is

$$\begin{aligned}
0 &= v^h_{w^h} + \lambda\left(\tau^{s,h}c^{r,h}_{w^h} - 1\right) + \frac{d\xi}{dw^h}\sum_{h'} v^{h'}_{\xi} + \lambda\tau^{s,h'}\cdot\overline{\overline{c}}^{r,h'}_{\xi} \\
&= v^h_{w^h} + \lambda\left(\tau^{s,h}c^{r,h}_{w^h} - 1\right) + \frac{\xi_{c^h}c^h_{w^h}}{1-\sum_{h'}\xi_{c^{h'}}c^{h'}_{\xi}}\sum_{h'} v^{h'}_{\xi} + \lambda\tau^{s,h'}\cdot\overline{\overline{c}}^{r,h'}_{\xi} \text{ by (76)} \\
&= v^h_{w^h} + \lambda\left(\tau^{s,h}c^{r,h}_{w^h} - 1\right) + \xi_{c^h}c^h_{w^h}\Xi \\
&= v^h_{w^h} + \lambda\left(\tau^{s,h}c^{r,h}_{w^h} - 1\right) - \lambda\tau^{\xi,h}c^h_{w^h} \\
&= \gamma^{\xi,h} - \lambda.
\end{aligned}$$

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

$$\begin{aligned}
g\left(\left\{\tau^{s,h}\right\}\right) &= \max_{\{w^h\}}\sum_h v^r\left(p+\tau^{s,h},w^h,\xi\right) + \lambda\sum_h\left[\tau^{s,h}\cdot\overline{\overline{c}}^r\left(p+\tau^{s,h},w^h,\xi\right) - w^h\right] \\
&= \max_{\{w^h\}}\sum_h v^r\left(p+\tau^{s,h},w^h,\xi\right) + \lambda\sum_h\left[p\cdot\overline{\overline{c}}^r\left(p+\tau^{s,h},w^h,\xi\right)\right]
\end{aligned}$$

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

We take the derivatives (72):

$$\begin{aligned}
g_{\tau^{s,h}}\left(\left\{\tau^{s,h'}\right\}\right) &= \left(\lambda - v^h_w\right)c^{r,h} + \lambda\tau^{s,h}c^{r,h}_p + \frac{d\xi}{d\tau^{s,h}}\sum_{h'}\left[v^{h'}_{\xi} + \lambda\tau^{s,h'}\cdot\overline{\overline{c}}^r_{\xi}\right] \\
&= \left(\lambda - v^h_w\right)c^{r,h} + \lambda\tau^{s,h}c^{r,h}_p + \frac{\xi_{c^h}c^h_{\tau^{s,h}}}{1-\sum_{h'}\xi_{c^{h'}}c^{h'}_{\xi}}\sum_{h'} v^{h'}_{\xi} + \lambda\tau^{s,h'}\cdot\overline{\overline{c}}^r_{\xi} \text{ by (75)} \\
&= \left(\lambda - v^h_w\right)c^{r,h} + \lambda\tau^{s,h}c^{r,h}_p + \xi_{c^h}c^h_{\tau^{s,h}}\Xi \\
&= \left(\lambda - v^h_w\right)c^{r,h} + \lambda\tau^{s,h}c^{r,h}_p - \lambda\tau^{\xi,h}c^h_p \\
&= \left(\lambda - v^h_w\right)c^{r,h} + \lambda\left(\tau^{s,h} - \tau^{\xi,h}\right)c^{r,h}_p \\
&= \left(\lambda - v^h_w\right)c^{r,h} + \lambda\left(\tau^{s,h} - \tau^{\xi,h}\right)\left[S^{r,h} - c^r_w c^h\right] \\
&= \left[\lambda - v^h_w - \lambda\left(\tau^{s,h} - \tau^{\xi,h}\right)c_w\right]c^h + \lambda\left(\tau^{s,h} - \tau^{\xi,h}\right)S^{r,h} \\
&= \lambda\left(\tau^{s,h} - \tau^{\xi,h}\right)S^{r,h}.
\end{aligned} \tag{77}$$

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

$$g_{\tau^{s,h}\tau^{s,h'}} = \lambda S^{r,h}\left(1_{h=h'} - \frac{d\tau^{\xi,h}\left(\left\{\tau^{s,h''}\right\}\right)}{d\tau^{s,h'}}\right). \tag{78}$$

*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 $\xi$, and $\Xi = \frac{\sum_h \left[\beta^h \frac{v_\xi^h}{v_w^h} + \lambda\tau\cdot\bar{\bar{c}}_\xi^h\right]}{1-\sum_h \xi_{c^h} c_\xi^h} = 1$, and $\tau^{\xi,h} = -\frac{1}{\lambda}\xi_{c^h}$.

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

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

so

$$g_{\tau^{s,h}\tau^{s,h\prime}} = \lambda S^{r,h} 1_{h=h'} + H S^{r,h}\xi_{c^h c^{h'}} S^{r,h'}. \tag{79}$$

That should generalize to additive externality: $u(c,\xi) = u(c) + \frac{1}{H}\xi$. Then $\Xi = \frac{\sum_h \left[\beta^h \frac{v_\xi^h}{v_w^h} + \lambda\tau\cdot\bar{\bar{c}}_\xi^h\right]}{1-\sum_h \xi_{c^h} c_\xi^h} = 1$. And $\tau^{\xi,h} = -\frac{1}{\lambda}\xi_{c^h}\left(\left\{\tau^{s,h}\right\}\right)$. When $\left\{\tau^{-h'}\right\}$ are held constant, varying $\tau^{h\prime}$ changing $c^{h'}$ and $\xi$, 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 $h'' \neq h'$,

$$\begin{aligned}
\frac{d\tau^{\xi,h}}{d\tau^{s,h'}} &= -\frac{1}{\lambda}\xi_{c^h c^{h'}}\frac{dc^{h'}}{d\tau^{h'}} = -\frac{1}{\lambda}\xi_{c^h c^{h'}} S^{r,h'} + \sum_{h''} -\frac{1}{\lambda}\xi_{c^h c^{h''}}\frac{\partial c^{h''}}{dw^{h''}}\frac{dw^{h''}}{d\tau^{h'}} \\
&= -\frac{1}{\lambda}\xi_{c^h c^{h'}}\left(S^{r,h'} - c_{w^{h'}}^{h'} c + c_{w^{h'}}^{h'}\frac{dw^{h'}}{d\tau^{h'}}\right) \\
&= -\frac{1}{\lambda}\xi_{c^h c^{h'}}\left[S^{r,h'} - c_{w^{h'}}^{h'}\left(c - \frac{dw^{h'}}{d\tau^{h'}}\right)\right].
\end{aligned}$$

as $\gamma^{\xi,h} = v_w^h + \lambda\left(\tau^{s,h} - \tau^{\xi,h}\right)\cdot c_w^h = \lambda$ implies $dv_w^h + \lambda\left(d\tau^{s,h}\right)\cdot c_w = 0$.□

# 10 The Nonlinear Income Tax Problem

Here are the notations we shall use.

$g(z)$ :social welfare weight

$h(z)$ (resp. $h^*(z)$): density (resp. virtual density) of earnings $z$

$H(z)$: cumulative distribution function of earnings

$n$ :agent's wage, also the index of his type

$q(z) = R'(z)$: marginal retention rate, locally perceived

$\boldsymbol{Q} = (q(z))_{z\geq 0}$: vector of marginal retention rates

$r_0$: tax rebate at 0 income

$r(z)$ :virtual income

$R(z) = z - T(z)$: retained earnings

$T(z)$: tax given earnings $z$

$z$: pre-tax earnings

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

$\eta$: income elasticity of earnings

$\pi$ :Pareto exponent of the earnings distribution

$\zeta^c$: compensated elasticity of earnings

$\zeta^c_{Q_{z*}}(z)$: compensated elasticity of earnings when the tax rate at $z^*$ changes.

$\zeta^u$ : uncompensated elasticity of earnings

## 10.1 Setup

**Agent's behavior** There is a continuum of agents indexed by skill $n$ with density $f(n)$ (we use $n$ rather than $h$, the conventional index in that literature). Agent $n$ has a utility function $u^n(c,z)$, where $c$ is his one-dimensional consumption, $z$ is his pre-tax income, and $u_z \leq 0$.[85]

The total income tax for income $z$ is $T(z)$, so that disposable income is $R(z) = z - T(z)$. We call $q(z) = R'(z) = 1 - T'(z)$ the local marginal "retention rate", $\boldsymbol{Q} = (q(z))_{z\geq 0}$ the ambient vector of all marginal retention rates, and $r_0 = R(0)$ the transfer given by the government to an agent earning zero income. We define the "virtual income" to be $r(z) = R(z) - zq(z)$. Equivalently $R(z) = q(z)z + r(z)$, so that $q(z)$ is the local slope of the budget constraint, and $r(z)$ its intercept.

We use a general behavioral model in a similar spirit to Section 2. The primitive is the income function $z^n(q, \boldsymbol{Q}, r_0, r)$, which depends on the local marginal retention rate $q$, the ambient vector of all marginal retention rates $\boldsymbol{Q}$, $r_0 = R(0)$ the transfer given by the government to an agent earning zero income, and the virtual income $r$. In the traditional model without behavioral biases we have $z^n(q, \boldsymbol{Q}, r_0, r) = \arg\max_z u^n(qz + r, z)$, so that $z^n$ does not depend on $\boldsymbol{Q}$ and $r_0$. With behavioral biases, this is no longer true in general. The income function is associated with the indirect utility function $v^n(q, \boldsymbol{Q}, r_0, r) = u^n(qz + r, z)_{|z=z^n(q,\boldsymbol{Q},r_0,r)}$. The earnings $z(n)$ of agent $n$ facing retention schedule $R(z)$ is then the solution of the fixed point problem $z = z^n(q(z), \boldsymbol{Q}, r_0, r(z))$. His consumption is $c(n) = R(z(n))$ and his utility is $v(n) = u^n(c(n), z(n))$.

**Planning problem** The objective of the planner is to design the tax schedule $T(z)$ in order to maximize the following objective function

$$\int_0^\infty W(v(n)) f(n)\, dn + \lambda \int_0^\infty (z(n) - c(n)) f(n)\, dn.$$

Like Saez (2001), we normalize $\lambda = 1$. We call $g(n) = W'(v(n)) v^n_r(q(z(n)), \boldsymbol{Q}, r_0, r(z(n)))$ the marginal utility of income. This is the analogue of $\beta^h$ in the Ramsey problem of Section 2, and

[85] If the agent's pre-tax wage is $n$, $L$ is his labor supply, and utility is $U^n(c, L)$, then $u^n(c, z) = U\left(c, \frac{z}{n}\right)$. Note that this assumes that the wage is constant (normalized to one). We discuss the impact of relaxing this assumption in the NBER working paper version of this paper.

we identify agents with their income level $z(n)$ instead of their skill $n$. Most of the time, we leave implicit the dependence of $n(z)$ on $z$ to avoid cluttering the notations. We now derive a behavioral version of the optimal tax formula in Saez (2001).

## 10.2 Saez Income Tax Formula with Behavioral Agents

### 10.2.1 Elasticity Concepts

Recall that the marginal retention rate is $q(z) = 1 - T'(z)$. Given an income function $z(q, \boldsymbol{Q}, r_0, r)$, we introduce the following definitions. We define the income elasticity of earnings

$$\eta = q z_r(q, \boldsymbol{Q}, r_0, r).$$

We also define the uncompensated elasticity of labor (or earnings) supply with respect to the actual marginal retention rate

$$\zeta^u = \frac{q}{z} z_q(q, \boldsymbol{Q}, r_0, r).$$

Finally, we define the compensated elasticity of labor supply with respect to the actual marginal retention rate

$$\zeta^c = \zeta^u - \eta.$$

We also introduce two other elasticities, which are zero in the traditional model without behavioral biases. We define the compensated elasticity of labor supply at $z$ with respect to the marginal retention rate $q(z^*)$ at a point $z^*$ different from $z$:

$$\zeta^c_{Q_{z^*}} = \frac{q}{z} z_{Q_{z^*}}(q, \boldsymbol{Q}, r_0, r).$$

We also define the earnings sensitivity to the lump-sum rebate at zero income[86]

$$\zeta^c_{r_0} = \frac{q}{z} z_{r_0}(q, \boldsymbol{Q}, r_0, r).$$

We shall call $\zeta^c_{Q_{z^*}}$ a "behavioral cross-influence" of the marginal tax rate at $z^*$ on the decision of an agent earning $z$. In the traditional model with no behavioral biases, $\zeta^c_{Q_{z^*}} = \zeta^c_{r_0} = 0$, not so with behavioral agents.[87][88]

All these elasticities a priori depend on the agent earnings $z$. As mentioned above, we leave this dependence implicit most of the time.

[86] Formulas would be cleaner without the multiplication by $q$ in those elasticities, but here we follow the public economics tradition.

[87] For instance, in the misperception model, in general, the marginal tax rate at $z^*$ affects the default tax rate and therefore the perceived tax rate at earnings $z$.

[88] In the language of Section 2.1, we use income-compensation based notion of elasticity, $\boldsymbol{S}^C$, rather than the utility–compensation based notion $\boldsymbol{S}^H$.

Just like in the Ramsey model, we define the "behavioral wedge"

$$\tau^b(q, \boldsymbol{Q}, r_0, r) = -\frac{qu_c(c,z) + u_z(c,z)}{v_r(q, \boldsymbol{Q}, r_0, r)}_{|z=z(q,\boldsymbol{Q},r_0,r),c=qz+r}.$$

We also define the renormalized behavioral wedge

$$\widetilde{\tau}^b(z) = g(z)\,\tau^b(z).$$

In the traditional model with no behavioral biases, we have $\tau^b(q, \boldsymbol{Q}, r_0, r) = \widetilde{\tau}^b(z) = 0$. But this is no longer true with behavioral agents.

We have the following behavioral version of Roy's identity (proven in Section 11.3.2):

$$\frac{v_q}{v_w} = z - \frac{\tau^b z}{q}\zeta^c, \qquad \frac{v_{Q_{z^*}}}{v_w} = -\frac{\tau^b z}{q}\zeta^c_{Q_{z^*}}. \tag{80}$$

As in Section 2, the general model can be particularized to a decision vs. experienced utility model, or to a misperception model.

**Misperception model** The agent may misperceive the tax schedule, including her marginal tax rate. We call $T^{s,n}(q, \boldsymbol{Q}, r_0)(z)$ the perceived tax schedule, $R^{s,n}(z) = z - T^{s,n}(q, \boldsymbol{Q}, r_0)(z)$ the perceived retention schedule, and $q^{s,n}(q, \boldsymbol{Q}, r_0)(z) = \frac{dR^{s,n}(q,\boldsymbol{Q},r_0)(z)}{dz}$ the perceived marginal retention rate. Faced with this tax schedule, the behavior of the agent can be represented by the following problem

$$\max_{c,z|R^{s,n}(\cdot)} u^n(c,z) \text{ s.t. } c = R(z). \tag{81}$$

This formulation implies that the agent's choice $(c, z)$ satisfies $c = R(z)$ and

$$q^{s,n}(z)\,u^n_c(c,z) + u^n_z(c,z) = 0, \tag{82}$$

instead of the traditional condition $q(z)\,u^n_c(c,z) + u^n_z(c,z) = 0$. This means that the agent correctly perceives consumption and income $(c, z)$ but misperceives his marginal retention rate $q^{s,n}(z)$. Together with $c = R(z)$, this characterizes the behavior of the agent.[89]

Accordingly, we define $z^n(q, q^s, r)$ to be the solution of $q^{s,n}u^n_c(c,z) + u^n_z(c,z) = 0$ with $c = qz + r$.[90] The income $z(n)$ of agent $n$ is then the solution of the fixed point equation

$$z = z^n\left(q(z), q^{n,s}(q, \boldsymbol{Q}, r_0)(z), r(z)\right),$$

[89] This is a sparse max problem with a non-linear budget constraint, which generalizes the sparse max with a linear budget constraint we analyzed in section 3.1. The true constraint is $c = R(z)$, but the perceived constraint is $c = R^{s,n}(q, \boldsymbol{Q}, r_0)(z)$.

[90] If there are several solutions, we choose the one that yields the greatest utility.

his consumption is $c(n) = R(z(n))$ and his utility is $v(n) = u^n(c(n), z(n))$.

Summing up, in the misperception model, the primitives are a utility function $u$ and a perception function $q^s(q, \boldsymbol{Q}, r_0)(z)$. This yields an income function $z(q, q^s, r)$. The general function $z(q, \boldsymbol{Q}, r_0, r)$ is then $z(q(z'), \boldsymbol{Q}, r_0, r) = z(q(z'), q^s(q, \boldsymbol{Q}, r_0)(z'), r)$ for any earnings $z'$.

One concrete example of misperception is $q^{s,n}(q, \boldsymbol{Q}, r_0) = q^s(q, \boldsymbol{Q}, r_0)$ with

$$q^s(q, \boldsymbol{Q}, r_0)(z) = mq(z) + (1-m)[\alpha q^d(\boldsymbol{Q}) + (1-\alpha)\frac{r_0 + \int_0^z q(z')\,dz'}{z}],$$

where $m \in [0,1]$ is the attention to the true tax (hence retention) rate, $\frac{r_0+\int_0^z q(z')dz'}{z}$ is the average retention rate (as in Liebman and Zeckhauser (2004)), and $\alpha \in [0,1]$. The default perceived retention rate might be a weighted average of marginal rates, e.g. $q^d(\boldsymbol{Q}) = \int q(z)\,\omega(z)\,dz$ for some weights $\omega(z)$.

As in the Ramsey case, it is useful to express behavioral elasticities as a function of an agent without behavioral biases. Call $z^r(q^s, r') = \arg\max_z u(q^s z + r', z)$ the earnings of a rational agent facing marginal tax rate $q^s$ and extra non-labor income $r'$. Then, $z(q, q^s, r) = z^r(q^s, r')$ where $r'$ solves $r' + q^s z^r(q^s, r') = r + q z^r(q^s, r')$. We call $S^r(q^s, r') = \frac{\partial z^r}{\partial q^s}(q^s, r') - \frac{\partial z^r}{\partial r'}(q^s, r')\, z^r(q^s, r')$ the rational compensated sensitivity of labor supply (it is just a scalar). We also define $\zeta^{cr} = \frac{qS^r}{z}$ as the compensated elasticity of labor supply of the agent if he were rational.

We define $m_{zz} = q^s_q(q, \boldsymbol{Q}, r_0)(z)$ as the attention to the own marginal retention rate and $m_{zz^*} = q^s_{Q_{z^*}}(q, \boldsymbol{Q}, r_0)(z)$ as the marginal impact on the perceived marginal retention rate at $z$ of an increase in the marginal retention rate at $z^*$. Then, we have the following concrete values for the elasticities of the general model (the derivation is in Section 11.3.2):

$$\zeta^c = \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{cr} m_{zz}, \qquad \zeta^c_{Q_{z^*}} = \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{cr} m_{zz^*}, \tag{83}$$

$$\tau^b = \frac{\tau - \tau^s}{1 - \eta\frac{\tau-\tau^s}{q}}. \tag{84}$$

If the behavioral agent overestimates the tax rate ($\tau - \tau^s < 0$), the term $\tau^b$ is negative. Loosely, we can think of $\tau^b$ as indexing an "underperception" of the marginal tax rate. In the traditional model without behavioral biases, $m_{zz^*} = 1_{z=z^*}$, $\tau^s = \tau$ and $\tau^b = 0$.

**Decision vs. experienced utility model** In the decision vs. experienced utility model, behavior is represented by the maximization of a subjective decision utility $u^s(c, z)$ subject to the budget constraint $c = R(z)$. We then have $\zeta^c_{Q_{z^*}} = 0$, and $\zeta^c$ and $\eta$ are the elasticities associated with decision utility $u^s$. The behavioral wedge is

$$\tau^b = \frac{\frac{u_c}{u^s_c}u^s_z - u_z}{v_r}. \tag{85}$$

**Other useful concepts and notations** 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 introduce the virtual density $h^*(z) = \frac{q(z)}{q(z)-\zeta^c z R''(z)} h(z)$.

We define the social marginal utility of income

$$\gamma(z) = g(z) + \frac{\eta(z)}{1-T'(z)}\left[\widetilde{\tau}^b(z) + \left(T'(z) - \widetilde{\tau}^b(z)\right)\frac{h^*(z)}{h(z)}\right]. \tag{86}$$

This definition is the analogue of the corresponding definition in the Ramsey model. It is motivated by Lemma 11.5, which shows that, if the government transfers a lump-sum $\delta K$ to an agent previously earning $z$, the objective function of the government increases by $\delta L(z) = (\gamma(z) - 1)\,\delta K$. The social marginal utility of income $\gamma(z)$ reflects a direct effect $g(z)$ of that transfer to the agent's welfare, and an indirect effect on labor supply captured—to the leading order as the agent receives $\delta K$, his labor supply changes by $\frac{\eta(z)}{1-T'(z)}\delta K$, which impacts tax revenues by $\frac{\eta(z)}{1-T'(z)}T'(z)\,\delta K$ and welfare by $\frac{\eta(z)}{1-T'(z)}\widetilde{\tau}^b(z)\,\delta K$; the terms featuring $\frac{h^*(z)}{h(z)}$ (in practice often close to 1) capture the fact that the agent's marginal tax rate changes as the agent adjusts his labor supply, which impacts tax revenues and welfare because misoptimization.

### 10.2.2 Optimal Income Tax Formula

We next present the optimal income tax formula. Section 11.3.1 presents the intermediary steps used in the derivation of this formula.

**Proposition 10.1** *Optimal taxes satisfy the following formulas (for all $z^*$)*

$$\begin{aligned}\frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1-T'(z^*)} = &\frac{1}{\zeta^c(z^*)}\frac{1-H(z^*)}{z^*h^*(z^*)}\int_{z^*}^{\infty}(1-\gamma(z))\frac{h(z)}{1-H(z^*)}dz \\ &- \int_0^{\infty}\frac{\zeta^c_{Q_{z^*}}(z)}{\zeta^c(z^*)}\frac{T'(z) - \tilde{\tau}^b(z)}{1-T'(z)}\frac{zh^*(z)}{z^*h^*(z^*)}dz.\end{aligned} \tag{87}$$

*This formula can also be expressed as a modification of the Saez (2001) formula*

$$\begin{aligned}&\frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1-T'(z^*)} + \int_0^{\infty}\omega(z^*,z)\frac{T'(z) - \tilde{\tau}^b(z)}{1-T'(z)}dz \\ &= \frac{1}{\zeta^c(z^*)}\frac{1-H(z^*)}{z^*h^*(z^*)}\int_{z^*}^{\infty}e^{-\int_{z^*}^{z}\rho(s)ds}\left(1-g(z)-\eta\frac{\tilde{\tau}^b(z)}{1-T'(z)}\right)\frac{h(z)}{1-H(z^*)}dz,\end{aligned} \tag{88}$$

*where $\rho(z) = \frac{\eta(z)}{\zeta^c(z)}\frac{1}{z}$ and*

$$\omega(z^*,z) = \left(\frac{\zeta^c_{Q_{z^*}}(z)}{\zeta^c(z^*)} - \int_{z'=z^*}^{\infty}e^{-\int_{z^*}^{z'}\rho(s)ds}\rho(z')\frac{\zeta^c_{Q_{z'}}(z)}{\zeta^c(z^*)}dz'\right)\frac{zh^*(z)}{z^*h^*(z^*)}.$$

The first term $\frac{1}{\zeta^c(z^*)}\frac{1-H(z^*)}{z^*h^*(z^*)}\int_{z^*}^{\infty}(1-\gamma(z))\frac{h(z)}{1-H(z^*)}dz$ on the right-hand side of the optimal tax formula (87) is a simple reformulation of Saez's formula, using the concept of social marginal utility of income $\gamma(z)$ rather than the marginal social welfare weight $g(z)$ The link between the two is in equation (86)). The second term $-\frac{1}{z^*}\int_0^{\infty}\frac{\zeta^c_{Q_{z^*}}(z)}{\zeta^c(z^*)}\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}z\frac{h^*(z)}{h^*(z^*)}dz$ on the right-hand side is new and captures a misoptimization effect together with the term $\frac{-\tilde{\tau}^b(z^*)}{1-T'(z^*)}$ on the left-hand side.

The intuition is as follows. First, suppose for concreteness that $\zeta^c_{Q_{z^*}}(z)>0$, then increasing the marginal tax rate at $z^*$ leads the agents at another income $z$ to perceive higher taxes on average, which leads them to decrease their labor supply and reduces tax revenues. Ceteris paribus, this consideration pushes towards a lower tax rate, compared to the Saez optimal tax formula. Second, suppose for concreteness that $\tilde{\tau}^b(z)<0$, then increasing the marginal tax rate at $z^*$ further reduces welfare. This, again, pushes towards a lower tax rate.

The modified Saez formula (88) uses the concept of the social marginal welfare weight $g(z)$ rather than the social marginal utility of income $\gamma(z)$. It is easily obtained from formula (87) using equation (86). When there are no income effects so that $\eta=\rho(z)=0$, the optimal tax formula (87) and the modified Saez formula (88) are identical. They coincide with the traditional Saez formula when there are no behavioral biases so that $\zeta^c_{Q_{z^*}}(z)=\omega(z^*,z)=\tilde{\tau}^b(z)=0$. In this case, the left-hand side of (88) is simply $\frac{T'(z^*)}{1-T'(z^*)}$ so that the formula solves for the optimal marginal tax rate $T'(z^*)$ at $z^*$.

The formula is expressed in terms of endogenous objects or "sufficient statistics": social marginal welfare weights $g(z)$, elasticities of substitution $\zeta^c(z)$, income elasticities $\eta(z)$, and income distribution $h(z)$ and $h^*(z)$. With behavioral agents, there are two differences. First, there are two additional sufficient statistic, namely the behavioral wedge $\tilde{\tau}^b(z)$ and the behavioral cross-elasticities $\zeta^c_{Q_{z^*}}(z)$. Second, it is not possible to solve out the optimal marginal tax rate in closed form. Instead, the modified Saez formula (88) at different values of $z^*$ form a system of linear equations in the optimal marginal tax rates $T'(z)$ for all $z$. The formula simplifies greatly in the case where behavioral biases can be represented by a decision vs. experienced utility model. Indeed, we then have $\omega(z^*,z)=0$ and $\tilde{\tau}^b(z)=g(z)\frac{u_c\frac{u^s_z}{u^s_c}-u_z}{v_r}$, so that there is no linear system of equations to solve out to recover $T'(z)$.

### 10.2.3 Marginal Tax Rate for Top Incomes

We start by revisiting the classic result that if the income distribution is bounded at $z_{\max}$, then the top marginal income tax rate should be zero. In our model, this needs not be the case. One simple way to see that is to consider the case of decision vs. experienced utility. The tax formula (87) prescribes $T'(z_{\max})=\tilde{\tau}^b(z_{\max})$ which is positive or negative depending on whether top earners overperceive or underperceive the benefits of work (underperceive or overperceive the costs of work).

We now derive a formula for the marginal rate at very high incomes when the income distribution is unbounded at the top. It proves convenient to consider a (high) $z_0$ above which we consider

that incomes are "top incomes", and the marginal rate is constant. We consider tax systems with constant marginal tax rates for $z \geq z_0$. We assume that $g(z) = \overline{g}$ for $z \geq z_0$. We call $\zeta_{\overline{q}}^c(z) = \int_{z_0}^{\infty} \zeta_{Q_{z^*}}^c(z) dz_*$ the sensitivity to the asymptotic tax rate. This is the elasticity of earnings of an individual at earnings $z < z_0$ to an increase to the top rate, arising perhaps because of a misperception of the tax environment. Concretely, think of the recent case of France where increasing the top rate to 75% might have created an adverse general climate with the perception that even earners the top income would pay higher taxes.

We call $\overline{\eta}$, $\overline{g}, \overline{\zeta}^c$ the asymptotic values for large incomes and $\pi$ the Pareto exponent of the earnings distribution (i.e. when $z$ is large, $1 - H(z) \propto z^{-\pi}$). We define the weighted means: $\mathbb{E}^z[\phi(z)] = \frac{\int \phi(z) h(z) z dz}{\int \phi(z) z dz}$ and $\mathbb{E}^*\left[\zeta_{\overline{q}}^c\right] = \frac{\int \zeta_{\overline{q}}^c(z) \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} h^*(z) dz}{\int \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} h^*(z) dz}$.

**Proposition 10.2** (Optimal tax rate for top incomes) *The optimal marginal rate $\overline{\tau}$ for top incomes is*

$$\overline{\tau} = \frac{1 - \overline{g} - \beta + \overline{\zeta}^c \pi \overline{g} \tau^b}{1 - \overline{g} - \beta + \overline{\zeta}^c \pi + \overline{\eta}}, \tag{89}$$

*where*

$$\beta = \mathbb{E}^z \left[ \frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)} \right] \pi \frac{\mathbb{E}[z]}{\mathbb{E}[z 1_{z \geq z_0}]} \mathbb{E}^*\left[\zeta_{\overline{q}}^c\right].$$

This generalizes the Saez (2001) formula which can be recovered in the particular case where $\beta = \tau^b = 0$. The intuition is as follows—the $\beta$ term reflects not only the fact that the top marginal tax rate affects not only top earners, but also the tax perceived by agents at all points of the income distribution with associated effects on tax revenues. The more increasing the top tax rate lowers all incomes (the higher $\zeta_{\overline{q}}^c(z)$), the higher $\beta$, and the lower the optimal top tax rate.

The $\tau^b$ terms are positive (resp. negative) when top earners overperceive (resp. underperceive) the marginal benefits of effort or underperceive (resp. overperceive) taxes. These terms lead to higher (resp. lower) optimal top rates compared to the Saez formula.

Consider the typical Saez calibration with $\zeta^c(\infty) = 0.2$, $\eta = 0$ and $\pi = 2$. If the typical tax is $T'(z) \simeq \frac{1}{3}$ so that $\mathbb{E}^z\left[\frac{T'(z)}{1 - T'(z)}\right] \simeq \frac{1}{2}$, we take $z_0$ to be at the top 1% quantile of the income distribution. Piketty and Saez (2003) (updated 2015) report that the income share of the top 1% is 20%, so that $\frac{\mathbb{E}[z]}{\mathbb{E}\left[z 1_{z \geq z_0}\right]} = \frac{1}{0.2}$. This implies that $\beta = \mathbb{E}^z\left[\frac{T'(z) - \tilde{\tau}^b(z)}{1 - T'(z)}\right] \pi \frac{\mathbb{E}[z]}{\mathbb{E}\left[z 1_{z \geq z_0}\right]} \mathbb{E}^*\left[\zeta_{\overline{q}}^c\right] = \frac{1}{2} 2 \frac{1}{0.2} \mathbb{E}^*\left[\zeta_{\overline{q}}^c\right] = 5E^*\left[\zeta_{\overline{q}}^c\right]$. Also, we take top earnings to be well calibrated, i.e. $\tau^b = 0$.

The average cross-influence $\mathbb{E}^*\left[\zeta_{\overline{q}}^c\right]$ does not appear to have ever been measured. It is assumed to be 0 in the traditional model. We propose the following thought experiment to gauge its potential magnitude. Suppose that increasing the top rate by 10% will decrease earnings outside the top bracket by $x = 1\%$. Then, $\mathbb{E}^*\left[\zeta_{\overline{q}}^c\right] = (1 - T'(z)) \frac{z_{\overline{q}}}{z} = \left(1 - \frac{1}{3}\right) \frac{x}{0.1} = 6.7x$, which gives an interpretable benchmark that we now use.

Take first the case where $\overline{g} = 0$, i.e. where the top optimal tax rate maximizes revenues raised from top earners. With rational agents ($x = 0$), the top marginal tax rate is $\overline{\tau} = 71\%$. If $x = 1\%$,

then $\overline{\tau} = 62\%$, and if $x = 2\%$, then $\overline{\tau} = 45\%$. If $x = -1\%$, then $\overline{\tau} = 77\%$.[91] When the weight on top earnings is higher, say $\overline{g} = 0.2$, the corresponding numbers for the top rate are: 67%, 53%, 25%, and 74%. This illustrates the potentially large importance of the behavioral cross-impact of the top tax rate, a sufficient statistic that is assumed to be zero in traditional analyses.

The behavioral wedge $\tau^b$ does not affect the optimal tax rate when $\overline{g} = 0$. When $\overline{g} = 0.5$, the optimal top rate increases from 56% to 67% when the internality goes from no misperception of taxes by top earners ($\tau^b = 0$) to underperception of taxes by top earners ($\tau^b = 0.5$).

### 10.2.4 Possibility of Negative Marginal Income Tax Rates

In the traditional model with no behavioral biases, negative marginal income tax rates can never arise at the optimum. With behavioral biases negative marginal income tax rates are possible at the optimum. To see this, consider for example the decision vs. experienced utility model with decision utility $u^s$ and assume that $u^s$ is quasilinear so that there are no income effects $u^s(c, z) = c - \phi\varkappa(z)$. We take experienced utility to be $u(c, z) = \theta c - \phi(z)$. Then the modified Saez formula (88) becomes

$$\frac{T'(z^*) - \widetilde{\tau}^b(z^*)}{1 - T'(z^*)} = \frac{1}{\zeta^c(z^*)} \frac{1 - H(z^*)}{z^* h^*(z^*)} \int_{z^*}^{\infty} (1 - g(z)) \frac{h(z)}{1 - H(z^*)} dz,$$

where $\tilde{\tau}^b(z) = -g(z)\phi'(z)\frac{\theta - 1}{\theta}$ by (85). When $\theta > 1$, we have $\widetilde{\tau}^b(z^*) < 0$, and it is possible for this formula to yield $T'(z^*) < 0$. This occurs if agents undervalue the benefits or overvalue the costs from higher labor supply. For example, it could be the case that working more leads to higher human capital accumulation and higher future wages, but that these benefits are underperceived by agents, which could be captured in reduced form by $\theta > 1$. Such biases could be particularly relevant at the bottom of the income distribution (see Chetty and Saez (2013) for a review of the evidence). If these biases are strong enough, the modified Saez formula could predict negative marginal income tax rates at the bottom of the income distribution. This could provide a behavioral rationale for the EITC program.[92] In parallel and independent work, Gerritsen (2016) and Lockwood (2017) derive a modified Saez formula in the context of decision vs. experienced utility model.Lockwood (2017) zooms in on the EITC program and provides an empirical analysis documenting significant present-bias among EITC recipients and shows that a calibrated version of the model goes a long way towards rationalizing the negative marginal tax rates associated with the EITC program.

This differs from alternative rationales for negative marginal income tax rates that have been put forth in the traditional literature. For example, Saez (2002) shows that if the Mirrlees model is extended to allow for an extensive margin of labor supply, then negative marginal income tax rates can arise at the optimum. We also provide a behavioral treatment of the Saez (2002) extensive margin of labor supply model next.

[91] We thank Thomas Piketty for suggesting to us that if workers are happier, and strike less, because taxes on the wealthy are high, then $x < 0$.

[92] The EITC program itself could be misperceived, see Chetty et al. (2013).

## 10.3 Mirrlees Problem 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 $c_i^s$, so that he decides according to

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

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

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

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

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

Hence, at the optimum:

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

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

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

# 11 Further Proofs and Derivations

## 11.1 General proofs and derivations

**Derivation of behavioral wedges in $\beta$-$\delta$ model in Section 2.1** Price of good 1 is normalized to be 1. For the consumer, $\max_{c_0,c_1} \ln c_0 + \beta\delta \ln c_1$ subject to $c_0 + \frac{c_1}{1+r} \leq w$. FOC: $\frac{1}{c_0} = \frac{\beta\delta/c_1}{1/(1+r)}$. Thus demand functions are $c_0 = \frac{1}{1+\beta\delta} w, c_1 = \frac{\beta\delta}{1+\beta\delta}(1+r) w$.

Subjectively perceived indirect utilitity and experienced indirect utility are

$$\begin{aligned}
v^s(w) &= \ln c_0 + \beta\delta \ln c_1 = \ln \frac{w}{1+\beta\delta} + \beta\delta \ln \frac{\beta\delta(1+r) w}{1+\beta\delta}, \\
v(w) &= \ln c_0 + \delta \ln c_1 = \ln \frac{w}{1+\beta\delta} + \delta \ln \frac{\beta\delta(1+r) w}{1+\beta\delta}.
\end{aligned}$$

To compute behavioral wedges, we have now $u_{c_o}^s = \frac{1}{c_0}, u_{c_1}^s = \frac{\beta\delta}{c_1}, u_{c_0} = \frac{1}{c_0}, u_{c_1} = \frac{\delta}{c_1}, v_w^s = \frac{1+\beta\delta}{w}, v_w = \frac{1+\delta}{w}$.

Therefore,

$$\tau_0^b = \frac{1/c_0}{\frac{1+\beta\delta}{w}} - \frac{1/c_0}{\frac{1+\delta}{w}} = \frac{\delta(1-\beta)}{1+\delta},$$
$$\tau_1^b = \frac{\beta\delta/c_1}{\frac{1+\beta\delta}{w}} - \frac{\delta/c_1}{\frac{1+\delta}{w}} = -\frac{1-\beta}{\beta(1+r)(1+\delta)}.$$

□

**Proof of Proposition 2.2** We observe that a tax $\tau_i$ modifies the externality as:

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

so $\frac{d\xi}{d\tau_i} = \frac{\sum_h \xi_{\boldsymbol{c}^h} \cdot \boldsymbol{c}_{q_i}^h}{1-\sum_h \xi_{\boldsymbol{c}^h} \cdot \boldsymbol{c}_\xi^h}$. The term $\frac{1}{1-\sum_h \xi_{\boldsymbol{c}^h} \cdot \boldsymbol{c}_\xi^h}$ 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 2.1)

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

Using Proposition 2.1,

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

□

**Proof of Proposition 2.4** We have, from Proposition 13.5

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

hence

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

Hence, Proposition 2.2 implies:

$$\sum_h \left(1 - \frac{\gamma^{\xi,h}}{\lambda}\right) \boldsymbol{C}^h = -\sum_h \left(\boldsymbol{S}^{C,h}\right)' \left(\boldsymbol{\tau} - \boldsymbol{\tau}^{\xi,h} - \widetilde{\boldsymbol{\tau}}^{b,h}\right) = -\sum_h \boldsymbol{M}^{h\prime} \boldsymbol{S}^{h,r} \left[\left[\boldsymbol{I} - \left(\boldsymbol{I} - \boldsymbol{M}^h\right) \frac{\beta^h}{\lambda}\right] \boldsymbol{\tau} - \boldsymbol{\tau}^{X,h}\right]. \tag{90}$$

□

**Derivation of** (46) To make use of Proposition 2.1, we need to express demand function, Slutsky matrix, behavioral wedge and social marginal utility of income for each consumer in terms of taxes, in the limit of small taxes.

First, we approximate demand function up to the second order and decompose it,

$$\boldsymbol{c}^h\left(\boldsymbol{p} + \boldsymbol{\tau}, w\right) = \boldsymbol{c}^{r,h}\left(\boldsymbol{p} + \boldsymbol{\tau}, w\right) + \check{\boldsymbol{c}}^h\left(\boldsymbol{p}, \boldsymbol{\tau}, w\right), \tag{91}$$

with

$$\check{\boldsymbol{c}}^h\left(\boldsymbol{p}, \boldsymbol{\tau}, w\right) = \hat{\boldsymbol{c}}^{u,h}\left(\boldsymbol{p}, w\right) + \hat{\boldsymbol{c}}^{M,h}\left(\boldsymbol{p}, w\right) \boldsymbol{\tau} + O\left(\left(\tilde{\eta}_1^h\right)^2\right),$$

and $\tilde{\eta}_1^h = \max\left(\|\boldsymbol{\tau}\|, \|\hat{\boldsymbol{c}}^{u,h}\left(\boldsymbol{p}, w\right)\|\right)$. We can also write:

$$\boldsymbol{c}^h\left(\boldsymbol{p} + \boldsymbol{\tau}, w\right) = \boldsymbol{c}^h\left(\boldsymbol{p}, w\right) + \boldsymbol{c}_{\boldsymbol{p}}^{r,h}\left(\boldsymbol{p}, w\right) \boldsymbol{M}^h \boldsymbol{\tau} + O\left(\left(\tilde{\eta}_1^h\right)^2\right),$$

where $\boldsymbol{M}^h$ is the attention matrix of consumer $h$:

$$\boldsymbol{M}^h = \boldsymbol{I} + \left(\boldsymbol{c}_{\boldsymbol{p}}^{r,h}\left(\boldsymbol{p}, w\right)\right)^{-1} \hat{\boldsymbol{c}}^{M,h}\left(\boldsymbol{p}, w\right).$$

We also observe that, thanks to envelope theorem (noting that $\left(\boldsymbol{p} + \boldsymbol{\tau}\right) \cdot \check{\boldsymbol{c}}^h\left(\boldsymbol{p}, \boldsymbol{\tau}, w\right) = 0$ by the budget constraint)

$$\begin{aligned} v^h\left(\boldsymbol{q}, w\right) &= u^h\left(\boldsymbol{c}^{r,h}\left(\boldsymbol{q}, w\right) + \check{\boldsymbol{c}}^h\left(\boldsymbol{p}, \boldsymbol{\tau}, w\right) + O\left(\left(\tilde{\eta}_1^h\right)^2\right)\right) \\ &= u^h\left(\boldsymbol{c}^{r,h}\left(\boldsymbol{q}, w\right)\right) + O\left(\left(\tilde{\eta}_1^h\right)^2\right). \end{aligned}$$

With these, we derive Slutsky matrices

$$\boldsymbol{S}^{C,h} = \frac{\partial \boldsymbol{c}^h}{\partial \boldsymbol{q}} + \boldsymbol{c}_w^h \boldsymbol{c}^{h\prime} = \boldsymbol{c}_{\boldsymbol{p}}^{r,h} + \hat{\boldsymbol{c}}^{M,h} + \boldsymbol{c}_w^h \boldsymbol{c}^{h\prime} + O\left(\tilde{\eta}_1\right) = \boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h} + O\left(\tilde{\eta}_1^h\right), \tag{92}$$

and behavioral wedges

$$\begin{aligned} \boldsymbol{\tau}^{b,h} &= \boldsymbol{q} - \frac{u_{\boldsymbol{c}}^h\left(\boldsymbol{c}^h\left(\boldsymbol{q}, w\right)\right)}{v_w^h\left(\boldsymbol{q}, w\right)} = \boldsymbol{q} - \frac{u_{\boldsymbol{c}}^h\left(\boldsymbol{c}^{r,h}\left(\boldsymbol{q}, w\right)\right) + u_{\boldsymbol{cc}}^h\left(\boldsymbol{q}, w\right) \check{\boldsymbol{c}}^h}{v_w^h\left(\boldsymbol{q}, w\right)} + O\left(\left(\tilde{\eta}_1^h\right)^2\right) \\ &= \Omega^h \check{\boldsymbol{c}}^h + O\left(\left(\tilde{\eta}_1^h\right)^2\right) = \boldsymbol{\tau}^{b,h,nat} + \Omega^h \hat{\boldsymbol{c}}^{M,h} \boldsymbol{\tau} + O\left(\left(\tilde{\eta}_1^h\right)^2\right), \end{aligned} \tag{93}$$

in which we define

$$\boldsymbol{\tau}^{b,h,nat} = \Omega^h \hat{\boldsymbol{c}}^{u,h}(\boldsymbol{p}, w), \qquad \Omega^h = -\frac{u^h_{\boldsymbol{cc}}(\boldsymbol{p}, w)}{v^h_w(\boldsymbol{p}, w)},$$

which are constants evaluated at 0 taxes.

For small $\lambda - b^h$, up to the second order, social welfare function, social marginal welfare weight and social marginal utility of income are

$$\begin{aligned}
W &= \sum_h \frac{b^h}{v^h_w(\boldsymbol{p}, w)} v^h(\boldsymbol{p} + \boldsymbol{\tau}, w), \\
\beta^h &= W_{v^h} v^h_w = b^h \frac{v^h_w(\boldsymbol{p} + \boldsymbol{\tau}, w)}{v^h_w(\boldsymbol{p}, w)} = b^h \left(1 + \frac{v^h_{w,\boldsymbol{p}}(\boldsymbol{p}, w)}{v^h_w(\boldsymbol{p}, w)} \cdot \boldsymbol{\tau}\right), \\
\gamma^h &= \beta^h + \lambda \boldsymbol{\tau} \cdot \boldsymbol{c}^h_w = b^h \left(1 + \left(\frac{v^h_{w,\boldsymbol{p}}(\boldsymbol{p}, w)}{v^h_w(\boldsymbol{p}, w)} + \frac{\lambda}{b^h} \boldsymbol{c}^h_w(\boldsymbol{p}, w)\right) \cdot \boldsymbol{\tau}\right).
\end{aligned}$$

By Roy's identity $v^h_{\boldsymbol{p}}(\boldsymbol{p}, w) = -\boldsymbol{c}^h(\boldsymbol{p}, w) v^h_w(\boldsymbol{p}, w)$, we have $\frac{v^h_{w,\boldsymbol{p}}(\boldsymbol{p},w)}{v^h_w(\boldsymbol{p},w)} = -\boldsymbol{c}^h_w(\boldsymbol{p}, w) - \boldsymbol{c}^h(\boldsymbol{p}, w) \frac{v^h_{ww}(\boldsymbol{p},w)}{v^h_w(\boldsymbol{p},w)}$,

$$\begin{aligned}
\frac{\gamma^h}{b^h} &= 1 + \left[\left(-1 + \frac{\lambda}{b^h}\right) \boldsymbol{c}^h_w(\boldsymbol{p}, w) - \boldsymbol{c}^h(\boldsymbol{p}, w) \frac{v^h_{ww}(\boldsymbol{p}, w)}{v^h_w(\boldsymbol{p}, w)}\right] \cdot \boldsymbol{\tau} \\
&= 1 - \frac{v^h_{ww}(\boldsymbol{p}, w)}{v^h_w(\boldsymbol{p}, w)} \boldsymbol{c}^h(\boldsymbol{p}, w) \cdot \boldsymbol{\tau} + O\left(\left(\tilde{\eta}^h_2\right)^2\right),
\end{aligned} \tag{94}$$

where we define $\tilde{\eta}^h_2 = \max\left(\left|b^h - \lambda\right|, \|\boldsymbol{\tau}\|, \|\hat{\boldsymbol{c}}^{u,h}(\boldsymbol{p}, w)\|\right)$. In addition, we have $\tilde{\boldsymbol{\tau}}^{b,h} = \frac{\beta^h}{\lambda} \boldsymbol{\tau}^b = \boldsymbol{\tau}^b + O\left(\left(\tilde{\eta}^h_2\right)^2\right)$.

We remember that good 0 is not taxed. So, in all expressions below, the subscript $> 0$ indicates the selection of the $(N-1) \times (N-1)$ sub-matrix corresponding to all goods except good 0. Plugging these results in the behavioral multi-person Ramsey formula (8),

$$\begin{aligned}
\mathbf{0} &= \sum_h \left[\left(1 - \frac{\gamma^h}{\lambda}\right) \boldsymbol{c}^h(\boldsymbol{q}, w) + \boldsymbol{S}^{C,h\prime}\left(\boldsymbol{\tau} - \tilde{\boldsymbol{\tau}}^{b,h}\right)\right]_{>0} \\
&= \sum_h \left(1 - \frac{b^h}{\lambda} + \frac{b^h}{\lambda} \frac{v_{ww}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)} \boldsymbol{c}^h(\boldsymbol{p}, w) \cdot \boldsymbol{\tau}\right) \left[\boldsymbol{c}^h(\boldsymbol{p}, w)\right]_{>0} \\
&\quad + \left[\sum_h \left(\boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h}(\boldsymbol{p}, w)\right)' \left(\left(\boldsymbol{I} - \Omega^h \hat{\boldsymbol{c}}^{M,h}(\boldsymbol{p}, w)\right) \boldsymbol{\tau} - \Omega^h \hat{\boldsymbol{c}}^{u,h}(\boldsymbol{p}, w)\right)\right]_{>0} + O\left(\sum_h \left(\tilde{\eta}^h_2\right)^2\right) \\
&= \sum_h \left[\frac{v_{ww}(\boldsymbol{p}, w)}{v_w(\boldsymbol{p}, w)} \boldsymbol{c}^h(\boldsymbol{p}, w) \left(\boldsymbol{c}^h(\boldsymbol{p}, w)\right)' + \left(\boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h}(\boldsymbol{p}, w)\right)' \left(\boldsymbol{I} - \Omega^h \hat{\boldsymbol{c}}^{M,h}(\boldsymbol{p}, w)\right)\right]_{>0} \boldsymbol{\tau} \\
&\quad + \left[\sum_h \left[\left(1 - \frac{b^h}{\lambda}\right) \boldsymbol{c}^h(\boldsymbol{p}, w) - \left(\boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h}(\boldsymbol{p}, w)\right)' \Omega^h \hat{\boldsymbol{c}}^{u,h}(\boldsymbol{p}, w)\right]\right]_{>0} + O\left(\sum_h \left(\tilde{\eta}^h_2\right)^2\right),
\end{aligned}$$

which is solved by optimal tax

$$\boldsymbol{\tau} = -\left[\sum_h \frac{v_{ww}}{v_w}\boldsymbol{c}^h\boldsymbol{c}^{h\prime} + \left(\boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h}\right)'\left(\boldsymbol{I} - \Omega^h\hat{\boldsymbol{c}}^{M,h}\right)\right]^{-1}_{>0}\left[\sum_h\left(1-\frac{b^h}{\lambda}\right)\boldsymbol{c}^h - \left(\boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h}\right)'\Omega^h\hat{\boldsymbol{c}}^{u,h}\right]_{>0} + O\left(\tilde{\eta}^2\right) \tag{95}$$

with $\tilde{\eta} = \sum_h \left|b^h - \lambda\right| + \|\hat{\boldsymbol{c}}^{u,h}(\boldsymbol{p}, w)\|$, $\boldsymbol{S}^{r,h} = \boldsymbol{c}^h_{\boldsymbol{p}} + \boldsymbol{c}^h_w\boldsymbol{c}^{h\prime}$, in which all the variables are evaluated at $(\boldsymbol{p}, w)$. [93]□

**Proof of Proposition 3.1** We start from the Ramsey planning problem in (16). Define

$$L = \gamma\sum_{i=1}^n\left[\frac{(c_i(\tau_i))^{1-1/\psi_i}-1}{1-1/\psi_i} - (p_i+\tau_i)c_i(\tau_i)\right] + \lambda\sum_{i=1}^n\tau_i c_i(\tau_i)$$

where $c_i = (p_i + m_i\tau_i)^{-\psi_i}$. The first-order condition with respect to $\tau_i$ is:

$$L_{\tau_i} = \gamma\left[[(c_i(\tau_i))^{-1/\psi_i} - (p_i+\tau_i)]\frac{\partial c_i}{\partial\tau_i} - c_i(\tau_i)\right] + \lambda\left[c_i(\tau_i) + \tau_i\frac{\partial c_i}{\partial\tau_i}\right] = 0$$

Note that $c_i(\tau_i)^{-1/\psi_i} = p_i + m_i\tau_i$ and $\partial c_i/\partial\tau_i = -\psi_i\frac{c_i}{p_i+m_i\tau_i}m_i$, we can rewrite the FOC as:

$$\begin{aligned} L_{\tau_i} &= \gamma\left[\left(\frac{\lambda}{\gamma} - 1 + m_i\right)\tau_i\frac{-\psi_i c_i(\tau_i)m_i}{p_i+m_i\tau_i}\right] + (\lambda-\gamma)c_i(\tau_i) \\ &= -\lambda\left(\Lambda + \frac{\gamma}{\lambda}m_i\right)\frac{\psi_i\tau_i c_i(\tau_i)m_i}{p_i+m_i\tau_i} + \lambda\Lambda c_i(\tau_i) = 0 \end{aligned}$$

Simplifying gives us:

$$\left(\Lambda + \frac{\gamma}{\lambda}m_i\right)\psi_i\tau_i m_i = \Lambda(p_i + m_i\tau_i)$$

which gives an explicit expression for $\tau_i$:

$$\frac{\tau_i}{p_i} = \frac{\Lambda}{\psi_i m_i}\frac{1}{\Lambda + (1-\Lambda)m_i - \Lambda/\psi_i} = \frac{\Lambda}{\psi_i m_i^2}\frac{1}{1+\Lambda\left(\frac{1-m_i-1/\psi_i}{m_i}\right)}.$$

□

**Derivation of (18), the approximate loss from taxation of inattentive agents** We define the behavioral elasticity $\alpha_i = m_i\psi_i$ and denote $y_i$ as the expenditure on good $i$ at zero tax

[93] $\boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h}$ can also be expressed as

$$\boldsymbol{S}^{r,h} + \hat{\boldsymbol{c}}^{M,h} = \boldsymbol{S}^{r,h} + \boldsymbol{c}^{r,h}_{\boldsymbol{p}}\left(\boldsymbol{M}^h - \boldsymbol{I}\right) = \boldsymbol{S}^{r,h} + \left(\boldsymbol{S}^{r,h} - \boldsymbol{c}^h_w\boldsymbol{c}^{h\prime}\right)\left(\boldsymbol{M}^h - \boldsymbol{I}\right) = \boldsymbol{S}^{r,h}\boldsymbol{M}^h - \boldsymbol{c}^h_w\boldsymbol{c}^{h\prime}\left(\boldsymbol{M}^h - \boldsymbol{I}\right).$$

rate. Demand is:

$$c_i(\tau_i) = \frac{y_i}{p_i}\left(1 - \alpha_i \frac{\tau_i}{p_i} + O\left(\left(\frac{\tau_i}{p_i}\right)^2\right)\right), \tag{96}$$

Hence have:

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

with

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

We have

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

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

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

$$\begin{aligned} f_i'(0) &= 0 \\ f_i''(0) &= F''(c_i(0)) c_i'(0)^2 \\ &= U_i''\left(\frac{y_i}{p_i}\right)\frac{y_i}{p_i^2}\alpha_i^2 \text{ using (96)} \\ &= -\frac{\alpha_i^2}{\psi_i}\frac{y_i}{p_i^2}, \end{aligned}$$

so

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

So the objective function is:

$$L = -\frac{1}{2}\sum_i \frac{\alpha_i^2}{\psi_i} y_i \left(\frac{\tau_i}{p_i}\right)^2 + \Lambda \sum_i \frac{\tau_i}{p_i} y_i + o\left(\|\boldsymbol{\tau}\|^2\right) + o(\|\boldsymbol{\tau}\|\Lambda). \tag{97}$$

□

**Derivation of (22), the endogenous social cost of public funds** Now the objective function of the government is

$$\begin{aligned} W(\boldsymbol{\tau}) &= \gamma u(\boldsymbol{C}(\boldsymbol{\tau})) + V(G(\boldsymbol{\tau})) \\ &= \gamma \sum_i \left[ \frac{\left[ (p_i + m_i \tau_i)^{-\psi_i} \right]^{1-1/\psi_i} - 1}{1 - 1/\psi_i} - (p_i + \tau_i)(p_i + m_i \tau_i)^{-\psi_i} \right] + V\left( \sum_i (p_i + m_i \tau_i)^{-\psi_i} \tau_i \right), \end{aligned}$$

where $V(G)$ is the utility from public goods $G$,

$$G(\boldsymbol{\tau}) = \sum_i (p_i + m_i \tau_i)^{-\psi_i} \tau_i.$$

A Taylor expansion gives a second order approximation:

$$\begin{aligned} W(\boldsymbol{\tau}) - W(0) = \sum_i c_i(0) p_i \left[ -\frac{1}{2} \psi_i m_i (m_i - 2) \left( \frac{\tau_i}{p_i} \right)^2 - \frac{\tau_i}{p_i} \right] \\ + V'(0) g + \frac{1}{2} V''(0) g^2 - V'(0) \sum_i c_i(0) p_i \psi_i m_i \left( \frac{\tau_i}{p_i} \right)^2 + o\left( \|\boldsymbol{\tau}\|^2 \right), \\ g = \sum_i \tau_i c_i(0) = \sum_i c_i(0) p_i \frac{\tau_i}{p_i}. \end{aligned}$$

We take $V'(0) = 1 + \Lambda_0$ with $\Lambda_0$ small, as we are taking the limit of small taxes. So,

$$\begin{aligned} W(\boldsymbol{\tau}) - W(0) = -\frac{1}{2} \sum_i c_i(0) p_i \psi_i m_i^2 \left( \frac{\tau_i}{p_i} \right)^2 + \Lambda_0 \sum_i c_i(0) p_i \frac{\tau_i}{p_i} + \frac{1}{2} V''(0) \left( \sum_i c_i(0) p_i \frac{\tau_i}{p_i} \right)^2 \\ + o\left( \|\boldsymbol{\tau}\|^2 \right) + O\left( \Lambda_0 \|\boldsymbol{\tau}\|^2 \right). \end{aligned}$$

Optimizing over $\tau_i$ gives $\frac{\tau_i}{p_i} = \frac{\Lambda}{\psi_i m_i^2}$ with $\Lambda = \Lambda_0 + V''(0) g$. Note that $\Lambda_0$ and $g$ are of the same order of magnitude.

Then we can solve for $\Lambda$:

$$\begin{aligned} g &= \sum_i c_i(0) p_i \frac{\tau_i}{p_i} = \alpha \Lambda, \\ \alpha &= \sum_i \frac{c_i(0) p_i}{\psi_i m_i^2}, \end{aligned}$$

which gives that

$$\Lambda = \frac{\Lambda_0}{1 - \alpha V''(0)} = \frac{V'(0) - 1}{1 - \alpha V''(0)}.$$

When $V''(0) < 0$, we find $\frac{\partial \Lambda}{\partial m_i} > 0$. So, when attention is lower, the marginal utility of public fund (at the optimum) is lower, mitigating the increase in taxes.□

**Proof of Proposition 3.3** The government's planning problem is

$$\sum_h U^h\left(c^h\right) - \left(p + \xi^h\right) c^h. \tag{98}$$

We call $c^{*h} = \arg\max_{c^h} U^h\left(c^h\right) - \left(p + \xi^h\right) c^h$ the quantity consumed by the agent at the first best.

To make things transparent, we specify

$$U^h(c) = \frac{a^h c - \frac{1}{2}c^2}{\Psi},$$

which using $U_c^h = \frac{a^h - c}{\Psi} = q^s$, implies a demand function $c^h(q^s) = a^h - \Psi q^s$.[94]

After some algebraic manipulations, social welfare compared to the first best can be written as

$$L(\tau) = -\frac{\Psi}{2}\sum_h \left(m^h \tau - \xi^h\right)^2. \tag{99}$$

The first best cannot be implemented unless all agents have the same ideal Pigouvian tax, $\xi^h/m^h$. Heterogeneity in attention creates welfare losses.

*Optimal Pigouvian tax.* At the optimum, $U_c^h\left(c^{h*}\right) = p + \xi^h$. 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 $\Psi = \frac{1}{U_{cc}^h\left(c^{h*}\right)}$ then. The social welfare is $L = \sum_h L^h = -\frac{\Psi}{2}\sum_h \left(m^h\tau - \xi^h\right)^2$ by (99).

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

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

[94]The expressions in the rest of this section are exact with this quadratic utility specification. For general utility functions, they hold provided that they are understood as the leading order terms in a Taylor expansion around an economy with no heterogeneity.

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

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

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

If there is no tax, the loss is (from equation (99))

$$L^{\text{no tax}} = -\frac{\Psi}{2}\sum_h \left(m^h \cdot 0 - \xi^h\right)^2 = -\frac{\Psi}{2}\sum_h \xi^{h2} = -\frac{1}{2}H\Psi\mathbb{E}\left[\xi^{h2}\right].$$

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

*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

$$\frac{1}{H}\sum_h U_c^h\left(c^*\right) = p + \frac{1}{H}\sum_h \xi^h. \tag{100}$$

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

$$L^h = \frac{1}{2}U_{cc}^h\left(c\right)\left(c^{h*} - c^*\right)^2 = -\frac{1}{2}\frac{1}{\Psi}\left(c^{h*} - c^*\right)^2.$$

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

The loss is:

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

The inequality $\frac{1}{2\Psi}var\left(c^{h*}\right) < \Psi\frac{\mathbb{E}\left[\xi^{h2}\right]\mathbb{E}\left[m^{h2}\right] - \left(\mathbb{E}\left[\xi^h m^h\right]\right)^2}{2\mathbb{E}\left[m^{h2}\right]}$ can also be written as $\frac{1}{2\Psi}var\left(c^{h*}\right) < \frac{\Psi}{2}var_{m^{h2}}\left(\xi^h/m^h\right)\mathbb{E}\left[m^{h2}\right]$, since

$$\mathbb{E}_{m^{h^2}}\left(\xi^h/m^h\right) = \frac{\sum_h m^{h^2}\xi^h/m^h}{\sum_h m^{h^2}} = \frac{\mathbb{E}\left[\xi^h m^h\right]}{\mathbb{E}\left[m^{h^2}\right]},$$

$$\begin{aligned} var_{m^{h^2}}\left(\xi^h/m^h\right) &= \frac{\sum_h m^{h^2}\left(\xi^h/m^h - \mathbb{E}_{m^{h^2}}\left(\xi^h/m^h\right)\right)^2}{\sum_h m^{h^2}} = \frac{\sum_h m^{h^2}\left(\frac{\xi^h}{m^h} - \frac{\mathbb{E}\left[\xi^h m^h\right]}{\mathbb{E}\left[m^{h^2}\right]}\right)^2}{\sum_h m^{h^2}} \\ &= \frac{\sum_h \xi^{h^2} - 2\sum_h \xi^h m^h \frac{\mathbb{E}\left[\xi^h m^h\right]}{\mathbb{E}\left[m^{h^2}\right]} + \sum_h m^{h^2}\left(\frac{\mathbb{E}\left[\xi^h m^h\right]}{\mathbb{E}\left[m^{h^2}\right]}\right)^2}{\sum_h m^{h^2}} = \frac{\mathbb{E}\left[\xi^{h^2}\right]\mathbb{E}\left[m^{h^2}\right] - \left(\mathbb{E}\left[\xi^h m^h\right]\right)^2}{\left(\mathbb{E}\left[m^{h^2}\right]\right)^2}. \end{aligned}$$

□

**Proof of Proposition 3.4** Equation (13) then yields the optimal tax:

$$\boldsymbol{\tau} = \left(\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^r\boldsymbol{M}^h\right]\right)^{-1}\mathbb{E}\left[\boldsymbol{M}^{h\prime}\right]\boldsymbol{S}^r\boldsymbol{\tau}^X. \tag{101}$$

with$\boldsymbol{\tau}^X = (\xi_*, 0)'$ .

When agents have uniform misperceptions ($\boldsymbol{M}^h = \boldsymbol{M}$), the optimal tax is $\boldsymbol{\tau} = \boldsymbol{M}^{-1}\boldsymbol{\tau}^X$. This implies $\tau_1 = \frac{\xi_*}{m_1} > 0$ and $\tau_2 = 0$. The principle of targeting applies. This is no longer true when misperceptions are not uniform.

We have $\left(\mathbb{E}\left[\boldsymbol{M}^{h'}\boldsymbol{S}^r\boldsymbol{M}^h\right]\right)_{ij} = S^r_{ij}\mathbb{E}\left[m^h_i m^h_j\right]$ and $\left(\mathbb{E}\left[\boldsymbol{M}^{h\prime}\boldsymbol{S}^r\right]\right)_{ij} = \mathbb{E}\left[m^h_i\right]S^r_{ij}$. Matrix inversion gives:

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

Because $\mathbb{E}\left[\boldsymbol{M}^{h'}\boldsymbol{S}^r\boldsymbol{M}^h\right]$ is a dimension $2\times 2$ 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), $\det\mathbb{E}\left[\boldsymbol{M}^{h'}\boldsymbol{S}^r\boldsymbol{M}^h\right] > 0$. 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, $sign\left(\tau_2\right) = -sign\left(S_{12}\right)$.

The quadratic case simply gives a constant matrix $\boldsymbol{S}^r$.□

**Proof of Proposition 3.5** We apply Proposition 2.4. Here,

$$\boldsymbol{M}^h = \boldsymbol{I}, \boldsymbol{S}^{r,h*} = \begin{pmatrix} -\psi_1\left(p_1+\tau_1\right)^{-\psi_1-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}^{r,h*}_{\boldsymbol{C_2}} \end{pmatrix}, \boldsymbol{S}^{r,h\neq h*} = \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}^{r,h\neq h*}_{\boldsymbol{C_2}} \end{pmatrix},$$

$$c_1^{h*} = \left(p_1+\tau_1\right)^{-\psi_1}, \boldsymbol{\tau}^{X,h*} = \begin{pmatrix} \frac{\gamma^{h*}}{\lambda}\xi^{h*} \\ \vdots \end{pmatrix}, \boldsymbol{\tau}^{X,h\neq h*} = \begin{pmatrix} 0 \\ \vdots \end{pmatrix}.$$

We plug the results above into (13). Suppose $\rho_p, \rho_r$ are the portions of agents $h^*$ and agents $h \neq h^*$ in the population with $\rho_p + \rho_r = 1$. Then

$$-\left[\sum_h \boldsymbol{M}^{h\prime}\boldsymbol{S}^{r,h}\left(\boldsymbol{I}-\left(\boldsymbol{I}-\boldsymbol{M}^h\right)\frac{\gamma^h}{\lambda}\right)\right]^{-1} = \frac{1}{H}\begin{pmatrix} \rho_p^{-1}\psi_1^{-1}\left(p_1+\tau_1\right)^{\psi_1+1} & \mathbf{0} \\ \mathbf{0} & -\left(\rho_p \boldsymbol{S}^{r,h*}_{\boldsymbol{C_2}} + \rho_r \boldsymbol{S}^{r,h\neq h*}_{\boldsymbol{C_2}}\right)^{-1} \end{pmatrix},$$

$$\sum_h\left[\left(1-\frac{\gamma^h}{\lambda}\right)\boldsymbol{C}^h - \boldsymbol{M}^{h\prime}\boldsymbol{S}^{r,h}\boldsymbol{\tau}^{X,h}\right] = H\begin{pmatrix} \rho_p\left[\left(1-\frac{\gamma^{h*}}{\lambda}\right)\left(p_1+\tau_1\right)^{-\psi_1} + \frac{\gamma^{h*}}{\lambda}\xi^{h*}\psi_1\left(p_1+\tau_1\right)^{-\psi_1-1}\right] \\ \vdots \end{pmatrix}$$

and get

$$\tau_1 = \left(1-\frac{\gamma^{h*}}{\lambda}\right)\frac{p_1+\tau_1}{\psi_1} + \frac{\gamma^{h*}}{\lambda}\xi^{h*},$$
$$\tau_1 = \frac{\frac{\gamma^{h*}}{\lambda}\xi^{h*} + \left(1-\frac{\gamma^{h*}}{\lambda}\right)\frac{p_1}{\psi_1}}{1+\left(\frac{\gamma^{h*}}{\lambda}-1\right)\frac{1}{\psi_1}}.$$

□

## 11.2 Helpful tools for the calculations

**Lemma 11.1** *(Losses from lack of equalization of marginal utilities). Suppose we have* $F(x) = \sum_i f^i(x_i)$*, and* $F^{FB} = \max_y F(y)$ *s.t.* $\sum y_i = \sum x_i$*. Then,*

$$F(x) - F^{FB} = \frac{1}{2}\frac{var\left(f^{i\prime}(x_i)\right)}{f''(x)} \tag{102}$$

*If the budget constraint is* $F^{FB} = \max_{y_i} F(y)$ *s.t.* $\sum B(y) = \sum x_i$*.*

**Proof**. At the FB, $f'^i(x_i) = \lambda$

$$F(x) = \sum_i f^i(x_i) = \sum_i f^i\left(x_i^{FB}\right) + f^{i\prime}\left(x_i^{FB}\right)\left(x_i - x_i^{FB}\right) + \frac{1}{2}f^{i\prime\prime}\left(x_i^{FB}\right)\left(x_i - x_i^{FB}\right)^2$$
$$F(x) - F^{FB} = \sum_i \lambda\left(x_i - x_i^{FB}\right) + \frac{1}{2}f^{i\prime\prime}\left(x_i^{FB}\right)\left(x_i - x_i^{FB}\right)^2$$
$$= \sum_i \frac{1}{2}f^{i\prime\prime}\left(x_i^{FB}\right)\left(x_i - x_i^{FB}\right)^2$$

and

$$F(x) - F^{FB} = \frac{1}{2}\sum_i \frac{1}{f^{i\prime\prime}\left(x_i^{FB}\right)}\left(f^{i\prime}(x_i) - f^{i\prime}\left(x_i^{FB}\right)\right)^2 \text{ as } f^{i\prime}(x_i) - f^{i\prime}\left(x_i^{FB}\right) = f'' \cdot \left(x_i - x_i^{FB}\right)$$
$$= \frac{1}{2}\sum_i \frac{var\left(f^{i\prime}(x_i)\right)}{f^{i\prime\prime}\left(x_i^{FB}\right)} \text{ as } \sum_i f^{i\prime}(x_i) - f^{i\prime}\left(x_i^{FB}\right) = f''\sum_i x_i - x_i^{FB} = 0$$

□

**Lemma 11.2** *Suppose that $F(x)$ is more general, then the loss is still linked to the variance of marginal utilities:*

$$F(x) - F^{FB} = \frac{1}{2}\left[F_x(x) - \lambda' I\right] F_{xx}^{-1}\left[F_x(x) - \lambda' I\right]$$
$$= \frac{1}{2}\sum_{i,j}\left[F_{x_i}(x) - \lambda' I\right]\left(F_{xx}^{-1}\right)_{i,j}\left[F_{x_j}(x) - \lambda' I\right]$$

*with $\lambda' = \langle F_{x_i}\rangle$.*

**Proof**. At the FB, $F_{x_i} = \lambda$

$$F(x) - F\left(x^{FB}\right) = \sum_i F_{x_i}\left(x_i - x_i^{FB}\right) + \frac{1}{2}\left(x_i - x_i^{FB}\right)F_{x_i x_j}\left(x_j - x_j^{FB}\right)$$
$$= 0 + \frac{1}{2}\left(x_i - x_i^{FB}\right)F_{x_i x_j}\left(x_j - x_j^{FB}\right)$$
$$= \frac{1}{2} y F_{xx} y$$

and $F_x(x) - F_x\left(x^{FB}\right) = F_{xx}\left(x - x^{FB}\right) = F_{xx}y$, so $y = F_{xx}^{-1}\cdot\left(F_x(x) - F_x\left(x^{FB}\right)\right)$. Hence,

$$F(x) - F^{FB} = \frac{1}{2} y F_{xx} y = \frac{1}{2}\left[F_x(x) - F_x\left(x^{FB}\right)\right] F_{xx}^{-1}\left[F_x(x) - F_x\left(x^{FB}\right)\right]$$
$$= \frac{1}{2}\left[F_x(x) - \lambda' I\right] F_{xx}^{-1}\left[F_x(x) - \lambda' I\right] + o\left(\varepsilon^2\right) \text{ with } \lambda' = \langle F_{x_i}\rangle$$
$$= \frac{1}{2}\sum_{i,j}\left[F_{x_i}(x) - \lambda' I\right]\left(F_{xx}^{-1}\right)_{i,j}\left[F_{x_j}(x) - \lambda' I\right]$$

which is a variance term.□

**Lemma 11.3** *Suppose we have $f(x,y)$ and a function $X(y) = \arg\max_x f(x,y)$. Call $(x^*, y^*) = \arg\max_{x,y} f(x,y)$ and $g(y) = f(X(y), y)$. We have:*

$$f(x,y) - f(x^*, y^*) = \left[f(x,y) - f(X(y), y)\right] + \left[f(X(y), y) - f(X(y^*), y^*)\right]$$
$$= \frac{1}{2}(x - X(y))\cdot f_{xx}(X(y), y)\cdot(x - X(y)) + \frac{1}{2}(y - y^*)\cdot g_{yy}\cdot(y - y^*)$$

*Also,*

$$\begin{aligned} g_y &= f_y\left(X\left(y\right), y\right) \\ g_{yy} &= f_{yy} + f_{xy}X'\left(y\right) \\ &= f_{yy} - f_{xy}f_{xx}^{-1}f_{xy} \end{aligned}$$

*with* $X'\left(y\right) = -f_{xx}^{-1}f_{xy}$.

## 11.3 Additional derivations for the Mirrlees Problem

### 11.3.1 Intermediary results for the Mirrlees problem

**Impact of a change in taxes on earnings and individual utility** We first study the impact of a small change $\delta q_{z^*}$ 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 $\delta K$. It will prove conceptually and notationally useful to define:

$$\overline{\overline{\zeta}}^{c}_{Q_{z^*}}\left(z\right) = \zeta^{c}_{Q_{z^*}}\left(z\right) + \zeta^{c}\left(z\right)\delta_z\left(z^*\right), \tag{103}$$

where $\delta_z$ is a Dirac distribution at point $z$. Hence, as $\zeta^{c}_{Q_{z^*}}\left(z\right)$ was a potentially smooth function of $z^*$, $\overline{\overline{\zeta}}^{c}_{Q_{z^*}}\left(z\right)$ 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 11.4** (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* $\delta K$*. The impact on earnings and agent's welfare is:*

$$\delta z = \frac{\eta\delta K + z\int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*}{q - \zeta^{c}zR''}, \tag{104}$$

$$\frac{\delta v}{v_r} = \delta K - z\frac{\tau^b}{q}\left(\zeta^{c}R''\delta z + \int_0^\infty \overline{\overline{\zeta}}^{c}_{Q_{z^*}}\delta q_{z^*}dz^*\right). \tag{105}$$

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 $z\zeta^{c}\left(z\right)\delta q_z + \int_{z^*=0}^\infty z\zeta^{c}_{Q_{z^*}}\left(z\right)\delta q_{z^*}dz^*$ (reintroducing in these equations the dependency on $z$), leading to

$$\begin{aligned} \delta z &= \frac{\eta\left(z\right)\delta K + z\zeta^{c}\left(z\right)\delta q_z + \int_0^\infty z\zeta^{c}_{Q_{z^*}}\left(z\right)\delta q_{z^*}dz^*}{q\left(z\right) - \zeta^{c}\left(z\right)zR''\left(z\right)}, \\ \frac{\delta v}{v_r} &= \delta K - z\frac{\tau^b}{q}\zeta^{c}\left(z\right)R''\left(z\right)\delta z - z\frac{\tau^b\left(z\right)}{q\left(z\right)}\zeta^{c}\left(z\right)\delta q_Z - \int_{z^*=0}^\infty z\frac{\tau^b\left(z\right)}{q\left(z\right)}\zeta^{c}_{Q_{z^*}}\left(z\right)\delta q_{z^*}dz^*. \end{aligned}$$

To interpret the economics of (104), start with an increase in income $\delta K$. 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 $\zeta^c z R''$ in the denominator of (104) 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 $\frac{\zeta^c}{q} z R''\delta z$. The final expression solves for $\delta z$ as a fixed point. The term $z\int_0^\infty \overline{=}^c_{Q_{z^*}} \delta q_{z^*} dz^*$ 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 (105), the term $\delta K$ is a mechanical income effect and is the only term present in the traditional model of Saez (2001). The term $-z\frac{\tau^b}{q}\left(\zeta^c R''\delta z + \int_0^\infty \overline{=}^c_{Q_{z^*}} \delta q_{z^*} dz^*\right)$ represent the welfare impact arising from changes in behavior (as the envelope theorem no longer applies) 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 ($-z\frac{\tau^b}{q}\int_0^\infty \overline{=}^c_{Q_{z^*}} \delta q_{z^*} dz^*$).

**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 11.5** *Under the conditions of the Lemma 11.4, the change in the government objective function associated with the agent is*

$$\delta L(z) = (\gamma(z) - 1)\,\delta K + \frac{T'(z) - \widetilde{\tau}^b(z)}{1 - T'(z)} \frac{h^*(z)}{h(z)} z \int_0^\infty \overline{=}^c_{Q_{z^*}}(z)\,\delta q_{z^*} dz^*, \tag{106}$$

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

$$\gamma(z) = g(z) + \eta(z)\frac{\widetilde{\tau}^b(z)}{1 - T'(z)} + \eta(z)\frac{T'(z) - \widetilde{\tau}^b(z)}{1 - T'(z)}\frac{h^*(z)}{h(z)}. \tag{107}$$

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 (106) is much like in the many-person Ramsey of Proposition 2.1. The term $(\gamma(z) - 1)\,\delta K$ is a mechanical effect, abstracting from changes in behavior. As the government gives (back) $\delta K$ 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 $\frac{T'(z)}{1-T'(z)}\frac{h^*}{h} z \int_0^\infty \overline{=}^c_{Q_{z^*}}(z)\,\delta q_{z^*} dz^*$:

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 11.6** *At the optimum,*

$$\int_0^\infty \left(1-\gamma(z)-\frac{T'(z)-\widetilde{\tau}^b(z)}{1-T'(z)}\frac{h^*(z)}{h(z)}z\zeta^c_{r_0}(z)\right)h(z)\,dz=0. \tag{108}$$

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

$$\frac{\partial L}{\partial \tau_{z^*}}=\int_{z^*}^\infty(1-\gamma(z))\,h(z)\,dz-\zeta^c(z^*)\frac{T'(z^*)-\widetilde{\tau}^b(z^*)}{1-T'(z^*)}z^*h^*(z^*)-\int_0^\infty \zeta^c_{Q_{z^*}}(z)\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}zh^*(z)\,dz. \tag{109}$$

This equation involves an equality between two generalized functions of $z^*$. This is the income tax equivalent of the formula in Proposition 2.1 for the many-person Ramsey. The three terms in (109) 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 ($\zeta^c(z^*)\frac{\widetilde{\tau}^b(z^*)}{1-T'(z^*)}z^*h^*(z^*)-\int_0^\infty\zeta^c_{Q_{z^*}}(z)\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)}zh^*(z)\,dz$) 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 (109) can be written in a more compact way as:

$$\frac{\partial L}{\partial \tau_{z^*}}=\int_{z^*}^\infty(1-\gamma(z))\,h(z)\,dz-\int_0^\infty \overline{\overline{\zeta}}^c_{Q_{z^*}}(z)\frac{T'(z)-\widetilde{\tau}^b(z)}{1-T'(z)}zh^*(z)\,dz. \tag{110}$$

### 11.3.2 Proofs for the Mirrlees results

**Notations and Derivation of relation (80)** We take the material from section 7.1. 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 $B(\boldsymbol{c},\boldsymbol{q})=q_1c_1-q_2c_2=c-qz$, so that the budget constraint is $B(\boldsymbol{c},\boldsymbol{q})\leq r$. Note that the Saez $r$ is also the $w$ in the rest of the paper (as the budget constraint is generally expressed as $B(\boldsymbol{c},\boldsymbol{q})\leq r$); 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 (37) gives

$$\boldsymbol{\tau}^b = (1, -q) - \frac{(u_c, u_z)}{v_r}. \tag{111}$$

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 7.1 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

$$\tau^b = \boldsymbol{\tau}^b(\boldsymbol{q}, r) \cdot (q, 1) = -\frac{q u_c + u_z}{v_r}. \tag{112}$$

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

$$\begin{aligned}
\boldsymbol{S}^C_q(\boldsymbol{q}, r) &= \boldsymbol{c}_q - \boldsymbol{c}_r z = (c_q - c_r z, z_q - z_r z) \\
&= (q z_q + z - (q z_r + 1) z, z_q - z_r z) = (q, 1)(z_q - z_r z) \\
&= (q, 1) \, z \frac{\zeta^c}{q}.
\end{aligned} \tag{113}$$

Proposition 7.1 implies:

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

□

**Proof of Elasticity relations (83) 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

$$B_{\boldsymbol{c}}(\boldsymbol{c}, \boldsymbol{q}) = (1, -q). \tag{114}$$

We use (43)

$$\begin{aligned}\boldsymbol{\tau}^b &= B_{\boldsymbol{c}}(\boldsymbol{c},\boldsymbol{q}) - \frac{B_{\boldsymbol{c}}(\boldsymbol{c},\boldsymbol{q}^s)}{B_{\boldsymbol{c}}(\boldsymbol{c},\boldsymbol{q}^s)\cdot \boldsymbol{c}_r(\boldsymbol{q},r)} \\ &= (1,-q) - \frac{(1,-q^s)}{(1,-q^s)\cdot(qz_r+1,z_r)} \text{ as } c(q,r) = qz(q,r)+r \text{ gives } c_r = qz_r+1. \\ &= (1,-q) - \frac{(1,-q^s)}{1+(q-q^s)z_r} \\ \boldsymbol{\tau}^b &= (1,-q) - \frac{(1,-q^s)}{1+(q-q^s)\frac{\eta}{q}}. \end{aligned} \tag{115}$$

Next, recall (112),

$$\begin{aligned}\tau^b &= \boldsymbol{\tau}^b(\boldsymbol{q},r)\cdot(q,1) \\ &= \left[(1,-q) - \frac{(1,-q^s)}{1+(q-q^s)\frac{\eta}{q}}\right]\cdot(q,1) = \frac{q^s-q}{1+(q-q^s)\frac{\eta}{q}} \\ &= \frac{\tau-\tau^s}{1-(\tau-\tau^s)\frac{\eta}{q}}. \end{aligned} \tag{116}$$

using $q = 1-\tau$, $q^s = 1-\tau^s$. Thus we have proven (84).

Next, we calculate $\zeta^c$. We call $\boldsymbol{e} = (0,1)$ the vector singling earnings on the vector $\boldsymbol{c} = (c,z)$. We apply (42) with $p_j = q$, the price of earnings. We have:

$$\begin{aligned}\boldsymbol{e}\cdot\boldsymbol{S}_j^H &= \boldsymbol{e}\cdot\boldsymbol{S}^r(\boldsymbol{p},r)\cdot\boldsymbol{p}_{p_j}^s(\boldsymbol{p},r) = z_q^r\frac{\partial q^s}{\partial q} = \frac{z}{q}\zeta^{c,r}m_{zz} \\ \boldsymbol{e}\cdot\boldsymbol{S}_j^H &= \frac{z}{q}\zeta^{c,r}m_{zz}. \end{aligned} \tag{117}$$

Next, using the notation $D_j$ of Proposition 7.1,

$$\begin{aligned} D_j &= -\boldsymbol{\tau}^b\cdot\boldsymbol{S}_j^H \text{ by (36)} \\ &= [B_{\boldsymbol{c}}(\boldsymbol{p},\boldsymbol{c}) - B_{\boldsymbol{c}}(\boldsymbol{p}^s,\boldsymbol{c})]\cdot\boldsymbol{S}_j^H \text{ by (44)} \\ &= [(1,q)-(1,q^s)]\cdot\boldsymbol{S}_j^H \text{ by (114)} \\ &= (q-q^s)\,\boldsymbol{e}\cdot\boldsymbol{S}_j^H \text{ as } \boldsymbol{e} = (0,1) \\ &= (q-q^s)\frac{z}{q}\zeta^{c,r}m_{zz} \text{ by (117)}. \end{aligned}$$

We record:

$$D_j = (q-q^s)\frac{z}{q}\zeta^{c,r}m_{zz}. \tag{118}$$

Next, we apply (40): $\boldsymbol{S}_j^C = \boldsymbol{S}_j^H + \boldsymbol{c}_r D_j$, which implies:

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

Now, as $\zeta^c = \frac{q}{z}\boldsymbol{e} \cdot \boldsymbol{S}_j^C$ is the compensated earnings elasticity (see e.g. (113)):

$$\begin{aligned}
\zeta^c &= \frac{q}{z}\boldsymbol{e} \cdot \boldsymbol{S}_j^C \\
\zeta^c &= \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{c,r} m_{zz}.
\end{aligned} \tag{119}$$

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

$$\zeta^c_{Q_{z^*}} = \left(1 - \eta\frac{\tau - \tau^s}{q}\right)\zeta^{c,r} m_{zz^*}. \tag{120}$$

Hence, we have proven (83).

**Proof of (85): Decision vs. Experienced utility model** The agent's optimization gives $qu_c^s + u_z^s = 0$. Equation (112) gives:

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

**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 Proposition 10.1** Let us now solve for the optimal $J$, which ensures $\frac{\partial L}{\partial Q_{z^*}} = 0$. We can write

$$-\frac{\partial L}{\partial \tau_{z^*}} = J(z^*) - \int_{z^*}^{\infty} a(z)\,dz - b(z^*) + \int_{z^*}^{\infty} J(z)\,\rho(z)\,dz. \tag{121}$$

We use the notations

$$\rho(z) = \frac{\eta}{\zeta^c}\frac{1}{z}, \tag{122}$$

$$a(z) = (1 - g(z))\, h(z) - \rho g(z)\, z \left(-\tau^b(z)\frac{\zeta^c}{q(z)}\right)(h^*(z) - h(z)), \tag{123}$$

$$b(z) = -\overline{\overline{F}}(z_*). \tag{124}$$

$a(z)$ 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 $\delta q_{z^*}$, coming from the distortions from perceptions

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

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

We note that

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

$$a = (1 - \gamma)\, h + \rho J. \tag{125}$$

We also have

$$J(z^*) = \int_{z^*}^\infty a(z)\, dz + b(z^*) - \int_{z^*}^\infty J(z)\,\rho(z)\, dz,$$

$$\dot{J} = -a + \dot{b} + J\rho,$$

$$\frac{d}{dz}\left[J(z)\, e^{-\int_0^z \rho(s)ds}\right] = e^{-\int_0^z \rho(s)ds}\left(-a(z) + \dot{b}(z)\right),$$

$$J(z)\, e^{-\int_0^z \rho(s)ds} = C + \int_z^\infty e^{-\int_0^{z'} \rho(s)ds}\left(a(z') - \dot{b}(z')\right)dz',$$

$$J(z) = \int_z^\infty e^{-\int_z^{z'} \rho(s)ds}\left(a(z') - \dot{b}(z')\right)dz'. \tag{126}$$

Integrating by parts, we get

$$\int_z^\infty e^{-\int_z^{z'}\rho(s)ds}\dot{b}\left(z'\right)dz' = \left[e^{-\int_z^{z'}\rho(s)ds}b\left(z'\right)\right]_z^\infty + \int_z^\infty e^{-\int_z^{z'}\rho(s)ds}\rho\left(z'\right)b\left(z'\right)dz'$$

$$= -b\left(z\right) + \int_z^\infty e^{-\int_z^{z'}\rho(s)ds}\rho\left(z'\right)b\left(z'\right)dz', \tag{127}$$

$$J\left(z\right) = b\left(z\right) + \int_z^\infty e^{-\int_z^{z'}\rho(s)ds}\left(a\left(z'\right) - \rho\left(z'\right)b\left(z'\right)\right)dz'. \tag{128}$$

We can rewrite this as

$$J\left(z^*\right) = -\overline{\overline{F}}\left(z^*\right) + \int_{z^*}^\infty e^{-\int_{z^*}^z\rho(s)ds}\left(a\left(z\right) + \rho\overline{\overline{F}}\left(z\right)\right)dz \tag{129}$$

$$= -z^*g\left(z^*\right)\left(-\tau^b\left(z^*\right)\frac{\zeta^c\left(z^*\right)}{q(z^*)}\right)h^*\left(z^*\right) - F\left(z^*\right)$$

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

Using

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

and rearranging gives

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

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

which can be rewritten to get the announced formula

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

$$= \frac{1}{\zeta^c}\frac{1-H\left(z^*\right)}{z^*h^*\left(z^*\right)}\int_{z^*}^\infty e^{-\int_{z^*}^z\rho(s)ds}\left(1-g\left(z\right)-\eta\frac{\tilde{\tau}^b\left(z\right)}{q\left(z\right)}\right)\frac{h\left(z\right)}{1-H\left(z^*\right)}dz.$$

□

**Proof of Proposition 10.2** We use (11.1):

$$\begin{aligned}
\frac{\partial L}{\partial \tau_{z^*}} &= \int_{z^*}^{\infty} (1-\gamma(z))\, h(z)\, dz - \zeta^c(z^*)\frac{T'(z^*)-\widetilde{\tau}^b(z^*)}{1-T'(z^*)} z^* h^*(z^*) - \int_0^{\infty} \zeta^c_{Q_{z^*}}(z)\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)} z h^*(z)\, dz \\
&= \int_{z^*}^{\infty} \left(1 - g(z) - \eta(z)\frac{T'(z)}{1-T'(z)}\right) h(z)\, dz - \zeta^c(z^*)\frac{T'(z^*)-\widetilde{\tau}^b(z^*)}{1-T'(z^*)} z^* h^*(z^*) \\
&- \int_0^{\infty} \zeta^c_{Q_{z^*}}(z)\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)} z h^*(z)\, dz \text{ using (86)} \\
&= (1-H(z_*))\left(1-\overline{g}-\overline{\eta}\frac{\overline{\tau}}{1-\overline{\tau}}\right) - \overline{\zeta}^c\frac{\overline{\tau}-\overline{g\tau}^b}{1-\overline{\tau}} z^* h^*(z^*) - \int_0^{\infty} \zeta^c_{Q_{z^*}}(z)\frac{T'(z)-\tilde{\tau}^b(z)}{1-T'(z)} z h^*(z)\, dz.
\end{aligned}$$

Recall that

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

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:

$$\frac{\overline{\tau}}{1-\overline{\tau}} = \frac{1-\overline{g}-\beta+\zeta^c\pi\overline{g}\frac{\overline{\tau}^b}{1-\overline{\tau}}}{\zeta^c\pi+\overline{\eta}}, \tag{130}$$

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:

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

□

**Proof of Lemma 11.4** We have

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

so

$$\begin{aligned} \delta r &= r'\left(z\right)\delta z + \delta r_{|\text{ constant } z} = -zq'\left(z\right)\delta z + \left(\delta K - z\delta q_z\right) \\ &= -zR''\delta z + \delta K - z\delta q_z. \end{aligned}$$

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

so

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

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

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

□

**Proof of Lemma 11.5** Observe that

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

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

$$zR'' = q\frac{h^* - h}{\zeta^c h^*}. \tag{131}$$

We have

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

We also have

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

Using Lemma 11.4, we can rewrite this as

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

Using equation (131) and Lemma 11.4, we can rewrite this as

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

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 11.1** We use the following notations:

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

We consider a change $\delta q_{z^*}$ at $z^*$. This leads to a lump-sum change $\delta K = 1_{z>z^*}\delta q_{z^*}$. Hence, Lemma 11.5 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

$$\begin{aligned}
\delta L &= \int_0^\infty \delta L(z) h(z)\, dz, \\
\frac{\delta L}{\delta q_{z^*}} &= \int_{z^*}^\infty (\gamma(z) - 1) h(z)\, dz + \int_0^\infty \left[ \frac{T'(z) - \widetilde{\tau}^b(z)}{q} \overline{\overline{\zeta}}^c_{Q_{z^*}} \right] \frac{h^*}{h} z h(z)\, dz, \\
\frac{\partial L}{\partial \tau_{z^*}} &= -\overline{\overline{\overline{F}}}(z_*) + \int_{z^*}^\infty (1 - \gamma(z)) h(z)\, dz.
\end{aligned} \tag{132}$$

We also have

$$\frac{\partial L}{\partial \tau_{z^*}} \equiv -\frac{\partial L}{\partial Q_{z^*}} = -\overline{\overline{\overline{F}}}(z^*) + \int_{z^*}^\infty (1 - \gamma(z)) h(z)\, dz. \tag{133}$$

□

**Proof of Lemma 11.6** Using Lemma 11.5, applied to a change $\delta r_0$ 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.□

# 12 Basic Behavioral Consumer Theory with Linear Budget Constraints

## 12.1 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:

$$v_{p_j} + v_w c^j = 0. \tag{134}$$

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

$$e_{p_i}(\boldsymbol{p}, u) = h^i, \tag{135}$$

and differentiating once more gives:

$$e_{p_i p_j} = h^i_{p_j}(\boldsymbol{p}, u) = S_{ij}. \tag{136}$$

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.

$$c^i_{p_j} + c^i_w c_j = h^i_{p_j} = S_{ij}. \tag{137}$$

## 12.2 Behavioral version with perceived prices

The sparse max demand

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

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

$$\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w'(\boldsymbol{p}, \boldsymbol{p}^s, w)), \tag{138}$$

where perceived budget $w'$ satisfies:

$$\boldsymbol{p}\cdot\boldsymbol{c}^r\left(\boldsymbol{p}^s,w'\left(\boldsymbol{p},\boldsymbol{p}^s,w\right)\right)=w. \tag{139}$$

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

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

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

We call

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

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

**Proposition 12.1** (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

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

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

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

□

**Proposition 12.2** *(*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 $v_{p_j^s}^s$, observe that:

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

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

**Proof**. 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. $p_j$ gives:

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

by the behavioral Shepard's lemma (Proposition 12.1).

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

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

again by the behavioral Shepard's lemma (Proposition 12.1).□

**Proposition 12.3** *(Marshallian demand) Given the consumption function $\boldsymbol{c}\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right)$, we have:*

$$\boldsymbol{c}_{p_j} = -\boldsymbol{c}_w c_j \tag{140}$$

$$\boldsymbol{c}_{p_j^s} = \boldsymbol{S}_j^r + \boldsymbol{c}_w D_j^s =: S \tag{141}$$

$$= \boldsymbol{S}_j^r + \boldsymbol{c}_w \left[\left(\boldsymbol{p}^s - \boldsymbol{p}\right) \cdot \boldsymbol{S}_j^r\right] \tag{142}$$

$$= \left(1 + \boldsymbol{c}_w \left(\boldsymbol{p}^s - \boldsymbol{p}\right)'\right) \boldsymbol{S}_j^r. \tag{143}$$

*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,* $\boldsymbol{c}_w = \frac{\mathbf{1}}{\boldsymbol{p} \cdot \boldsymbol{c}_{w'}^r \left(\boldsymbol{p}^s, w'\right)} \boldsymbol{c}_{w'}^r = \frac{v_w}{\lambda} \boldsymbol{c}_{w'}^r$.

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. $dp^s = \chi\left(p - p^s\right)$, $\chi > 0$. Then,

$$d\boldsymbol{c} = \boldsymbol{c}_{p^s} dp^s = S dp^s + \boldsymbol{c}_w dE$$
$$dE = \left[\left(\boldsymbol{p}^s - \boldsymbol{p}\right) \cdot \boldsymbol{S}^r \cdot dp^s\right] \geq 0.$$

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 $\boldsymbol{c}\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right) = \boldsymbol{h}^r\left(\boldsymbol{p}^s, v\left(\boldsymbol{p}, \boldsymbol{p}^s, w\right)\right)$, which implies

$$\boldsymbol{c}_w = \boldsymbol{h}_u^r v_w, \qquad \boldsymbol{c}_{p_j} = \boldsymbol{h}_u^r v_{p_j}. \tag{144}$$

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:

$$\begin{aligned}\boldsymbol{c}_{p_j^s}(\boldsymbol{p},\boldsymbol{p}^s,w)&=\boldsymbol{h}_{p_j^s}+\boldsymbol{h}_u v_{p_j^s}\\&=\boldsymbol{S}_j^r+\boldsymbol{c}_w\frac{v_{p_j^s}}{v_w}\text{ using }\boldsymbol{c}_w=\boldsymbol{h}_u^r v_w\\&=\boldsymbol{S}_j^r+\boldsymbol{c}_w D_j^s\text{ using Proposition 12.2.}\end{aligned}$$

We have (139): so $\boldsymbol{p}\cdot\boldsymbol{c}_{w'}^r\frac{\partial w'}{\partial w}=1$, and $\frac{\partial w'}{\partial w}(\boldsymbol{p},\boldsymbol{p}^s,w)=\frac{1}{\boldsymbol{p}\cdot\boldsymbol{c}_{w'}^r(\boldsymbol{p}^s,w')}$. So $\boldsymbol{c}(\boldsymbol{p},\boldsymbol{p}^s,w)=\boldsymbol{c}^r(\boldsymbol{p}^s,w'(\boldsymbol{p},\boldsymbol{p}^s,w))$, we have $\boldsymbol{c}_w^s=\boldsymbol{c}_{w'}^r\frac{\partial w'}{\partial w}$.□

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 12.4** *(Envelope theorem, modified) Call $\lambda$ 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 $\boldsymbol{p}\cdot\boldsymbol{c}_{w'}^r\frac{\partial w'}{\partial w}=1$, and

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

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

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

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

$$v_w=u'(\boldsymbol{c}^s)\cdot\boldsymbol{c}_w=\lambda p^s\cdot\boldsymbol{c}_w=\lambda p^s\cdot\left(\frac{\boldsymbol{c}_{w'}^r(\boldsymbol{p}^s,w')}{\boldsymbol{p}\cdot\boldsymbol{c}_{w'}^r(\boldsymbol{p}^s,w')}\right)=\lambda\frac{p^s\cdot\boldsymbol{c}_{w'}^r(\boldsymbol{p}^s,w')}{\boldsymbol{p}\cdot\boldsymbol{c}_{w'}^r(\boldsymbol{p}^s,w')}$$

$$v_w=\frac{\boldsymbol{\lambda}}{\boldsymbol{p}\cdot\boldsymbol{c}_{w'}^r(\boldsymbol{p}^s,w')}.$$

□

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

$$v_{p_j^s}=u_{\boldsymbol{c}}\boldsymbol{c}_{p_j^s}=\lambda\boldsymbol{p}^s(1+\boldsymbol{c}_w p')\boldsymbol{S}_j^r=\lambda\boldsymbol{p}^s\boldsymbol{c}_w p'\boldsymbol{S}_j^r=v_w p'\boldsymbol{S}_j^r=v_w D_j^s.$$

**Proposition 12.5** *(Expenditure function – second derivatives) Given* $e^s(\boldsymbol{p}, \boldsymbol{p}^s, u) = \boldsymbol{p} \cdot \boldsymbol{h}(\boldsymbol{p}^s, u)$*, we have*

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

*The first derivatives of the expenditure functions were calculated in Proposition 12.1.*

**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 12.1.

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

Differentiating more,

$$e^s_{p_i p_j} = 0 \tag{145}$$

$$e^s_{p_i p^s_j} = h^i_{p^s_j}(\boldsymbol{p}^s, u) = S^r_{ij}, \tag{146}$$

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

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

□

## 12.3 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 $\boldsymbol{p}^s(\boldsymbol{p}, w)$.

**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:*

$$\boldsymbol{p}^s(\boldsymbol{p}, w) = \frac{u_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p}, w))}{v_w(\boldsymbol{p}, w)}. \tag{147}$$

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

$$\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{c}^s(\boldsymbol{p}, \boldsymbol{p}^s(\boldsymbol{p}, w), w). \tag{148}$$

**Proof**. The demand of an agent misperceiving prices, $\boldsymbol{c}^s(\boldsymbol{p},\boldsymbol{p}^s,w)$, is characterized by $u_{\boldsymbol{c}}(\boldsymbol{c}^s(\boldsymbol{p},\boldsymbol{p}^s,w)) = \lambda\boldsymbol{p}^s$ for some $\lambda$, 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 $\boldsymbol{p}^s = \boldsymbol{p}^s(\boldsymbol{p},w)$. 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 $\lambda$ not necessarily the lowest value possible.□

We note that for any $\boldsymbol{p}^s(\boldsymbol{p},w) = ku_{\boldsymbol{c}}(\boldsymbol{c}(\boldsymbol{p},w))$ 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 $m_i \neq m_j$.

# 13 Complements on Basic Consumer Theory with Nonlinear Budget Constraints

Here we give complements to Section 7.

## 13.1 Rational Agent

Primal is: $v(\boldsymbol{p},w) = \max_{\boldsymbol{c}} u(\boldsymbol{c})$ s.t. $B(\boldsymbol{p},\boldsymbol{c}) \leq w$ 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. $u(c) \geq \overline{u}$ 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: $B_{\boldsymbol{c}}\boldsymbol{h}_{p_i}=B_{\boldsymbol{c}}\boldsymbol{c}_{p_i}+B_{\boldsymbol{c}}\boldsymbol{c}_wB_{\boldsymbol{p}_i}=-B_{p_i}+B_{p_i}=0$,

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

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

## 13.2 Misperceiving Agent

We now study

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

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 $\lambda$, 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 $\boldsymbol{p}^s$

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

Define the dual expenditure function

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

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 13.1** (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 $B_{\boldsymbol{c}}^{s}\boldsymbol{h}_{\boldsymbol{p}_{j}^{s}}^{r}=0$. □

**Proposition 13.2** (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.* $\frac{v_{p_{j}^{s}}^{s}}{v_{w}^{s}}=\sum_{i}\left(\boldsymbol{B}_{i}^{s}-\boldsymbol{B}_{i}\right)\boldsymbol{S}_{ij}^{r}$.

**Proof**. 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 13.1).

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 13.1).□

**Proposition 13.3** *(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:

$$\begin{aligned}
\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p}^s, w) &= \boldsymbol{h}_{\boldsymbol{p}_j^s} + \boldsymbol{h}_u v_{\boldsymbol{p}_j^s} \\
&= \boldsymbol{S}_j^r + \boldsymbol{c}_w \frac{v_{p_j^s}}{v_w} \text{ using } \boldsymbol{c}_w = \boldsymbol{h}_u v_w \\
&= \boldsymbol{S}_j^r + \boldsymbol{c}_w \left[ (B_{\boldsymbol{c}}^s - B_{\boldsymbol{c}}) \cdot \boldsymbol{S}_j^r \right] \text{ using Proposition 13.2.}
\end{aligned}$$

□

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 13.4** *(Envelope theorem, modified) Call $\lambda$ 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

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

□

## 13.3 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 maximizes instead $\text{smax}_{\boldsymbol{c}|\boldsymbol{p}^s} u^s(\boldsymbol{c})$ 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:

$$\boldsymbol{c}(\boldsymbol{p},w) = \boldsymbol{h}^{r,s}(\boldsymbol{p}^s(\boldsymbol{p},w), v^s(\boldsymbol{p},w)). \tag{162}$$

**Proposition 13.5** (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*

$$\boldsymbol{S}^s_j(\boldsymbol{p},w) = \boldsymbol{S}^{r,s}(\boldsymbol{p},w) \cdot \boldsymbol{p}^s_{p_j}(\boldsymbol{p},w). \tag{163}$$

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

$$\begin{aligned}
\boldsymbol{S}^C_j(\boldsymbol{p},w) &= \boldsymbol{S}^s_j(\boldsymbol{p},w) + \boldsymbol{c}_w\left(\frac{v^s_{p_j}}{v^s_w} + B_{p_j}\right)\\
\boldsymbol{S}^H_j(\boldsymbol{p},w) &= \boldsymbol{S}^s_j + \boldsymbol{c}_w\left(\frac{v^s_{p_j}}{v^s_w} - \frac{v_{p_j}}{v_w}\right)\\
\boldsymbol{S}^s_j(\boldsymbol{p},w) &= \boldsymbol{c}_{p_j} - \boldsymbol{c}_w\frac{v^s_{p_j}}{v^s_w}.
\end{aligned}$$

*We can write*

$$-D_j = \overline{\boldsymbol{\tau}}^b \cdot \boldsymbol{S}^s_j,$$

*with:*

$$\overline{\boldsymbol{\tau}}^b = (B_{\boldsymbol{c}}(\boldsymbol{p},\boldsymbol{c}) - B_{\boldsymbol{c}}(\boldsymbol{p}^s,\boldsymbol{c})) + \left(\frac{u^s_{\boldsymbol{c}}}{v^s_w} - \frac{u_{\boldsymbol{c}}}{v_w}\right). \tag{164}$$

*Finally, $B_{\boldsymbol{c}}(\boldsymbol{p}^s,\boldsymbol{c}) \cdot \boldsymbol{S}^s_j = 0$.*

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

**Proof**.

Use $\boldsymbol{c}(\boldsymbol{p}, w) = \boldsymbol{h}^{r,s}\left(\boldsymbol{p}^s, v^s\left(\boldsymbol{p}, w\right)\right)$:

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

Using (32) and (33) gives:

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

We have

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

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,

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

□