**Hicksian aggregation across consumers**

**or (the rebound problem after derivation went wrong)**

**Felipe Del Canto, 9 de Septiembre de 2019**

# 1. Do you remember this?

Consider a two period economy composed by a continuum of agents that consume $n+1$ goods: an essential good $z$ (i.e., water) and $n$ different goods grouped in the vector $\mathbf{x}$. Good $z$ is valued at price $q \in \mathbb{R}_{++}$ and the $\mathbf{x}$ goods are valued at prices $\mathbf{p} \in \mathbb{R}^n_{++}$. In what follows and for simplicity of notation, the subindices $t$ for all variables will be omitted unless needed. Each individual in this setting is identified with a pair $(y, \alpha)$, where $y \in [m, M]$ is her income and $\alpha \in (0,1)$ determines the form of her pseudo-Cobb-Douglas (PCD)[1] utility function, that is,

$$u_\alpha(\mathbf{x}, z) := u_{(y,\alpha)}(\mathbf{x}, z) = \sum_{j=1}^{n} x_j^\alpha z^{1-\alpha} = \left( \sum_{j=1}^{n} x_j^\alpha \right) z^{1-\alpha}.$$

Consumers choose how much of the $\mathbf{x}$ goods to consume in each period according to

$$\begin{aligned} \mathbf{x}(\mathbf{p}, q, y) := \arg\max_{\mathbf{x},z} \quad & u_\alpha(\mathbf{x}, z), \\ \text{s.a.} \quad & \mathbf{p}\mathbf{x} + qz = y. \end{aligned} \tag{1}$$

Individuals in this economy (that is, pairs $(y, \alpha)$) are distributed according to $F$ (with marginal distributions $F_y$ and $D_\alpha$) with support $S := [m, M] \times (0,1)$. In this setting, is it possible to aggregate consumption of the $\mathbf{x}$ goods into a single good $X$ given by:

$$g_\alpha(\mathbf{x}) = \left( \sum_{j=1}^{n} x_j^\alpha \right)^{1/\alpha}.$$

In that case, we have

$$U(g_\alpha(\mathbf{x}), z) := g_\alpha(\mathbf{x})^\alpha z^{1-\alpha},$$

[1] The choice of this name will be clear shortly.

which is the usual Cobb-Douglas utility function. Hence, by defining $\varepsilon := (1-\alpha)^{-1}$ and

$$P_\alpha(\mathbf{p}) := \left( \sum_{j=1}^{n} p_j^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}},$$

we have that the category demand

$$\begin{aligned} X(P(\mathbf{p}), q, y) := \arg\max_{X,z} \quad & U(X, z) \\ \text{s.a.} \quad & P_\alpha(\mathbf{p})X + qz = y \end{aligned} \tag{2}$$

satisfies

$$X(P_\alpha(\mathbf{p}), q, y) = g_\alpha\big(\mathbf{x}(\mathbf{p}, q, y)\big). \tag{3}$$

Note that the functions $g_\alpha$ and $P_\alpha$ depend explicitly on the value of $\alpha$. Thus, in order to correctly aggregate the goods $\mathbf{x}$ a different aggregator and price index for every person is needed.

In this context, consider now the following situation. An investigator at time $t = 0$ has data available on disaggregate consumption of the $\mathbf{x}$ and $z$ goods, income $y$ and prices $(\mathbf{p}, q)$. For simplicity she assumes that all agents have the same preferences, which means that all differences in consumption are accounted by differences in income. The last assumption also implies that in order to obtain the category demand of each agent and the price index, she must set a parameter $\overline{\alpha}$ which in principle could be chosen

- based on previous evidence (i.e., the expenditure fraction of category $X$) or,
- optimally in order to minimize discrepancies between actual and predicted category demands in $t = 0$.

For the second method, the choice of $\overline{\alpha}$ should minimize the expectation of

$$\mathcal{D}(\alpha) = \big| g_\alpha(\mathbf{x}(\mathbf{p}, q, y)) - g_{\overline{\alpha}}(\mathbf{x}(\mathbf{p}, q, y) \big|,$$

which by (3) is equivalent to

$$\mathcal{D}(\alpha) = \big| X(P_\alpha(\mathbf{p}), q, y) - X(P_{\overline{\alpha}}(\mathbf{p}), q, y) \big|,$$

and this difference can be easily computed by taking advantage of the structure of problem (2),

$$\mathcal{D}(\alpha) = y \left| \frac{\alpha}{P_\alpha(\mathbf{p})} - \frac{\overline{\alpha}}{P_{\overline{\alpha}}(\mathbf{p})} \right|.$$

Thus, $\overline{\alpha}$ should satisfy

$$\overline{\alpha} = \arg\min_{\hat{\alpha}\in[0,1]} \int_S y \left| \frac{\alpha}{P_\alpha(\mathbf{p})} - \frac{\overline{\alpha}}{P_{\overline{\alpha}}(\mathbf{p})} \right| dF(y,\alpha).$$

If, as I assumed in previous settings, $\alpha$ and $y$ are independent random variables, then

$$\overline{\alpha} = \arg\min_{\hat{\alpha}\in[0,1]} \int_0^1 \left| \frac{\alpha}{P_\alpha(\mathbf{p})} - \frac{\overline{\alpha}}{P_{\overline{\alpha}}(\mathbf{p})} \right| dF_\alpha(\alpha). \tag{4}$$

Intuitively, $\overline{\alpha} \in (0,1)$ because $F_\alpha$ has support over the open interval and choosing a parameter away from that support should not be optimal.[2] As a final comment, observe that choosing $\overline{\alpha}$ needs knowledge about the distribution $F_\alpha$, which may not be available.

The objective of our researcher is to estimate category demands in $t = 1$. Whichever was her choice of $\overline{\alpha}$ she intends to use this parameter in the second period to

[2] A formal argument for this claim would be nice.