Integrability and Generalized Separability

Thibault Fally*
UC Berkeley ARE

December 2017

Abstract

This paper examines demand systems where the demand for a good depends only on its own price, consumer income, and a single aggregator synthesizing information on all other prices. As indicated by Gorman (1972), symmetry of the Slutsky substitution terms implies that such demand can take only one of two simple forms. Conversely, here we show that only weak conditions ensure that such demand systems are integrable, i.e. can be derived from the maximization of a utility function. This paper further studies useful properties, special cases and applications of such demand systems.

Keywords: Separable demand, Single aggregator, Integrability, Recoverability, Non-homothetic preferences.
JEL Classification: D11, D40, L13

*UC Berkeley, Agricultural and Resource Economics Department, NBER and CEPR. Many thanks to Philippe Aghion, Dave Donaldson, Federico Etro, Ben Faber, Ethan Ligon, Kiminori Matsuyama, Marti Mestieri, Andres Rodriguez-Clare, Ina Simonovska and Brian Wright for helpful comments and suggestions. Contact: fally@berkeley.edu

1 Introduction

The integrability problem, recognizing demand systems that can be derived from utility functions and rationalized, has long been a central issue in economic theory with earliest contributions from Antonelli (1886) and numerous applications to various fields, including micro and macroeconomics, econometrics, industrial organization and international trade. Theorists have provided broad sufficient and necessary conditions for demand patterns to be integrable, notably Hurwicz and Uzawa (1971) who provide conditions based on the Slutsky substitution matrix, which must be symmetric and semi-definite negative for all prices and income levels.

Applied theorists and practitioners have often focused on less general cases for the sake of simplicity, tractability and empirical applications. Typical demand used in practice depends only on a few variables, namely consumer income, a good's own price, and a single aggregator (scalar) that is itself a function of the vector of prices and income. Such aggregator can be, for instance, a price index or a Lagrange multiplier. Many examples fall into that category. Demand systems derived from directly-additive utility have attracted particular attention given their tractability and ease of interpretation, where marginal utility provides such an aggregator. Indirectly-additive preferences (for which the indirect utility is additively separable), implicitly-additive preferences and quasi-linear preferences are alternative options that have often been considered.

Here we study a class of demand systems that includes all these cases and follows what Pollak (1972) characterized as "generalized separability", where demand for good i satisfies:

$$q_i = q_i(p_i/w, \Lambda) \tag{1}$$

where w refers to consumer income, p_i the price of good i, and where Λ is a scalar (aggregator) function of all prices and income. Without providing a complete proof, Gorman (1972, 1995) indicates that such demand system can take either of two main forms[1] if we impose the Slutsky matrix to be symmetric:

$$q_i = \frac{D_i(F(\Lambda)p_i/w)}{H(\Lambda)} \qquad \text{or:} \qquad q_i = A_i(\Lambda)(p_i/w)^{-\sigma(\Lambda)} \tag{2}$$

where, in both cases, Λ is a scalar that is adjusted so that the budget constraint is satisfied, and can thus be defined is an implicit function of prices and income. As little is known about the general forms of these demand systems, they have not been used in the applied literature in spite of their usefulness.[2]

[1] As well as other cases that can be ruled out under additional restrictions on price sensitivity.

[2] With very few exceptions, including Bertoletti and Etro (2017a).

The objectives of this paper are threefold. First, I provide conditions under which demand with generalized separability have to take these two forms (i.e. a formal statement of Gorman's claims mentioned above). These functional forms however do not imply that these demand systems are well defined and integrable (counter-examples are provided).

The second and main contribution is to show that any demand function that takes the form in equation (2) can be integrated under weak restrictions. These restrictions ensure not only that the Slutsky substitution matrix be symmetric but also semi-definite negative, or equivalently a quasi-concave and well defined utility function.[3] In the first case where demand satisfies $q_i = D_i(F(\Lambda)p_i/w)/H(\Lambda)$, integrability is guaranteed if D_i is monotonically decreasing in p_i, and demand q_i is decreasing in Λ. In the second case with common price elasticities across goods, i.e. where $q_i = A_i(\Lambda)(p_i/w)^{-\sigma(\Lambda)}$, we need the demand shifters $A_i(\Lambda)$ to increase fast enough with Λ to ensure that the associated utility is well defined. In that case, there is a one-to-one mapping between Λ and utility. Notice that the price elasticity $\sigma(\Lambda)$ does not have to remain constant or monotonic across indifference curves; it can increase or decrease with Λ, i.e. indifference curves can become flatter or more convex as income goes up.

The third objective is to illustrate the additional properties of such demand functions and practical applications of these results. The first case corresponds to directly-additive utility when $H(\Lambda)$ is constant and indirectly-additive utility when $F(\Lambda)$ is constant. It is however very useful to extend to more flexible demand and price shifters (H and F), while keeping flexible price effects (through D_i). This also generalizes the results of Matsuyama and Ushchev (2017) who focus on the homothetic case where $F(\Lambda) = 1/H(\Lambda) = \Lambda$. In this more general demand system, the elasticity of the aggregator Λ to prices and income depends tightly on these demand shifters, which in turn influence the price and income elasticity of demand. The second case features Allen-Uzawa substitution elasticities that do not vary across goods but may vary with the demand aggregator Λ, and generalizes implicitly-additive utility functions previously defined in the literature (Hanoch 1975, Comin et al 2016) that impose a constant elasticity of substitution $\sigma(\Lambda) = \sigma$. This case allows for more flexible income patterns, but less flexible price effects.[4]

The demand systems in both cases can have various applications. They are particularly useful in the case of monopolistic competition. In the limit where firms have small market shares, they choose their price by taking as given other prices and quantities. It is then

[3] This utility representation provided here is derived from Gorman (1987). The latter is however more restrictive and Gorman does not show that such representation has the required properties of a utility function.

[4] Note that implicitly-additive utility depends on a single aggregator only in the case where the Allen-Uzawa elasticity of substitution is common across goods. In other cases, such as Kimball (1995), two aggregators are needed. See also Preckel et al. (2005) for another form of implicitly additive utility and generalization of CES preferences.

practical to have a single industry-wide indicator Λ that uniquely determines the locus of the demand curve for a variety i w.r.t its own price p_i.[5] Such a price aggregator Λ can be used as an index of competition, and may shift downward or upward along demand curves as income and competition grows, with flexible implications for markups. The results also extend to inverse demand functions with a single aggregator in quantities.

This leads to a generalization of various demand systems, such as the constant relative income elasticities (CREI) used in Fieler (2011) and Caron et al. (2014). These demand functions encompass most examples from Mrázová and Neary (2013), e.g. bi-power and inverse bi-power demand functions, or Weyl and Fabinger (2013), e.g. Bulow-Pfleiderer demand.

In cases where demand D_i for a particular good goes to zero, we obtain a choke price the depend on wealth and the price aggregator. The demand functions considered here can be used to generate choke prices that depend on income more flexibly than commonly used preferences in macroeconomics and international trade. In particular it generalizes Bertoletti and Etro (2017b) and Bertoletti et al. (2016) where choke prices are proportional to income (with elasticity one). An application illustrates the gains from trade with variable markups and non-homothetic preferences, following Arkolakis et al. (2015). Here I show that the more flexible demand functions described here (more flexible demand shifter H) allow us to have a continuum of cases with various degrees of feedback from markup adjustments to consumer welfare, which matter for the quantification of the gains from trade.

Recent work (such as Handbury, 2013, and Faber and Fally, 2017) show that price elasticities vary significantly with income, and model income effects in the elasticity of substitution by relying on a numeraire good. With the second type of demand functions described here, we can generate observed relationships between income and the elasticity of substitution without relying on a numeraire but still having Allen-Uzawa elasticities of substitution $\sigma(\Lambda)$ that are common across goods.

The paper is related to a vast literature studying functional forms restrictions of utility and demand systems. Ligon (2016) focuses on cases where the aggregator Λ corresponds to the Lagrange multiplier associated with the budget constraint, and shows that $\Lambda-$separability implies directly-additive utility with even more specific functional forms. Nocke and Schutz (2017) study the ("quasi-") integrability of quasi-linear demand systems, i.e. without income effects. The discussion on the existence of aggregators also mirrors the restrictions associated with the rank of a demand system (Gorman, 1981; Lewbel, 1991, 2010; LaFrance and Pope, 2006; Lewbel and Pendakur, 2009). The rank of a demand system corresponds to the number of vectors and homothetic price aggregators needed to recover Engel curves (see Lewbel, 1991).

[5] Bertoletti and Etro (2017a) contemporaneous work formalizes this insight and covers this demand system as an example.

Here, the single aggregator Λ is generally not homogeneous one in prices (it also depends on income) and the two demand systems studied here do not have restrictions in terms of rank. Finally, Blackorby et al. (1978) study functional forms implied by various definitions of separability, and find that same functional structure as case 2 is obtained when imposing stronger forms of separability that imply equality among Allen-Uzawa elasticities of substitution.

2 Generalized separability

2.1 Functional forms

Additively-separable utility allows us to obtain demand as a simple function of a good's own price p_i and a single aggregator, the Lagrange multiplier. While practical, direct additive separability puts strong constraints on the structure of demand, such as Pigou's law on the relationship between price elasticity and income elasticity, with the adverse consequence that utility with constant elasticity of substitution (CES) is the only directly-separable utility function that is homothetic.

In an attempt to generalize the concept of separability, Gorman (1972) and Pollak (1972) define generalized separability as demand that would take the form:

$$q_i = q_i(p_i/w, \Lambda) \tag{3}$$

where q_i refers to demand for good i (quantity) and Λ is implicitly defined by the budget constraint:

$$\sum p_i q_i = p_i q_i(p_i/w, \Lambda) = w$$

i.e. such that total expenditures equal a fixed revenue w. Note that, generally, Λ is not a Lagrange multiplier, except for the case where demand can be derived from a directly-additive separable utility (Ligon 2016).

In an unpublished paper by Gorman (printed in Gorman, 1995) mentioned by Pollak (1972), Gorman indicates that a demand system defined as above needs to take specific forms in order to satisfy Slutsky's symmetry condition. With a few additional restrictions, this result can be formulated as follows:[6]

[6]Gorman's sketch of proof had many shortcuts, as he himself noted: "Throughout this paper I have talked as if my claims were definitely proven. Of course this is not so: my arguments are far from rigorous" (Gorman, 1995). Here I impose somewhat stronger assumptions on the form of demand and price effects in order to avoid a few weird cases. In particular, the assumption that expenditure shares are not just a function of Λ allows me to avoid what Gorman calls "the abnormal case".

Proposition 1 *If demand satisfies the following conditions:*

i) generalized separability (equation 3);

ii) there are at least four goods,

iii) holding Λ constant, $(p_i/w)q_i(p_i/w,\Lambda)$ is not constant over the range of prices p_i

Then demand can be written as either:

$$\begin{array}{lrcll} \textit{case 1:} & q_i(p_i/w,\Lambda) & = & \dfrac{D_i(F(\Lambda)p_i/w)}{H(\Lambda)} & \textit{for all goods } i \textit{ and all } p_i, w, \Lambda \\ \textit{case 2:} & q_i(p_i/w,\Lambda) & = & A_i(\Lambda)(p_i/w)^{-\sigma(\Lambda)} & \textit{for all goods } i \textit{ and all } p_i, w, \Lambda \\ \textit{+ case 2':} & q_i(p_i/w,\Lambda) & = & A_i\Lambda^{\rho}(p_i/w)^{-\sigma} & \textit{for } \underline{\textit{all but one}} \textit{ good } i \end{array}$$

or a combination of cases 2 and 2' depending on Λ. In all cases, Λ adjusts and is implicitly defined such that budget constraint is satisfied, i.e. such that $\sum_i q_i(p_i/w,\Lambda)p_i = w$.

Since the third case is not very interesting (CES for all but one good), the remainder of the paper focuses on cases 1 and 2, setting aside case 2'. Note that there may be alternative functional forms under generalized separability if we allow for price-insensitive expenditures shares, which Gorman calls "abnormal" goods. Assumption iii) allows us to avoid such cases.

Finally, note that functional forms are unique up to a constant term and a monotonic transformation of Λ:

Proposition 2 *Uniqueness of functional forms, except for the CES case:*

> *Case 1: $H(\Lambda)$ and $F(\Lambda)$ are uniquely determined by demand patterns, up to a constant term and a strictly-monotonic transformation of Λ*
>
> *Case 2: $A_i(\Lambda)$ and $\sigma(\Lambda)$ are uniquely defined by demand patterns, up to a strictly-monotonic transformation of Λ*

To prove this result, in case 1), note that price effect depend tightly on ε_{Di} (see sections 3.2 and 3.3), hence price elasticities can be used to determine D_i. One can then identify functions F and H by examining variations in ε_{Di} depending on income (except in the CES case).[7] In case 2), σ corresponds to price elasticities and $A_i(\Lambda)$ can be determined by examining income expansion path. Note that for any function H, F, A_i and $\sigma(\Lambda)$, one would obtain the same demand patterns after the change in variable $\Lambda' = g(\Lambda)$ with any one-to-one mapping g.

[7] Price and income elasticities depend on the ratio of ε_F and ε_H which is uniquely determined once we know these elasticities. Taking a strictly monotonic transformation of Λ does not change this ratio and yields the same consumption patterns (assuming that this new aggregator is also such that the budget constraint is satisfied).

2.2 Integrability

2.2.1 First case with homogeneous demand shifters

Let us now examine the reciprocal of Proposition 1. Under which conditions are these demand systems integrable, i.e. can be derived from a rational utility-maximizing consumption behavior? These functional forms, imposed by the symmetry of the Slutsky matrix, do not necessarily correspond to rational consumer behavior (see counter-examples in Appendix B). However it turns out that only weak additional conditions are sufficient to guarantee that the demand systems described in Proposition 1 are integrable.

Suppose that demand is given by:

$$q_i = \frac{D_i(F(\Lambda)p_i/w)}{H(\Lambda)} \tag{4}$$

where Λ is implicitly determined by the budget constraint $\sum_i p_i D_i(F(\Lambda)p_i/w)/H(\Lambda) = w$, which can be rewritten:

$$H(\Lambda) \;=\; \sum_i (p_i/w)\, D_i(F(\Lambda)p_i/w) \tag{5}$$

Let us denote by $\varepsilon_{Di} = \frac{\partial \log D_i}{\partial \log x}$ the elasticity of D_i in its argument, and $\varepsilon_H = \frac{\partial \log H}{\partial \log \Lambda}$ and $\varepsilon_F = \frac{\partial \log F}{\partial \log \Lambda}$ the elasticity of H and F in Λ.

To ensure integrability, we impose the following sufficient regularity restrictions:

Regularity assumptions [**A1**] on functions D_i and H:

i) D_i is differentiable, $\varepsilon_{Di} < 0$ unless $D_i = 0$

ii) H and F are differentiable and $\varepsilon_F\, \varepsilon_{Di} < \varepsilon_H$ for all i, Λ and p_i/w

iii) For any set of normalized prices $y_i = p_i/w$:

$\max_{\Lambda,i}\{y_i D_i(F(\Lambda)y_i)/H(\Lambda)\} > 1 \quad$ and $\quad \min_\Lambda \{D_i(F(\Lambda)y_i)/H(\Lambda)\} = 0$

Assumptions i) and ii) imply that the solution in Λ to equation (5) is always unique, but they are also needed to show that utility is quasi-concave and that the Slutsky substitution matrix is definite negative. The first counter-example provided in Appendix B illustrates what happens when $\varepsilon_F\, \varepsilon_{Di} < \varepsilon_H$ is not satisfied. However, condition ii) on elasticities does not ensure that there exists a Λ to satisfy the budget constraint. By continuity of H and D_i, assumption iii) is then sufficient to ensure that such a solution exists.[8] The first part of condition iii) is similar

[8] Note that condition iii) actually implies that $D_i(F(\Lambda)y_i)/H(\Lambda)$ converges to zero when $\Lambda \to +\infty$, for any good i, and goes to infinity when $\Lambda \to 0$ for at least one good i.

to non-satiety conditions imposed on utility while the second part ensures that demand is well defined for very low incomes.

Under these conditions, we obtain:

Proposition 3 *If H and D_i satisfy the regularity conditions [A1], the demand described in equations (4) and (5) is integrable, i.e. can be derived from a utility function.*

I provide two alternative proofs. First, we can reconstruct a utility function that depends on one implicitly-defined aggregator.[9] We can show that demand can be derived from the maximization of:

$$U(q) = \sum_i \int_{y=0}^{y=H(z)q_i} D_i^{-1}(x)dx \; - \; \int_{z_0}^{z} H'(z)F(z)dz \tag{6}$$

for an arbitrary $z_0 \geq 0$ and where z is itself a function of the consumption vector x implicitly defined by:[10]

$$F(z) = \sum_i q_i D_i^{-1}(H(z)q_i) \tag{7}$$

As a slight abuse of notation, we define $D_i^{-1}(0) = a_i$ if $D_i(y) = 0$ for all $y \geq a_i$ (which yields a choke price) and $D_i^{-1}(x) = 0$ for all $x \geq b_i$ if $D_i(0) = b_i$. Regularity conditions [A1] are needed to ensure that equation (7) always has a unique solution in z and that the utility function is quasi-concave. Note that equation (7) can be seen as a first-order condition such that the derivative of the expression above for U has a zero derivative in z, and such that:

$$\frac{\partial U}{\partial q_i} = D_i^{-1}(H(z)q_i) \tag{8}$$

The two aggregators Λ (function of prices and income) and z (function of quantities q) coincide for the optimal consumption basket.

An alternative proof of Proposition 3 checks that the Slutsky substitution matrix is semi-definite negative under these restrictions, so that we can apply the integrability theorem of Hurwitz and Uzawa (1971). Thanks to Proposition 1, we already know that it is symmetric

[9] This utility representation was pointed out by Gorman (1987) with a more restrictive formulation and no formal proof that such utility function is well defined and quasi-concave. Gorman formulated this as a maximization: $U = \max_z \{\sum_i u_i(zq_i) - \Phi(z)\}$ but this approach omits very useful cases (e.g. cases with $0 < \beta < 1$ in applications 3, 4, 9 and 10 in Section 4) where the second order condition of this maximization is not satisfied yet the utility function remains quasi-concave with z implicitly defined by equation 7.

[10] Note that Λ and z coincide in equilibrium, but z is defined as a function of the vector of quantities while Λ is defined as a function of the vector of prices and income.

but this does not ensure negativity. As one could expect, the conditions ensuring the negative definitiveness of the Slutsky matrix are the same as those providing the quasi-concavity of the utility function above.

The case $F(\Lambda) = 1/H(\Lambda)$ coincides with Matsuyama and Ushchev (2017) for homothetic preferences. With the change in variable: $F(\Lambda) = 1/H(\Lambda) = w\Lambda'$ (where Λ' is again implicitly determined by the budget constraint), a demand system such that $q_i = \Lambda' w D_i(\Lambda' p_i)$ can be rationalized as long as $\varepsilon_{Di} + 1$ has the same sign for all goods (which is equivalent to A1-ii). If $\varepsilon_{Di} + 1$ is negative, we define Λ' instead by $w/\Lambda' = F(\Lambda) = 1/H(\Lambda)$.

More generally, as long as $F'(\Lambda) \neq 0$ for all Λ (which is satisfied for most applications provided in Section 4), we can reformulate demand as:

$$q_i = \widetilde{H}(\Lambda w) D_i(\Lambda p_i) \tag{9}$$

$$\text{with: } \widetilde{H}(\Lambda w) \sum_i p_i D_i(\Lambda p_i) = w \tag{10}$$

Proposition 3 applies to such demand forms, which can be see as a straightforward corollary with the change in variable: $\Lambda' = F(\Lambda)/w$ and the transformation $1/\widetilde{H}(\cdot) = H(F^{-1}(\cdot))$. Denote $\varepsilon_{\widetilde{H}} = \frac{\partial \log \widetilde{H}(h)}{\partial \log h}$ and define:

Regularity assumptions [**A2**] on functions D_i and $\widetilde{H}$:

i) D_i is differentiable and $\varepsilon_{Di} < 0$

ii) $\widetilde{H}$ is differentiable and there exists $\epsilon > 0$ such that $\varepsilon_{\widetilde{H}} + \varepsilon_{Di} < -\epsilon$ for all p_i, w and Λ.

With these restrictions, the strictly negative upper bond on the total elasticity $\varepsilon_{\widetilde{H}} + \varepsilon_{Di}$ ensures existence of a solution in Λ and z so we no longer need the equivalent of part iii) of restrictions [A1]. We obtain:

Corollary *A demand system defined as in equations (9) and (10) above with restrictions [A2] is integrable.*

This corollary applies to most of the applications discussed in Section 4.

In general, Proposition 3 does not require $F(\Lambda)$ to be monotonic in Λ. If $F'(\Lambda) > 0$, an increase in Λ (tightness of the budget constraint) leads to a downward shift in the partial demand curve D_i. When $F'(\Lambda) < 0$, we would instead have an upward shift in D_i, which needs to be compensated by a large enough decrease in the demand shifter $H(\Lambda)$.

2.2.2 Second case with iso-elastic Allen-Uzawa substitution

Now, consider the second case of Proposition 1. Let us assume that demand is given by:

$$p_i q_i = w\left(G_i(\Lambda)\, p_i/w\right)^{1-\sigma(\Lambda)} \tag{11}$$

where Λ is an implicit function of the vector of normalized prices p_i/w such that the budget constraint is satisfied:

$$\sum_i (G_i(\Lambda) p_i/w)^{1-\sigma(\Lambda)} = 1 \tag{12}$$

We obtain that such demand is integrable with a few monotonicity restrictions:

Proposition 4 *Suppose that demand can be written as in equation (11) where G_i and σ are continuous and where Λ is implicitly defined by (12). Suppose also that $\sigma(\Lambda) \neq 1$ for all Λ. This demand system is integrable if, for any Λ, either condition i) or ii) is satisfied:*

i) $\sigma(\Lambda)$ is weakly increasing in Λ and $G_i(\Lambda)$ is strictly increasing in Λ

ii) $\sigma(\Lambda)$ is decreasing in Λ and, for each Λ_0, there exists $\alpha_i > 0$ such that $\sum_i \alpha_i = 1$ and such that $G_i(\Lambda)\, \alpha_i^{\frac{1}{\sigma(\Lambda)-1}}$ is strictly increasing in Λ in a neighborhood of Λ_0

In each case, demand can be derived from a utility function that is implicitly defined by (and is the unique solution of):

$$\sum_i (q_i/G_i(U))^{\frac{\sigma(U)-1}{\sigma(U)}} = 1 \tag{13}$$

with $\Lambda = U$ for the demand q_i described above.

The constant elasticity case $\sigma(\Lambda) = \sigma$ corresponds to implicitly additive utility as in Hanoch (1975), Comin, Laskhari and Mestieri (2015). This does not imply CES since we can have non-trivial income effects through the demand shifter $G_i(\Lambda)$. The main contribution of this proposition is to generalize to variable elasticity of substitution.

When both $\sigma(\Lambda)$ and $G_i(\Lambda)$ are all differentiable, condition ii) can be rewritten after solving for the minimum α_i that would satisfy this monotonicity condition. Condition ii) of Proposition 4 is formally equivalent to imposing:[11]

$$\sum_i \exp\left(\frac{(\sigma(\Lambda)-1)^2 G_i'(\Lambda)}{\sigma'(\Lambda) G_i(\Lambda)}\right) < 1 \tag{14}$$

[11] In general, note that condition ii) need not hold for *any* set of α_i's, it is sufficient that it holds for a single set of α_i's. In particular, using $\alpha_i = 1/N$ (where N denotes the number of goods), a sufficient condition is that $G_i(\Lambda)\, N^{\frac{1}{1-\sigma(\Lambda)}}$ strictly increases in Λ.

See Appendix for the proof of equivalence. We show in Appendix that this condition is necessary if we want utility to always be strictly increasing in quantities.

The proof of Proposition 4 mainly consists in showing that Λ is well-defined, i.e. that the budget constraint has a unique solution in Λ, and that utility is also uniquely defined by equation 13 (quasi-concavity and monotonicity are then almost immediate, as in Comin *et al* 2015).

Why do we need different conditions depending on whether $\sigma(\Lambda)$ decreases with Λ? In the first case where $\sigma(\Lambda)$ increases with Λ, indifference curves become flatter as we move away from the origin (with increases in income and Λ). In that case, indifference curves are most likely to cross around the intercepts (when only one good is consumed). Monotonicity in $G_i(\Lambda)$ is then sufficient to ensure that indifference curves do not cross. In the second case, where the elasticity of substitution $\sigma(\Lambda)$ decreases with Λ, the indifference curves are more curved as we move away from the origin. In this case, indifference curves are most likely to be close to each other and intersect in the middle.

Counter-examples in Appendix to further illustrate the role of each condition, showing that equations (12) and (13) admit multiple solutions in Λ and U if the conditions i) and ii) above are not satisfied (showing incidentally that monotonicity in demand shifters $G_i(\Lambda)$ is not sufficient in the second case).

3 Properties

3.1 Inverse demand function

In the first case, one can express the inverse demand system as:

$$p_i/w = (1/F(z))\ D_i^{-1}(H(z)q_i) \tag{15}$$

where z is implicitly defined as described in equation (7) as function of the vector of consumption. When $F(z)H(z)$ is constant and preferences are homothetic, this aggregator z is homogeneous of degree one in quantities q_i. Note however that there is no one-to-one mapping between z and utility in this case. This inverse demand formulation highlights the symmetric role of H vs. F. While H is the demand shifter in the direct demand function (equation 15), now F is the demand shifter in the inverse demand function.

In the second case of Proposition 1, where demand is determined by equation (11), the inverse demand can be written $p_i = wq_i^{-\frac{1}{\sigma(U)}} G_i(U)^{\frac{1}{\sigma(U)}-1}$ where U is defined an implicit function of the vector of consumption q_i. As in the previous case, the functional form of the inverse demand is nearly identical to the functional form of the direct demand system, but here U corresponds to utility (see Proposition 3), thanks to the constant elasticity of substitution.

3.2 Price effects

In the first case, with demand as in (4), the elasticity of Λ w.r.t price i is:

$$\frac{\partial \log \Lambda}{\partial \log p_i} = W_i \, \frac{1 + \varepsilon_{Di}}{\varepsilon_H - \varepsilon_F \bar{\varepsilon}_D}$$

where $\bar{\varepsilon}_D = \sum_i W_i \varepsilon_{Di}$ denotes the weighted average of partial price elasticities $\varepsilon_{Di} = \frac{\partial \log D_i}{\partial \log x}$, and where W_j is the expenditure share of good j. Given our assumption $\varepsilon_H > \varepsilon_F \bar{\varepsilon}_D$, the elasticity $\frac{\partial \log \Lambda}{\partial \log p_i}$ is positive when ε_{Di} is larger than -1, and negative otherwise. Hence, in the usual case where goods are sufficiently price elastic ($\varepsilon_{Di} < -1$), the aggregator Λ increases with the intensity of competition.

The price elasticity of Marshallian demand is then:

$$\frac{\partial \log q_i}{\partial \log p_j} = \varepsilon_{Di} \, . \, \mathbb{1}_{(i=j)} - \frac{W_j (1 + \varepsilon_{Dj})(\varepsilon_H - \varepsilon_F \varepsilon_{Di})}{\varepsilon_H - \varepsilon_F \bar{\varepsilon}_D}$$

where W_j is the expenditure share of good j and $\mathbb{1}_{(i=j)}$ is a dummy equal to one when $i = j$. Given our restriction $\varepsilon_H > \varepsilon_F \varepsilon_{Di}$, the cross-price elasticity ($i \neq j$) is positive if and only if $\varepsilon_{Dj} < -1$. The own-price elasticity is always negative, which rules out Giffen goods.

The own price elasticity is mainly determined by the shape of function D_i when that good has a negligible market share:

$$\frac{\partial \log q_i}{\partial \log p_i} \approx \varepsilon_{Di}$$

Since we impose very few constraints on ε_{Di} the patterns of price elasticities can be very flexible.

In the second case, when demand corresponds to equation (11), price effects are more simple. The price elasticity of the Hicksian demand corresponds to $\sigma(\Lambda)$, since we can also interpret Λ as utility (Proposition 3).

3.3 Income effects

In the first case (demand as in equation 4), from the budget constraint, we obtain that the elasticity of Λ w.r.t income is:

$$\frac{\partial \log \Lambda}{\partial \log w} = - \frac{1 + \bar{\varepsilon}_D}{\varepsilon_H - \varepsilon_F \bar{\varepsilon}_D}$$

Given our assumption $\varepsilon_H > \varepsilon_F \varepsilon_{Di}$, it is negative when $\bar{\varepsilon}_D > -1$ (when demand is not very price elastic), and positive otherwise.

The income elasticity of good i is:

$$\frac{\partial \log q_i}{\partial \log w} = 1 + \frac{(\varepsilon_H + \varepsilon_F)(\bar{\varepsilon}_D - \varepsilon_{Di})}{\varepsilon_H - \varepsilon_F \bar{\varepsilon}_D} \tag{16}$$

Using this expression, one can see that homotheticity implies that either $\varepsilon_H = -\varepsilon_F$ or $\varepsilon_{Di} = \bar{\varepsilon}_D$ for all i.

As pointed out by Pigou (1910) and Deaton (1974), own-price elasticities and income elasticities are tightly linked when demand is derived from a directly-additive utility (which corresponds to the case where $\varepsilon_H = 0$). Here we obtain: $\frac{\partial \log q_i}{\partial \log w} = \frac{\varepsilon_{Di}}{\bar{\varepsilon}_D}$ with directly separable utility. With $\varepsilon_H \neq 0$, the relationship between income elasticity and price elasticity is muted and we obtain a demand system that is more in line with Deaton (1974) empirical results. Note that $\frac{\partial \log q_i}{\partial \log w} - 1$ has the same sign as $\bar{\varepsilon}_D - \varepsilon_{Di}$ when $\varepsilon_F + \varepsilon_H > 0$, and flipped otherwise. This property can seem attractive, as empirical evidence indicates that price-elastic goods are not necessarily more income elastic.

In the second case, demand corresponding to equation (11) yields even more flexible income effects. Changes in $G_i(\Lambda)$ in Λ need not be related to changes in $\sigma(\Lambda)$. Starting with the special case where $\sigma(\Lambda) = \sigma$ is constant, which corresponds to Hanoch (1975), the effect of income on Λ is such that:

$$\frac{\partial \log \Lambda}{\partial \log w} = \frac{1 - \sigma}{\bar{\varepsilon}_G} \tag{17}$$

where $\bar{\varepsilon}_G$ is an average of elasticities $\varepsilon_{Gi} = \frac{\Lambda G_i'(\Lambda)}{G_i(\Lambda)}$ weighted by expenditures shares. We obtain the income elasticity:

$$\frac{\partial \log q_i}{\partial \log w} = \sigma + (1 - \sigma) \left(\frac{\varepsilon_{Gi}}{\bar{\varepsilon}_G} \right) \tag{18}$$

Good i is income-elastic if $\sigma < 1$ and $\varepsilon_{Gi} > \bar{\varepsilon}_G$ or if $\sigma > 1$ and $\varepsilon_{Gi} < \bar{\varepsilon}_G$. In the more general case where $\sigma(\Lambda)$ is not constant, function G_i plays a similar role and dictates income effects. Moreover, depending on how income affects Λ (with depends on both the sign of $\sigma'(\Lambda)$ and whether $\sigma(\Lambda)$ is smaller than unity), one can have σ increase or decrease with income.

One constraint, however, links the price elasticity and the income elasticity. Both in the cases where $\sigma(\Lambda)$ is fixed or increasing in Λ, the price elasticity imposes a lower bound on income elasticities of demand:[12]

$$
\begin{aligned}
\frac{\partial \log q_i}{\partial \log w} &> \sigma(\Lambda) \qquad \text{if } \ \sigma(\Lambda) < 1 \\
\frac{\partial \log q_i}{\partial \log w} &< \sigma(\Lambda) \qquad \text{if } \ \sigma(\Lambda) > 1
\end{aligned}
$$

Given the two expressions above for income elasticities (22 and 18), we obtain the following Proposition to describe the conditions for homotheticity:

Proposition 5 *Homotheticity:*
i) The demand system in (4) is homothetic if and only if $H(\Lambda)F(\Lambda)$ is constant or if it is CES.
ii) The demand system in (11) is homothetic only in the CES case.

In the homothetic case, assuming $F(z) = 1/H(z) = z$, we obtain that a utility representation is then given by:

$$
U(q) = \sum_i \int_{y=0}^{y=q_i/z} D_i^{-1}(x)dx \ + \ \log(z) \tag{19}
$$

where z is such that $\sum_i (q_i/z)D_i^{-1}(q_i/z) = 1$.

3.4 Additivity and indirect utility

Here we examine the structure of the indirect utility function, particularly for the first case. Pollak (1972) already shows that indirect utility can be expressed as:

$$
V(p,w) = \widetilde{V}(p,w,\Lambda) = \sum_i \int_{y_i F(\Lambda)}^{y_{i0}} D_i(y)dy \ + \ \int_{\Lambda_0}^{\Lambda} F'(l)H(l)dl \tag{20}
$$

where $\Lambda = \Lambda(p,w)$ is implicitly defined as above. Conveniently, we can verify that $\frac{\partial \widetilde{V}}{\partial \Lambda} = 0$, hence we obtain simple expressions for marginal (indirect) utility from income and price changes:

$$
\begin{aligned}
\frac{\partial V(p,w)}{\partial p_i} &= -\frac{F(\Lambda)}{w} D_i(F(\Lambda)p_i/w) \ = \ -\, q_i(p_i/w,\Lambda) \ \cdot \ \frac{F(\Lambda)H(\Lambda)}{w} \\
\frac{\partial V(p,w)}{\partial w} &= \frac{F(\Lambda)}{w^2} \sum_j p_j D_j(F(\Lambda)p_i/w) \ = \ \frac{F(\Lambda)H(\Lambda)}{w}
\end{aligned}
$$

[12] When $\sigma(\Lambda)$ decreases in Λ, the income elasticity exceeds $\sigma(\Lambda)$ if and only if $\exp\left(\frac{(\sigma(\Lambda)-1)^2 G_i'(\Lambda)}{\sigma'(\Lambda)G_i(\Lambda)}\right) < W_i$. This need not be satisfied, even if inequality (14) is imposed.

From the derivative w.r.t income, one can interpret the product of the two shifters $F(\Lambda)H(\Lambda)$ as the marginal utility of income (in log). This term plays a key role for welfare evaluation, especially when we want to compute equivalent or compensating variations. Suppose that we have a change in the vector of prices from p_0 to p_1, and denote $\widehat{W}_0 = \frac{w-CV}{w}$ and $\widehat{W}_1 = \frac{w+EV}{w}$ compensating variations and equivalent variations in relative terms. Using the expression above for indirect utility, one can compute $\widehat{W}_x$ as the change in income that satisfies:

$$\int_{u=1}^{\widehat{W}_x} F(\Lambda)H(\Lambda)\, d\log u \;=\; \int_{p=p_0}^{p_1} F(\Lambda)H(\Lambda) \sum_i \frac{p_i q_i(p_i/w, \Lambda)}{w}\, d\log p_i$$

where, in the right-hand side, $\Lambda = \Lambda(p, w)$ is a function of prices as price changes (income being kept fixed). In the left-hand side, $\Lambda = \Lambda(p_x, uw)$ is a function of prices and income as income adjusts, using either initial prices if $x = 0$ (equivalent variations) or final prices if $x = 1$ (compensating variations) as reference prices. Note that, when preferences are homothetic, $F(\Lambda)H(\Lambda)$ is constant and the left-hand side is simply equal to $\widehat{W}_0 \;=\; \widehat{W}_1$.

Regarding the second form of demand with iso-elastic substitution (equation 11), the indirect utility function corresponds to Λ, implicitly defined by Equation 12.

When are these preferences directly or indirectly additive? Recall that preferences are directly additive if utility can be written as $U(q) = f(\sum_i u_i(q_i))$ where f and $\{u_i\}$ are scalar functions. Preferences are indirectly additive if indirect utility can be written as $V(p, w) = g(\sum_i v_i(p_i/w))$ where g and $\{v_i\}$ are scalar functions.

In the first case when demand is given by (4), we can use equation (6) to describe utility and equation (20) for indirect utility to conclude that:

i) Utility is directly-additive if and only if H is constant or demand is CES.

ii) Indirect utility is additive if and only if F is constant or demand is CES.

Note that aside from the case with direct additivity, Λ differs from the Lagrange multiplier associated with the budget constraint but can still be interpreted as reflecting the tightness of the budget constraint.[13]

For the second case of demand (equation 11) with implicitly-additive utility, the CES case is the only directly-additive or indirectly-additive case.

[13] In fact, Ligon (2016) shows that directly-additive utility is the only case where Λ can be equal to, or be a function of, the Lagrange multiplier associated with the budget constraint.

3.5 Oligopoly and monopolistic competition

It is far easier to examine firm choices under imperfect competition when demand for a good depends only on its own price and a single aggregator Λ. Under Bertrand competition, each firm would take as given the price of other firms and account for the own price effect on the single aggregator Λ. Under monopolistic competition, equilibrium is even more simple as firms would take such an aggregator Λ as given. But even if we take Λ as given, one can see that the shape of the demand is flexible enough to allow for all sorts of behaviors and markups. Such demand systems can lead to flexible pass-through from costs to prices as in Weyl and Fabinger (2013).

If we assume that partial demand curves D_i become more elastic as prices increase ("second law of demand"), effects of wealth and competition on markups depend tightly on the changes in the price shifter $F(\Lambda)/w$.

$$\frac{\partial \log\left(F(\Lambda)/w\right)}{\partial \log w} = -\frac{\varepsilon_H + \varepsilon_F}{\varepsilon_H - \varepsilon_F \bar{\varepsilon}_D}$$

The case $\varepsilon_H + \varepsilon_F > 0$ is the most natural, as it implies that $F(\Lambda)/w$ decreases with income and allows richer markets to be less "tough", i.e. generate higher markups when markets are segmented. This is consistent for instance with empirical findings by Simonovska (2015).

To summarize demand patterns and markup, it can also be useful to take Mazrova and Neary (2013) approach and examine the "demand manifold", i.e. the combinations of demand elasticity and curvature across consumption baskets. Demand with generalized separability allows for very flexible "demand manifold". A shock in Λ can be interpreted as a combination of a translation along the price axis if $F'(\Lambda) \neq 0$ and a translation along the quantity axis if $H'(\Lambda) \neq 0$.

Conversely, note that one could consider the inverse demand function with a single quantity aggregator, which would be more practical for the analysis of Cournot oligopoly equilibrium where firms take other firms' quantities as given. In both cases, direct and inverse demand have the same functional forms and similar properties, as discussed in Section 3.1.

3.6 Choke prices

The demand function as in Proposition 2 can be accommodated to yield choke prices, i.e. a price threshold p_i^* such that $D_i(F(\Lambda)p_i/w) = 0$ for all $p_i \geq p_i^*$, which arises as soon as $D_i(y) = 0$ for large enough y's.

Choke prices arise naturally in various situations. They are useful in international trade to explain why less efficient firms are less likely to export to a specific market (without having to rely on export fixed costs) and to obtain gravity equations as shown in Melitz and Ottaviano (2008) and Arkolakis et al (2015) among others.

Of particular interest is how choke prices depend on income. Income is irrelevant with homothetic preferences, but this prediction contradicts multiple evidence (e.g. Hummels and Klenow 2007) showing that richer consumer buy a greater variety of goods and that richer countries import a larger variety of products. Here, we can obtain demand systems where the choke price can flexibly depend on income. The elasticity of choke price w.r.t income depends both on the shape of F and H.

A special case corresponds to indirectly-additive preferences (H is constant). In this case, choke prices are linear in income, as documented by Bertoletti et al (2017). Here we can obtain more flexible relationships between choke price and income using the more general demand functions described in Proposition 2 – see comment 8 on Bertoletti et al (2017) below. In most cases, choke prices increases with income (not necessarily linearly if H is not constant) but we can even obtain a decreasing relationship between choke prices and income if the elasticities of $H(\Lambda)$ and $F(\Lambda)$ satisfy: $\varepsilon_H + \varepsilon_F < 0$ and $\varepsilon_F > 0$.

4 Practical cases and applications

For the second case with uniform elasticity of substitution across goods:

1. In various contexts, one has associated a lower elasticity of substitution for richer consumers (in line with empirical evidence) while keeping the practicality of CES preferences. Proposition 4 case ii) allows us to do just that. Handbury (2016) and Faber and Fally (2017) assume that the consumption of the outside good influences elasticities. Here, one can circumvent such assumption by defining utility implicitly in a similar fashion as in Hanoch (1975) and Comin et al (2016).

2. Comin et al (2016) provide an excellent application of the case with constant elasticity of substitution σ that does not depend on income. In the calibration of their model, each industry is associated with a distinct structural parameter driving income effect, while keeping constant elasticities of substitution among industries. Here we show that it can be extended to elasticities of substitution that can potentially change with real income. Moreover, these income effects in substitution would not have to be tied (in terms of functional form) to income effects in consumption shares across industries.

   For instance, one application could be to rationalize the rise of profits and fixed costs relative to variable costs. If σ decreases with utility (and thus income), growth would be associated with larger markups and larger variable profits, and under free entry with larger shares of fixed costs over total costs. If fixed costs are more intensive in capital, this would rationalize an increasing share of capital in GDP.

For the first case with non-uniform elasticity of substitution:

3. Consider iso-elastic functions H and F such that demand can be written:

$$q_i = (w\Lambda)^{\beta} D_i(\Lambda p_i)$$

   This leads to practical applications in Trade, as shown in detail in Appendix. By imposing weaker constraints on income effects than with additively-separable preferences, such demand system leads to a more general formula for the welfare gains from trade. While Arkolakis et al (2015) focus on the two polar cases with $\beta \in \{0, 1\}$ (homothetic vs. additively-separable utility), this more general structure shows that such finding can be extended to intermediate cases where $\beta \in (0, 1)$.

4. Iso-elastic functions $D_i(x) = A_i x^{-\sigma_i}$, $F(\Lambda) = \Lambda$ and $H(\Lambda) = \Lambda^{-\beta}$ with $\beta < 1$ lead to another interesting case. It is equivalent to CES only if the σ_i's are identical.

   This corresponds to self-dual addilog preferences previously examined by Houthakker (1965), Pollak (1972) and just recently by Bertoletti and Etro (2017a). With direct separability (when H is constant), we obtain preferences with "Constant Relative Income Elasticities" (Fieler 2011, Caron et al., 2017, 2014). When $\beta = 0$ (CRIE). In the latter case with $\beta = 0$, income elasticities are:

$$\frac{\partial \log q_i}{\partial \log w} = \frac{\sigma_i}{\bar{\sigma}} \tag{21}$$

   where $\bar{\sigma}$ is the expenditure-weighted average of σ_j. With $\beta \neq 0$, we obtain a generalization of such preferences where income elasticities are given by:

$$\frac{\partial \log q_i}{\partial \log w} = 1 + (1 - \beta)\Big(\frac{\sigma_i - \bar{\sigma}}{\bar{\sigma} - \beta}\Big) \tag{22}$$

   These demand systems are non-homothetic except for the case where $\beta = 1$ (which does not necessarily imply CES) or when $\sigma_j = \sigma_i$ for all i, j.

5. Augmented bi-power form, as in Mazrova and Neary (2013)

   Consider the demand system, assuming $\sigma > \nu > 0$:

$$q_i(p_i/w, \Lambda) \;=\; \gamma(\Lambda)[p_i/w]^{-\nu} + \delta(\Lambda)[p_i/w]^{-\sigma}$$

Under which conditions is that demand system integrable? Defining $F(\Lambda) = [\gamma(\Lambda)/\delta(\Lambda)]^{\frac{1}{\sigma-\nu}}$ and $H(\Lambda) = \delta(\Lambda)^{-1}F(\Lambda)^{-\sigma} = \gamma(\Lambda)^{-\frac{\sigma}{\sigma-\nu}}\delta(\Lambda)^{\frac{\nu}{\sigma-\nu}}$, one can recover the form of demand systems as in Mazrova and Neary (2013) by applying Proposition 3.

To apply Proposition 3, one would need q_i to be decreasing in Λ, regardless of prices. Hence a sufficient condition is that both $\delta(\Lambda)$ and $\gamma(\Lambda)$ decrease with Λ. If those conditions are not satisfied, we can see that Λ will not be well defined.[14] Taking very low prices, we have $q_i(p_i/w, \Lambda) \approx \delta(\Lambda)[p_i/w]^{-\sigma}$ and we can see that the equation $\sum_i \delta(\Lambda)[p_i/w]^{1-\sigma} = 1$ can lead to multiple solutions in Λ if δ is not monotonic. Conversely, if we have very high prices, $q_i(p_i/w, \Lambda) \approx \gamma(\Lambda)[p_i/w]^{-\nu}$, and the equation $\sum_i \gamma(\Lambda)[p_i/w]^{1-\sigma} = 1$ can lead to multiple solutions in Λ if γ is not monotonic. Hence monotonicity in both $\gamma(\Lambda)$ and $\delta(\Lambda)$ is required to ensure that Λ is well defined.

6. The previous example builds on a symmetric case (symmetric across all goods). More generally, one can also consider asymmetric bi-power forms. With $D_i(y_i) = \gamma_i y_i^{-\nu_i} + \delta_i y_i^{-\sigma_i}$, one can obtain recoverable demand systems as long as $F(\Lambda)$ and $H(\Lambda)$ are common across all goods, i.e. if demand takes the form:

$$q_i(p_i/w, \Lambda) = \gamma_i H(\Lambda)[F(\Lambda)p_i/w]^{-\nu_i} + \delta_i H(\Lambda)[F(\Lambda)p_i/w]^{-\sigma_i}$$

with $\frac{\partial \log H}{\partial \log \Lambda} < \min\{\nu_i, \sigma_i\}$ to ensure that Λ is well defined.

7. Similarly, one can consider inverse demand functions that are bi-power, with the same extensions as above.

8. Mrazova et al. (2017) introduce CREMR demand (with constant revenue elasticity of marginal revenue) where the own-price effects are such that the distribution of sales and productivity belong to the same family (e.g. lognormal+lognormal, Pareto+Pareto, etc.). They rationalize their demand system with a directly-additive utility function.

Using our results, one can further generalize such demand system (though only in the superconvex case where there is no restriction on minimum quantities) with more flexible income effects while keeping the CREMR properties linked to price effects. If we specify:

$$D_i^{-1}(x) = \frac{\beta_i}{x}(x - \gamma_i)^{\frac{\sigma-1}{\sigma}}$$

with $\gamma_i < 0$, and $H(z)$ such that $\varepsilon_H \geq -1$, the following (inverse) demand systems is

[14]If both δ and γ are instead increasing in Λ, we can just replace Λ by $1/\Lambda$.

integrable (see Section 3.1):

$$p_i(q_i, z) = w/z \; D_i^{-1}(H(z)q_i)$$

where z is implicitly defined by the budget constraint $\sum_i \frac{p_i q_i}{w}$=1 (see Section 3.1), which can be written:

$$z \; = \; \sum_i \beta_i (H(z)q_i - \gamma_i)^{\frac{\sigma-1}{\sigma}} \; / \; H(z)$$

The directly additive case described in Mrazova et al. (2017) corresponds to the case where $H(z)$ is constant. The homothetic case corresponds to $H(z) = 1/z$.

9. With $F(\Lambda) = 1$, as noted in Section 3.4, this case corresponds to indirectly-additive preferences, with a simple demand function:

$$q_i = D_i(p_i/w) / \sum_j D_j(p_j/w)$$

As discussed in Bertoletti and Etro (2017) and Bertoletti, Etro and Simonovska (2017), those preferences are very practical for monopolistic competition under non-homothetic preferences.

As noted in Bertoletti et al (2017), a key property of indirectly-additive preferences is a choke price (assuming that demand drops to zero at sufficiently high prices) that depends linearly with income (instead of the marginal utility of income). Here, with $F(\Lambda) = \Lambda^\gamma$ and $H(\Lambda) = \Lambda^{-\beta}$, we obtain a simple generalization of this result. Assume that demand for each good i takes the form:

$$q_i(p_i, \Lambda) = \Lambda^\beta D(\Lambda^\gamma p_i/w)$$

This demand system is well defined if $\beta < -\gamma \varepsilon_D$. Assuming $D(x) = 0$ if $x \geq a$, we obtain a choke price $p^* = aw\Lambda^{-\gamma}$. The elasticity of the choke price w.r.t income w is then:

$$\frac{\partial \log p^*}{\partial \log w} = 1 - \gamma \frac{\partial \log \Lambda}{\partial \log w} = \frac{\beta - \gamma}{\beta + \gamma \bar{\varepsilon}_D}$$

This elasticity can take a wide range of value, including any value between 0 and 1. The special case with $\gamma = 0$ corresponds to Bertoletti et al (2017) where the choke price is linear in income. The special case with $\gamma = \beta$ corresponds to homothetic preferences where the choke price is independent of income.

10. Conditionally-linear demand:

$$q_i = (\alpha_i - \gamma_i \Lambda p_i)/H(w\Lambda)$$

is integrable as long as $\Lambda p_i \leq \frac{\alpha_i}{2\beta_i}$ and $\varepsilon_H > -1$.

This generalizes Ottaviano et al. (2002), Melitz and Ottaviano (2008) and Mayer et al. (2014) based on quasi-linear preferences. This conditionally-linear demand system nevertheless yields very simple expressions for markups in monopolistic competition when Λ is taken as given (limit case with many firms).

A practical case is to impose $H(w\Lambda) = (w\Lambda)^{-\beta}$ with $\beta < 1$, but even in this most simple case there is no fully closed-form solution since utility (direct and indirect) still depends on functions z and Λ that are implicitly defined.[15] While $z^{-\beta} q_i < \alpha_i$ and $p_i \Lambda < \alpha_i/\gamma_i$, utility and indirect utility take the form:

$$\begin{aligned} U(q) &= \frac{\beta z^{1-\beta}}{1-\beta} + \sum_i \frac{z^{-\beta} q_i (2\alpha_i - z^{-\beta} q_i)}{2\gamma_i} \\ V(p,w) &= \frac{(w\Lambda)^{1-\beta}}{1-\beta} - \sum_i \frac{p_i \Lambda (2\alpha_i - \gamma_i p_i \Lambda)}{2} \end{aligned}$$

where z and Λ are implicitly defined as the solutions of $z = \sum_i q_i(\alpha_i - z^{-\beta} q_i)/\gamma_i$ and $\sum_i (w\Lambda)^\beta p_i(\alpha_i - \gamma_i \Lambda p_i) = w$ respectively.

11. Counter-examples where demand depends on more than one aggregator: Kimball (1995) and QMOR (Feenstra 2015). QMOR demand can be expressed as:

$$q_i = \alpha u \left(\frac{p_i}{e(p)}\right)^{r-1} \left[1 - \left(\frac{p^*}{p_i}\right)^{r/2}\right]$$

which is a function of two aggregators: p^* and $e(p)$. Given that $e(u)u = w$, we also have:

$$\frac{p_i q_i}{w} = \alpha \left(\frac{p_i}{e(p)}\right)^{r} \left[1 - \left(\frac{p^*}{p_i}\right)^{r/2}\right]$$

[15] Note that the conditions in Proposition 2 are not always satisfied if $x \leq \frac{\alpha_i}{2\gamma_i}$ and the elasticity of H is close to -1 (D_i would have an elasticity below unity in absolute value). In a monopolistic competition framework, a trick is to replace D_i by $\frac{\alpha^2}{4\gamma_i x}$ for $x < \frac{\alpha}{2\gamma_i}$. Such a function D_i would satisfy the conditions of Proposition 1 as long as the elasticity of H remains between zero and unity. In equilibrium, none of the firms would set an infinite markup, hence equilibrium prices are such that $\Lambda p_i > \frac{\alpha_i}{2\gamma_i}$ and none of the firms would end up on the non-linear portion of the demand curve.

Note that the price elasticity can be expressed as a function of p^* only.

Similarly, demand with Kimball preferences with an implicit aggregator depends in fact on two aggregators (it also depend on the Lagrange multiplier associated with the budget constraint) hence they are not a special case of the demand systems described here.

5 Concluding remarks

Economists have often focused on demand systems where prices are conveniently summarized by a single aggregator, and where demand depends solely on such an aggregator, wealth and a good's own price ("generalized separability", following Pollak 1972 terminology). Here I show that such demand system can take only one of two forms when price effects are not trivial. This result was already known by Pollak (1972) and Gorman (1972) but not formally demonstrated and is not well known today in spite of its usefulness. In addition, here we show that these two types of demand systems are integrable (i.e. can be derived from well-behaved utility functions) under fairly mild regularity restrictions that guarantee the semi-definite negativity of the Slutsky substitution matrix (and quasi-concavity of utility).

The first case of demand allows for very flexible price effects, yet with somewhat more restricted income effects. While the magnitude of income effects is flexible, the ranking in income elasticities is still tightly linked to the ranking in price elasticities (a weaker form of Pigou's Law on price and income elasticities). The second case of demand allows for very flexible income effects but more restricted price effects. Allen-Uzawa substitution elasticities have to be constant across goods but here I show that they may vary (increase or decrease) with utility and thus vary indirectly with income.

There are numerous applications. This can be used to generalize various demand systems used in recent work across different fields. Here I provide several illustrations related to structural transformation, trade and industrial organization. Such demand systems are particularly well suited to the study of firm behavior under monopolistic competition or under Bertrand competition where other firms' prices are taken as given. Conversely, the results described here also extend to inverse demand functions with a single aggregator in quantities, which would be particularly practical for Cournot oligopoly analysis.

Appendix A – Proofs of Propositions 1, 3 and 4

Proof of Proposition 1

Consider the demand system:

$$q_i = (w/p_i).W_i(p_i/w, \Lambda) \tag{23}$$

where W_i denotes expenditure shares as a function of normalized prices p_i/w (demand is homogeneous of degree zero in prices and income) and Λ (by assumption), and where Λ is implicitly determined by the budget constraint:

$$\sum_i W_i(p_i/w, \Lambda) = 1 \tag{24}$$

We denote normalized prices by $y_i \equiv p_i/w$ such that $q_i = (1/y_i).W_i(y_i, \Lambda)$ and with Λ an implicit function of the vector of y_i such that: $\sum_j W_j(y_j, \Lambda) = 1$.

For $i \neq j$, the Slutsky substitution coefficient is:

$$\begin{aligned}
s_{ij} &= \frac{\partial q_i}{\partial p_j} + q_j \frac{\partial q_i}{\partial w} \\
&= \frac{w}{p_i}\frac{\partial W_i}{\partial \Lambda}\left[\frac{\partial \Lambda}{\partial p_j} + q_j \frac{\partial \Lambda}{\partial w}\right] - \frac{q_j}{w}\frac{\partial W_i}{\partial y_i} - \frac{q_j q_i}{w} \\
&= \frac{\partial \log W_i}{\partial \log \Lambda}\left[\frac{q_i}{p_j}\frac{\partial \log \Lambda}{\partial \log p_j} + \frac{q_i q_j}{w}\frac{\partial \log \Lambda}{\partial \log w}\right] - \frac{q_j q_i}{w}\frac{\partial \log W_i}{\partial \log y_i} - \frac{q_j q_i}{w} \\
&= \frac{q_i \epsilon_{\Lambda i}}{p_j}\frac{\partial \log \Lambda}{\partial \log p_j} + \frac{q_i q_j \epsilon_{\Lambda i}}{w}\frac{\partial \log \Lambda}{\partial \log w} - \frac{q_j q_i \epsilon_{yi}}{w} - \frac{q_j q_i}{w}
\end{aligned}$$

where we denote:

$$\epsilon_{yi} \equiv \frac{\partial \log W_i}{\partial \log y_i} \qquad \text{and} \qquad \epsilon_{\Lambda i} \equiv \frac{\partial \log W_i}{\partial \log \Lambda}$$

To compute the derivatives of Λ in w and p_i, we differentiate (24) w.r.t w, which gives:

$$\frac{\partial \Lambda}{\partial w}\left[\sum_i \frac{\partial W_i}{\partial \Lambda}\right] - \left[\sum_i \frac{p_i}{w^2}\frac{\partial W_i}{\partial y_i}\right] = 0$$

$$\Longleftrightarrow \frac{\partial \log \Lambda}{\partial \log w}\left[\sum_i W_i \epsilon_{\Lambda i}\right] - \left[\sum_i W_i \epsilon_{yi}\right] = 0$$

We obtain:

$$\frac{\partial \log \Lambda}{\partial \log w} = \frac{\sum_i W_i \epsilon_{yi}}{\sum_i W_i \epsilon_{\Lambda i}} = \frac{\overline{\epsilon_y}}{\overline{\epsilon_\Lambda}}$$

where $\overline{\epsilon_y}$ and $\overline{\epsilon_\Lambda}$ denotes the expenditure-weighted averages of ϵ_{yi} and $\epsilon_{\Lambda i}$.

Similarly, differentiating (24) w.r.t price p_j, we get:

$$\frac{\partial \Lambda}{\partial p_j}\left[\sum_i \frac{\partial W_i}{\partial \Lambda}\right] \;+\; \frac{1}{w}\frac{\partial W_j}{\partial y_j} = 0$$

$$\Longleftrightarrow \frac{\partial \log \Lambda}{\partial \log p_j}\left[\sum_i W_i \epsilon_{\Lambda i}\right] \;+\; W_j \epsilon_{yj} = 0$$

which gives:

$$\frac{\partial \log \Lambda}{\partial \log p_j} = -\frac{W_j \epsilon_{yj}}{\overline{\epsilon_\Lambda}}$$

Incorporating the expressions for the derivatives of Λ, the Slutsky coefficients become:

$$\begin{aligned}
s_{ij} &= -\frac{q_i \epsilon_{\Lambda i}}{p_j}\frac{W_j \epsilon_{yj}}{\overline{\epsilon_\Lambda}} + \frac{q_i q_j \epsilon_{\Lambda i}}{w}\frac{\overline{\epsilon_y}}{\overline{\epsilon_\Lambda}} \;-\; \frac{q_j q_i \epsilon_{yi}}{w} \;-\; \frac{q_j q_i}{w} \\
&= -\frac{q_i q_j \epsilon_{\Lambda i}}{w}\frac{\epsilon_{yj}}{\overline{\epsilon_\Lambda}} + \frac{q_i q_j \epsilon_{\Lambda i}}{w}\frac{\overline{\epsilon_y}}{\overline{\epsilon_\Lambda}} \;-\; \frac{q_j q_i \epsilon_{yi}}{w} \;-\; \frac{q_j q_i}{w} \\
&= \frac{q_i q_j}{w \overline{\epsilon_\Lambda}}\left[-\epsilon_{yj}\epsilon_{\Lambda i} + \epsilon_{\Lambda i}\overline{\epsilon_y} \;-\; \epsilon_{yi}\overline{\epsilon_\Lambda} \;-\; \overline{\epsilon_\Lambda}\right]
\end{aligned}$$

Rearranging, one can see that the Slutsky substitution matrix is symmetrical if and only if, for all $i \neq j$:

$$\epsilon_{yi}\epsilon_{\Lambda j} - \epsilon_{\Lambda j}\overline{\epsilon_y} \;+\; \epsilon_{yj}\overline{\epsilon_\Lambda} \;=\; \epsilon_{yj}\epsilon_{\Lambda i} \;-\; \epsilon_{\Lambda i}\overline{\epsilon_y} \;+\; \epsilon_{yi}\overline{\epsilon_\Lambda}$$

Subtracting $\epsilon_{yj}\epsilon_{\Lambda j}$ on both sides, rearranging and factorizing, this can be rewritten:

$$(\epsilon_{yj} - \overline{\epsilon_y})(\epsilon_{\Lambda j} - \epsilon_{\Lambda i}) \;=\; (\epsilon_{yj} - \epsilon_{yi})(\epsilon_{\Lambda j} - \overline{\epsilon_\Lambda})$$

This holds for any pair of goods i and j. Picking any three goods i, j, k from the consumption basket, we have:

$$\begin{aligned}
(\epsilon_{yj} - \overline{\epsilon_y})(\epsilon_{\Lambda j} - \epsilon_{\Lambda i}) &= (\epsilon_{yj} - \epsilon_{yi})(\epsilon_{\Lambda j} - \overline{\epsilon_\Lambda}) \\
(\epsilon_{yi} - \overline{\epsilon_y})(\epsilon_{\Lambda i} - \epsilon_{\Lambda k}) &= (\epsilon_{yi} - \epsilon_{yk})(\epsilon_{\Lambda i} - \overline{\epsilon_\Lambda}) \\
(\epsilon_{yk} - \overline{\epsilon_y})(\epsilon_{\Lambda k} - \epsilon_{\Lambda j}) &= (\epsilon_{yk} - \epsilon_{yj})(\epsilon_{\Lambda k} - \overline{\epsilon_\Lambda})
\end{aligned}$$

Adding up, the average terms $\overline{\epsilon_y}$ and $\overline{\epsilon_\Lambda}$ disappear and we obtain:

$$\epsilon_{\Lambda i}\epsilon_{yj} + \epsilon_{\Lambda k}\epsilon_{yi} + \epsilon_{\Lambda j}\epsilon_{yk} = \epsilon_{\Lambda j}\epsilon_{yi} + \epsilon_{\Lambda i}\epsilon_{yk} + \epsilon_{\Lambda k}\epsilon_{yj}$$

Subtracting $\epsilon_{\Lambda j}\epsilon_{yj}$ on both sides and factorizing, we obtain:

$$(\epsilon_{\Lambda i} - \epsilon_{\Lambda j})(\epsilon_{yj} - \epsilon_{yk}) = (\epsilon_{\Lambda j} - \epsilon_{\Lambda k})(\epsilon_{yi} - \epsilon_{yj}) \tag{25}$$

Let us denote by $\mathcal{Y}(\Lambda)$ the set of feasible y (vector of normalized prices for which the aggregator takes a value Λ):

$$\mathcal{Y}(\Lambda) = \left\{ y \, ; \;\; s.t. \; \sum_i W_i(y_i, \Lambda) = 1 \right\}$$

and denote by $\mathcal{Y}_i(\Lambda)$ its projection onto the i axis, i.e. the set of feasible y_i for a given Λ.

For a given Λ, to satisfy equation (25) for any triplet of goods, three cases arise naturally:
- case 2A: $\epsilon_{yj} = \epsilon_{yi}$ for all goods, for each $y \in \mathcal{Y}(\Lambda)$.
- case 2B: $\epsilon_{yj} = \epsilon_{yi}$ for all but one good $i0$, for all $y \in \mathcal{Y}(\Lambda)$.
- case 1: there exists three goods $i \neq j \neq k$ such that $\epsilon_{yi} \neq \epsilon_{yj} \neq \epsilon_{yk}$ for some $y \in \mathcal{Y}(\Lambda)$.

Case 2A is easier to consider. Recall that, in general, ϵ_{yj} is a function of y_j and Λ. The goal is to show that

- $W_i(y_i, \Lambda) = A_i(\Lambda) y_i^{\epsilon(\Lambda)}$ for all $y_i \in \mathcal{Y}_i(\Lambda)$ for some functions $A_i(\Lambda)$ and $\epsilon(\Lambda)$.
- $\mathcal{Y}_i(\Lambda) = \left\{ y_i \; ; \; A_i(\Lambda) y_i^{\epsilon(\Lambda)} < 1 \right\}$

Suppose that $y^* \in \mathcal{Y}(\Lambda)$, i.e. such that $\sum W_i(y_i^*, \Lambda) = 1$ and denote $\epsilon(\Lambda) = \epsilon_{yj}(y_j, \Lambda)$. Also suppose for now that $\epsilon(\Lambda)$ is strictly positive (the same proof applies to the other case after a simple change in variable). Define $A_i(\Lambda)$ such that $W_i(y_i^*, \Lambda) = A_i(\Lambda) y_i^{*\epsilon(\Lambda)}$.

Denote by $(\underline{y_i}, \overline{y_i})$ the maximum interval included in $\mathcal{Y}_i(\Lambda)$ containing y_i^* and such that, for each $y_i \in (\underline{y_i}, \overline{y_i})$, there exists $\tilde{y} \in \mathcal{Y}(\Lambda)$, with $\tilde{y}_j \in (\underline{y_j}, \overline{y_j})$ in each of its argument, with $\tilde{y}_i = y_i$, and such that $W_j(\tilde{y}_j, \Lambda) = A_j(\Lambda)(\tilde{y}_j)^{\epsilon(\Lambda)}$.

By contradiction, suppose that this is not the case: suppose that for a good $i = I$ we have $\underline{y_I} > 0$. We need to show that it is then possible to construct a new vector y' such that $W_j(y'_j, \Lambda) = A_j(\Lambda)(y'_j)^{\epsilon(\Lambda)}$ and $\sum_j A_j(\Lambda)(y'_j)^{\epsilon(\Lambda)} = 1$, and such that $y'_I < \underline{y_I}$.

Since we deal with bounded intervals, we can construct a series $y^{(n)}$ that satisfies these four conditions:
i) $W_j(y_j^{(n)}, \Lambda) = A_j(\Lambda)(y_j^{(n)})^{\epsilon(\Lambda)}$;
ii) $\sum_j W_j(y_j^{(n)}, \Lambda) = 1$;
iii) each term converge to a finite (possibly zero) value denoted y_j^{∞};
iv) and such that $y_I^{(n)}$ converges to $\underline{y_I} > 0$.

By continuity (with the abuse of notation $W_i(0, \Lambda) = 0$ in the limit case), we must have $W_j(y_j^{\infty}, \Lambda) = A_j(\Lambda)(y_j^{\infty})^{\epsilon(\Lambda)}$ and $\sum_j W_j(y_j^{\infty}, \Lambda) = 1$. Denote by K_0 the set of goods k such that $y_k^{\infty} = 0$. If K_0 is empty, pick a good $k \neq I$. Since $\underline{y_I} > 0$, we know that such a good k satisfies $\underline{y_k} < 1/A_k$. Next, since $\underline{y_I} > 0$ and $\underline{y_k} < 1/A_k$ for $k \in K_0$, and since the derivative of W_I is non-zero at $\underline{y_I}$, we can find small enough but positive $\nu_I > 0$ and $\nu_k > 0$ (for $k \in K_0$) such that:

$$W_I(y_I^{\infty} - \nu_I, \Lambda) + \sum_{k \in K_0} W_k(y_k^{\infty} + \nu_k, \Lambda) + \sum_{j \notin K_0, k \neq I} W_j(y_j^{\infty}, \Lambda) = 1$$

Moreover, given that the derivative of W_i is strictly positive on each interval $(y_I^{\infty} - \nu_I, y_I^{\infty})$ and $(y_k^{\infty}, y_k^{\infty} + \nu_k)$, we can construct a continuum of other vectors y on such intervals that satisfy the same condition as above. For all these vectors, the elasticity ν_{yj} must be kept constant so we must have: $W_j(y_j^\infty, \Lambda) = A_j(\Lambda)(y_j^\infty)^{\epsilon(\Lambda)}$. This contradicts the assumption that $\underline{y_I} > 0$ and proves that $\underline{y_i} = 0$ for all i.

Conversely, one can show that $\overline{y_i} = 1/A_i$ otherwise one can construct a new vector y' such as y^∞ with a i^{th} argument strictly above $\overline{y_i}$ if $\overline{y_i} < 1/A_i$.

Here we have imposed $\epsilon(\Lambda) > 0$. The same arguments can be applied to the case with $\epsilon(\Lambda) < 0$ with the change in variable $y_i' = 1/y_i$.

Case 2B Here we assume that all but one good $i0$ has a common elasticity $\epsilon(\Lambda)$. Applying the same arguments as in the previous case, we obtain that $W_j(y_j, \Lambda) = A_j(\Lambda)y_j^{\epsilon(\Lambda)}$ for all y_j such that $A_i(\Lambda)y_i^{\epsilon(\Lambda)} < 1 - s_{i0}$ where s_{i0} is the minimum expenditure share of good $i0$ for which its elasticity differs from other goods.

On these intervals, we have:

$$\epsilon_{\Lambda j}(y_j, \Lambda) - \epsilon_{\Lambda k}(y_k, \Lambda) = \left(\frac{\epsilon_{\Lambda i0} - \epsilon_{\Lambda j}}{\epsilon_{yi0} - \epsilon_{yj}}\right)(\epsilon_{yj} - \epsilon_{yk}) = 0$$

which holds for any two goods $j, k \neq i0$, we obtain that they also share the same elasticity with respect to Λ: $\epsilon_{\Lambda j}(y_j, \Lambda) = \epsilon_{\Lambda k}(y_k, \Lambda) = \rho(\Lambda)$. Hence these elasticities do not depend on y_i or y_k. This implies that both ϵ and ρ are constant, and that for any good $k \neq i$ on these intervals we must have:

$$W_k(y_k, \Lambda) = A_k \Lambda^\rho y_k^\epsilon \qquad\qquad k \neq i0$$

Case 1 is more involved. Pick any three goods i, j, k from the consumption basket. From equation (25), we have:

$$(\epsilon_{\Lambda i} - \epsilon_{\Lambda j})(\epsilon_{yj} - \epsilon_{yk}) = (\epsilon_{\Lambda j} - \epsilon_{\Lambda k})(\epsilon_{yi} - \epsilon_{yj})$$

In this case, there exists at least three goods $i \neq j \neq k$ such that $\epsilon_{yj} \neq \epsilon_{yi} \neq \epsilon_{yk}$ for some $y \in \mathcal{Y}(\Lambda)$. For these goods, we obtain:

$$\frac{\epsilon_{\Lambda i} - \epsilon_{\Lambda j}}{\epsilon_{yi} - \epsilon_{yj}} = \frac{\epsilon_{\Lambda j} - \epsilon_{\Lambda k}}{\epsilon_{yj} - \epsilon_{yk}} \tag{26}$$
$$\frac{\epsilon_{\Lambda i} - \epsilon_{\Lambda j}}{\epsilon_{yi} - \epsilon_{yj}} = \frac{\epsilon_{\Lambda k} - \epsilon_{\Lambda i}}{\epsilon_{yk} - \epsilon_{yi}} \tag{27}$$

Notice that the left-hand side of both equations potentially depend on y_i and y_j but the right-hand side of equation (26) does not depend on y_i and the right-hand side of equation (27) does not depend on y_j. Since there are at least four goods (by assumption) and the expenditure share of these other goods vary with prices for a given Λ, we obtain that equations (26) and (27) hold for a neighborhood of y_i, y_j and y_k (these can vary over that neighborhood while holding Λ constant).

Denote by $f(\Lambda) = \frac{\epsilon_{\Lambda i} - \epsilon_{\Lambda j}}{\epsilon_{yi} - \epsilon_{yj}}$ the left-hand side of equations (26) and (27).

Picking any good h instead of k, equation (25) also applies and yields:

$$\epsilon_{\Lambda j}(y_j,\Lambda)-\epsilon_{\Lambda k}(y_h,\Lambda)=f(\Lambda)\left(\epsilon_{yj}(y_j,\Lambda)-\epsilon_{yh}(y_h,\Lambda)\right) \tag{28}$$

Given that there are at least four goods, this should hold over a non-trivial range of y_h (same argument as above). Taking the derivative w.r.t. y_h, we get:

$$\frac{\partial \epsilon_{\Lambda h}}{\partial \log y_h}=f(\Lambda)\,\frac{\partial \epsilon_{yh}}{\partial \log y_h} \tag{29}$$

Given the symmetry of the cross-derivative (we assume that demand is twice differentiable), this can be rewritten:

$$\frac{\partial \epsilon_{yh}}{\partial \log \Lambda}=f(\Lambda)\,\frac{\partial \epsilon_{yh}}{\partial \log y_h} \tag{30}$$

By continuity, the conditions required for this equation hold over a neighborhood of Λ. Take reference Λ_0 and consider the function: $M(y,\Lambda)=\epsilon_{yh}(y,\Lambda)-\epsilon_{yh}(yF(\Lambda),\Lambda_0)$ where $F(\Lambda)=\exp\left(\int_{\Lambda_0}^{\Lambda} f(t)\frac{dt}{t}\right)$. Note that $F(\Lambda_0)=1$ hence $M(y,\Lambda_0)=0$. For other Λ, we obtain:

$$\begin{aligned}
\frac{\partial M}{\partial \log \Lambda} &= \frac{\partial \epsilon_{yh}(y,\Lambda)}{\partial \log \Lambda}-\frac{\partial \log F}{\partial \log \Lambda}\,\frac{\partial \epsilon_{yh}(yF(\Lambda),\Lambda_0)}{\partial \log y_h}\\
&= \frac{\partial \epsilon_{yh}(y,\Lambda)}{\partial \log \Lambda}-f(\Lambda)\,\frac{\partial \epsilon_{yh}(yF(\Lambda),\Lambda_0)}{\partial \log y_h}\\
&= 0
\end{aligned}$$

Hence, $M(y,\Lambda)=0$ over the neighborhood where equation (30) holds. This implies:

$$\epsilon_{yh}(y_h,\Lambda)=\epsilon_{yk}(y_hF(\Lambda),\Lambda_0)$$

Now, for a given reference point y^* and Λ_0, let us construct D_i as:

$$D_h(y_h)=W_i(y_h^*,\Lambda_0)\,\exp\left[\int_{y_h^*}^{y_h}\epsilon_{yh}(y,\Lambda_0)dy\right]$$

Integrating over y from a reference point y^* in the region where this equality holds, we obtain that demand can be written as:

$$\begin{aligned}
W_h(y_h,\Lambda) &= W_h(y_h^*,\Lambda)\,\exp\left[\int_{y_h^*}^{y_h}\epsilon_{yh}(y,\Lambda)\frac{dy}{y}\right]\\
&= W_h(y_h^*,\Lambda)\,\exp\left[\int_{y_h^*}^{y_h}\epsilon_{yh}(yF(\Lambda),\Lambda_0)\frac{dy}{y}\right]\\
&= W_h(y_h^*,\Lambda)\,\exp\left[\int_{y_h^*F(\Lambda)}^{y_hF(\Lambda)}\epsilon_{yh}(y,\Lambda_0)\frac{dy}{y}\right]
\end{aligned}$$

$$
\begin{aligned}
&= W_h(y_h^*, \Lambda)\,.\,\frac{D_h(y_hF(\Lambda))}{D_h(y_h^*F(\Lambda))} \\
&= H_h(\Lambda)\, D_h(y_hF(\Lambda))
\end{aligned}
$$

where H_h is a function of Λ defined as:

$$
H_h(\Lambda) \equiv \frac{D_h(y_h^*F(\Lambda))}{W_h(y_h^*, \Lambda)}
$$

for which, by definition of D_h, we have: $H_h(\Lambda_0) = \frac{D_h(y_h^*F(\Lambda_0))}{W_h(y_h^*, \Lambda_0)} = \frac{D_h(y_h^*)}{W_h(y_h^*, \Lambda_0)} = 1$.

Examining elasticities w.r.t. prices and Λ, we can then show that H_h is in fact identical across goods. To check this, with $W_h(y_h, \Lambda) = D_h(y_hF(\Lambda))/H_i(\Lambda)$, we have:

$$
\epsilon_{yh} = \epsilon_{Dh}
$$

where ϵ_{Dh} denotes the elasticity of D_h, while the elasticity in Λ is:

$$
\epsilon_{\Lambda h} = f(\Lambda)\, \epsilon_{Dh} \; - \; \epsilon_{Hh}
$$

where ϵ_{Hh} denotes the elasticity of H_h in Λ. Thanks for condition (28), we obtain that for any two goods h and k:

$$
\begin{aligned}
\epsilon_{Hk} - \epsilon_{Hh} &= \epsilon_{\Lambda h} - \epsilon_{\Lambda k} - f(\Lambda)\,(\epsilon_{Dh} - \epsilon_{Dk}) \\
&= \epsilon_{\Lambda h} - \epsilon_{\Lambda k} - f(\Lambda)\,(\epsilon_{yh} - \epsilon_{yk}) \\
&= 0
\end{aligned}
$$

This implies that $H_h(\Lambda)$ and $H_k(\Lambda)$ remain proportional for any two goods k and h. Given that $H_h(\Lambda_0) = H_k(\Lambda_0) = 1$, H_h must be identical across all goods h.

Hence, demand in case 1 can be written for any good h as:

$$
W_h(y_h, \Lambda) = D_h(y_hF(\Lambda)) \,/\, H(\Lambda)
$$

Combinations of cases: Locally, for a given Λ and around it, one must be in one of these three cases. A remaining question is whether demand can be a mixture of these three cases as Λ varies. To finish the proof of Proposition 1, we show that we cannot combine cases 1 with cases 2A and 2B. Hence the functional form of case 1 needs to hold globally across all Λ's, while we can potentially have a combination of 2A and 2B across Λ.

Combination of cases 1+2A Here we show that we cannot have a combination of cases 1 and 2A globally. First, note that for a given Λ, case 1 and 2 are mutually exclusive by definition. Hence, if we have a mixture of cases 1 and 2, it must occurs along different Λ's. By contradiction, suppose that there exists Λ^* such that, at least locally,

$$
W_i(y_i, \Lambda) = D_i(F(\Lambda)y_i) \,/\, H(\Lambda) \qquad \text{if } \Lambda < \Lambda^*
$$

$$W_i(y_i, \Lambda) = A_i(\Lambda) y_i^{1-\sigma(\Lambda)} \qquad \text{if } \Lambda > \Lambda^*$$

By continuity, at the limit where $\Lambda = \Lambda^*$, we must have:

$$\frac{\partial \log D_i(F(\Lambda^*)y)}{\partial \log y} = 1 - \sigma(\Lambda^*)$$

Since it must hold for any i and any y, it implies that $F^*(y) = 0$ and that $1 - \sigma(\Lambda^*) = 0$, which contradicts our assumption that $W_i(y_i, \Lambda)$ is not locally constant across y_i for any given Λ.

Combinations of cases 1+2B Here we show that we cannot have a combination of cases 1 and 2B globally, using the same arguments as above. Note again that for a given Λ, case 1 and 2B are mutually exclusive by definition. Hence, if we have a mixture of cases 1 and 2B, it must occurs along different Λ's.

By contradiction, suppose that there exists Λ^* such that, at least locally, such that for all but one good we have:

$$W_i(y_i, \Lambda) = D_i(F(\Lambda) y_k)/H(\Lambda) \qquad \text{if } \Lambda < \Lambda^*$$
$$W_i(y_i, \Lambda) = \Lambda^{-\rho} A_i y_i^{1-\sigma} \qquad \text{if } \Lambda > \Lambda^*$$

Again, by continuity, at the limit where $\Lambda = \Lambda^*$, we must have:

$$\frac{\partial \log D_i(F(\Lambda^*)y)}{\partial \log y} = 1 - \sigma$$

Since it must hold for any i and any y, it implies that either demand is CES or $F^*(y) = 0$ and that $1 - \sigma(\Lambda^*) = 0$. This contradicts the assumption that $W_i(y_i, \Lambda)$ is not locally constant across y_i for any given Λ.

Proof of Proposition 3

Define $\widetilde{U}(q,z)$ as:

$$\widetilde{U}(q,z) = \sum_i u_i(H(z)q_i) \;-\; \int_{z_0}^{z} F(z)H'(z)dz$$

where:

$$u_i(q_i) = \int_{q=0}^{q_i} D_i^{-1}(x)dx$$

and $u_i' = D_i^{-1}$. Recall that D_i is strictly decreasing unless $D_i = 0$. As noted in the text, as an abuse of notation, we define $D_i^{-1}(0) = a_i$ if $D_i(y) = 0$ for all $y \geq a_i$ (which yields a choke price) and $D_i^{-1}(x) = 0$ for all $x \geq b_i$ if $D_i(0) = b_i$.

In turn, we want to define z as an implicit function of q such that:

$$F(z) = \sum_i q_i u_i'(H(z)q_i) \tag{31}$$

We proceed in three steps. First we show that equation (31) admits a solution $z(q)$ for each q and that this solution is unique. Second we show that utility defined as $U(q) = \widetilde{U}(q, z(q))$ is well-behaved and quasi-concave. Finally, we show that maximizing U leads to the demand function in the text, and that the single aggregator Λ is also well defined.

Step 1: Implicit function $z(q)$. Here we show that for any vector q of consumption, there is a z such that equation (31) holds.

With restrictions [A1] iii), we have assumed that for each good i and each y_i, there is a Λ such that $D_i(F(\Lambda)y_i)/H(\Lambda)$ is arbitrarily small. Take $y_i = 1/(Nq_i)$ where N denotes the number of goods. For each i, there is a z such that:

$$D_i(F(z)/(Nq_i))/H(z) < q_i$$

Since the left-hand side decreases with z (this is implied by restriction [A1]-ii), we can take the maximum z of all z_i's such that it holds for all goods i with a common z. This inequality is equivalent to:

$$\begin{aligned} D_i(F(z)/(Nq_i)) &< H(z)q_i \\ \Leftrightarrow \quad F(z)/(Nq_i)) &> u_i'(H(z)q_i) \\ \Leftrightarrow \quad F(z)/N &> q_i u_i'(H(z)q_i) \end{aligned}$$

Going from the first to second inequality above is guaranteed by the assumption that D_i decreases strictly. Adding across all goods (given that $q_i u_i'$ is positive for all goods), we obtain that for each vector q there exists a z such that:

$$F(z) > \sum_i q_i u_i'(H(z)q_i)$$

Next, with restrictions [A1] iii), we have also assumed that there is at least one good i such that, for each y_i, there is a Λ such that $y_i D_i(F(\Lambda)y_i)/H(\Lambda)$ is larger than 1. Taking $y_i = 1/q_i$, there is a $z = \Lambda$ such that:

$$\begin{aligned} (1/q_i)D_i(F(z)/q_i)/H(z) &> 1 \\ \Leftrightarrow \quad D_i(F(z)/q_i) &> H(z)q_i \\ \Leftrightarrow \quad F(z)/(Nq_i)) &< u_i'(H(z)q_i) \\ \Leftrightarrow \quad F(z) &< q_i u_i'(H(z)q_i) \end{aligned}$$

Hence, summing across goods, we also have:

$$F(z) < \sum_i q_i u_i'(H(z)q_i)$$

We have shown that for each q there is a z such that $F(z) > \sum_i q_i u_i'(H(z)q_i)$ and a z such that $F(z) < \sum_i q_i u_i'(H(z)q_i)$. By continuity, we conclude that equation (31) has a solution.

Then, using part ii) of restrictions [A1], we can see that $D_i(F(\Lambda)y_i)/H(\Lambda)$ strictly decreases with Λ. This implies that $q_i u_i'(H(z)q_i)$ also strictly decreases with z, and that the right-hand side of equation (31), $\sum_i q_i u_i'(H(z)q_i)$, decreases faster (or increases slower) than the left-hand side of equation (31), $F(z)$. Hence the solution to equation (31) is unique.

Step 2: Quasi-concavity. The second step is to show that utility defined as $U(q) = \widetilde{U}(q, z(q))$ is quasi-concave. First, we need to get the first and second derivatives.

Derivatives in z. Here we consider the properties of $z = z(q)$, the solution of equation (31). Taking the derivative of equation (31), we get:

$$\frac{\partial z}{\partial q_i}\left[F' - H'\sum_i q_i^2 u_i''\right] = u_i' + Hq_i u_i''$$

and thus:

$$\frac{\partial z}{\partial q_i} = \frac{u_i' + Hq_i u_i''}{\Delta}$$

with $\Delta \equiv F' - H' \sum_i q_i^2 u_i''$.

Showing that Δ is positive. Note that $\frac{u_i'}{u_i'' Hq_i} = \varepsilon_{Di}$ and thus:

$$\begin{aligned} \Delta &= F' - H'\sum_i q_i^2 u_i'' \\ &= (F/z)\left(\varepsilon_F - \varepsilon_H \frac{\sum_i Hq_i^2 u_i''}{\sum_i q_i u_i'}\right) \\ &= (F/z)\left(\varepsilon_F - \varepsilon_H \frac{\sum_i q_i u_i'(1/\varepsilon_{Di})}{\sum_i q_i u_i'}\right) \end{aligned}$$

We can see that, if $\varepsilon_{Di} < 0$ and $\varepsilon_F \varepsilon_{Di} < \varepsilon_H$ for all i, then $\varepsilon_F - \varepsilon_H (1/\varepsilon_{Di}) > 0$ for all i and we always have $\Delta > 0$.

This implies that the derivatives of z are always well defined, and knowing $\Delta > 0$ will be useful below.

Derivatives in U. First derivatives:

$$\frac{\partial U}{\partial q_i} = H\,u_i'(Hq_i) + \frac{\partial z}{\partial q_i}\left[H'\sum_i q_i u_i'(Hq_i) - H'F\right] = H\,u_i'(Hq_i)$$

Second derivatives:

$$\begin{aligned}\frac{\partial^2 U}{\partial q_i^2} &= \frac{\partial z}{\partial q_i}\,(u_i' + Hq_i u_i'')H' \;+\; H^2\,u_i'' \\ \frac{\partial^2 U}{\partial q_i \partial q_j} &= \frac{\partial z}{\partial q_j}\,(u_i' + Hq_i u_i'')H'\end{aligned}$$

and thus, incorporating the derivatives in z:

$$\begin{aligned}\frac{\partial^2 U}{\partial q_i^2} &= (u_i' + Hq_i u_i'')^2\,H'/\Delta \;+\; H^2\,u_i'' \\ \frac{\partial^2 U}{\partial q_i \partial q_j} &= (u_i' + Hq_i u_i'')\,(u_j' + Hq_j u_j'')\,H'\,/\Delta\end{aligned}$$

Negative definiteness. To show that utility is quasi-convex, we need to show that the bordered Hessian is semi-definite negative, i.e we need to show:

$$\sum_{i,j} t_i t_j \frac{\partial^2 U}{\partial q_i \partial q_j} = \left(\sum_i t_i\,(u_i' + Hq_i u_i'')\right)^2 H'\,/\Delta \;+\; \sum_i t_i^2\,H^2\,u_i'' \;<\; 0$$

for any t_i such that:

$$\sum_i t_i \frac{\partial U}{\partial q_i} = \sum_i t_i H u_i' \;=\; 0$$

The objective function above is homogeneous of degree 2. We can renormalize the sum $\sum_i t_i(u_i' + Hq_i u_i'')$ up to any constant without loss of generalization.

First step is to find the optimal vector of t_i's that maximizes the left-hand side of the inequality above. It is equivalent to consider the maximization:

$$\max\left\{\sum_i t_i^2\,u_i''\right\}$$

under the constraint: $\sum_i t_i(u_i' + Hq_i u_i'') = \mathit{constant}$ and $\sum_i t_i u_i' \;=\; 0$.

This leads to t_i proportional to:

$$t_i \;\sim\; \frac{u_i'}{Hu_i''} + \mu q_i$$

(and second-order condition is fine, objective function is concave since $u_i'' < 0$ for all i). Given that we must have $0 = \sum_i t_i u_i' = \sum_i \frac{u_i'^2}{Hu_i''} + \mu \sum_i q_i u_i'$, μ must be:

$$\begin{aligned}
\mu &= -\frac{\sum_i \frac{u_i'^2}{Hu_i''}}{\sum_i q_i u_i'} \\
&= -\frac{\sum_i q_i u_i' \frac{u_i'}{q_i H u_i''}}{\sum_i q_i u_i'} \\
&= -\frac{\sum_i q_i u_i' \, \varepsilon_{Di}}{\sum_i q_i u_i'} \\
&= -\bar{\varepsilon}_D
\end{aligned}$$

where $\varepsilon_{Di} = \frac{u_i'}{q_i H u_i''}$ and $\bar{\varepsilon}_D$ is its weighted average.

Next, using the optimal $t_i = \frac{u_i'}{Hu_i''} - \bar{\varepsilon}_D q_i = q_i \varepsilon_{Di} - q_i \bar{\varepsilon}_D$, a sufficient and necessary condition for negative semi-definiteness is:

$$\left(\sum_i (q_i \varepsilon_{Di} - q_i \bar{\varepsilon}_D)(u_i' + q_i H u_i'') \right)^2 H' / \Delta \; + \; H^2 \sum_i (q_i \varepsilon_{Di} - q_i \bar{\varepsilon}_D)^2 \, u_i'' < 0$$

Since $\Delta > 0$, this condition can be rewritten:

$$\begin{aligned}
& \left(\sum_i (q_i \varepsilon_{Di} - q_i \bar{\varepsilon}_D)(u_i' + q_i H u_i'') \right)^2 H' \;<\; - \, H^2 \, \Delta \, \sum_i (q_i \varepsilon_{Di} - q_i \bar{\varepsilon}_D)^2 \, u_i'' \\
\Leftrightarrow & \left(\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'' \right)^2 H' \;<\; - \, H^2 \, \Delta \, \sum_i (q_i \varepsilon_{Di} - q_i \bar{\varepsilon}_D)^2 \, u_i'' \\
\Leftrightarrow & \left(\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'' \right)^2 H' \;<\; - \, H \, \Delta \, \left(-\bar{\varepsilon}_D \sum_i q_i u_i' + \bar{\varepsilon}_D^2 \, H \, \sum_i q_i^2 u_i'' \right) \\
\Leftrightarrow & \left(\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'' \right)^2 H' \;<\; \bar{\varepsilon}_D \; H \, \Delta \, \left(\sum_i q_i u_i' - \bar{\varepsilon}_D \, H \, \sum_i q_i^2 u_i'' \right)
\end{aligned}$$

Note that $(\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'')$ is negative (unless all price elasticities ε_{Di} are identical):

$$\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'' < 0 \iff \frac{\sum_i q_i u_i'}{\sum_i q_i u_i' \frac{1}{(-\varepsilon_{Di})}} < \frac{\sum_i q_i u_i' (-\varepsilon_{Di})}{\sum_i q_i u_i'}$$

(In the second inequality, the left hand side corresponds to a harmonic average while the right-hand-side corresponds to an arithmetic average of a positive variable $-\varepsilon_{Di} > 0$).

Hence, using $\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'' < 0$ and also that $\Delta \equiv F' - H' \sum_i q_i^2 u_i''$ the previous inequality is equivalent to:

$$
\begin{aligned}
\Leftrightarrow \quad & H' \left(\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'' \right) > \bar{\varepsilon}_D \ H \, \Delta \\
\Leftrightarrow \quad & H' \left(\sum_i q_i u_i' - \bar{\varepsilon}_D H \sum_i q_i^2 u_i'' \right) > \bar{\varepsilon}_D \ H \ \left(F' - H' \sum_i q_i^2 u_i'' \right) \\
\Leftrightarrow \quad & H' \sum_i q_i u_i' > \bar{\varepsilon}_D \ H \, F'
\end{aligned}
$$

Given that $F = \sum_i q_i u_i'$, this gives:

$$
\begin{aligned}
\Leftrightarrow \quad & H' \, F > \bar{\varepsilon}_D \ H \, F' \\
\Leftrightarrow \quad & \bar{\varepsilon}_D \varepsilon_F < \varepsilon_H
\end{aligned}
$$

This holds, given that $\varepsilon_{Di}\varepsilon_F < \varepsilon_H$ is assumed in part ii) of restrictions [A1] for each good i.

Step 3: Marshallian demand and price aggregator. Maximizing $U(q)$ under the budget constraint $\sum_i p_i q_i = w$ leads to:

$$
\frac{\partial U}{\partial q_i} \; = \; H(z) \, u_i'(H(z) q_i) \; = \; \mu p_i
$$

where μ henceforth denotes the Lagrange multiplier associated with the budget constraint. Summing across goods, we can see that μ is such that:

$$
\mu \; = \; \frac{1}{w} \sum_i \mu p_i q_i \; = \; \frac{1}{w} \sum H \, q_i \, u_i'(H q_i) \; = \; \frac{H(z) F(z)}{w}
$$

(note that it only depends on quantities and prices through z). Using $H(z) \, u_i'(H(z) q_i) \; = \; \mu p_i$, we obtain:

$$
u_i'(H(z) q_i) = \frac{\mu p_i}{H(z)} = \frac{F(z) p_i}{w}
$$

and thus, given the definition of u_i':

$$
H(z) q_i = D_i(\mu p_i / H(z)) = D_i(F(z) p_i / w)
$$

and:

$$
q_i = D_i(F(z) p_i / w) / H(z)
$$

The final step is to show that z can be written as $z = \Lambda$ where Λ is implicitly defined as a function of all normalized prices p_i / w.

To see this, notice that q_i must satisfy the budget constraint:

$$w = \sum_i q_i p_i = \sum_i p_i D_i(F(z)p_i/w)/H(z)$$

which can be rewritten:

$$H(z) = \sum_i (p_i/w) D_i(F(z)p_i/w)$$

This equation in z has a unique solution, which we denote Λ and is a function of all p_i/w:

- To prove uniqueness, we use restriction [A1] part ii) which implies that $D_i(F(z)p_i/w)/H(z)$ is strictly decreasing in z. Hence the right-hand side of the equation above decreases strictly faster with z (or increases strictly slower) than the left-hand side.

- To prove existence (for a given set of prices and income), we use restriction [A1] part iii) which assumes that $D_i(F(z)p_i/w)/H(z)$ can be arbitrarily small and that there is at least one good for which $D_i(F(z)p_i/w)/H(z)$ is larger for unity for some z. By continuity, there must be at least one solution $\Lambda = z$ to the equation $H(z) = \sum_i (p_i/w) D_i(F(z)p_i/w)$.

Alternative proof of Proposition 3 using the Slutsky Matrix

An alternative proof of proposition 3 is to show that the Slutsky matrix is symmetric and semi-definite negative, and then apply Hurwicz and Uzawa (1971) theorem. We have already proved symmetry in Proposition 1 but semi-definitiveness is yet to be checked. Consider the demand system:

$$q_i = D_i(F(\Lambda)p_i/w)\,/\,H(\Lambda) \tag{32}$$

where Λ is implicitly determined by the budget constraint $\sum_i p_i D_i(\Lambda p_i/w)\,/\,H(\Lambda) = w$, which can be rewritten:

$$\sum_i (p_i/w) D_i(F(\Lambda)p_i/w)\,/\,H(\Lambda) = 1 \tag{33}$$

Slutsky substitution coefficients. For $i \neq j$, the Slutsky term is:

$$\begin{aligned} s_{ij} &= \frac{\partial q_i}{\partial p_j} + q_j \frac{\partial q_i}{\partial w} \\ &= \frac{\partial q_i}{\partial \Lambda}\left[\frac{\partial \Lambda}{\partial p_j} + q_j \frac{\partial \Lambda}{\partial w}\right] - q_i q_j \frac{\varepsilon_{Di}}{w} \end{aligned}$$

where $\bar{\varepsilon}_D = \sum_i (p_i q_i/w)\varepsilon_{Di}$ and $\varepsilon_{Di} = \frac{\partial \log D_i}{\partial \log x}$. The first term is:

$$\frac{\partial q_i}{\partial \Lambda} = \frac{q_i(\varepsilon_F \varepsilon_{Di} - \varepsilon_H)}{\Lambda}$$

Then, we get the derivative of Λ w.r.t. p_j and w. From:

$$\sum_i (p_i/w) D_i(F(\Lambda)p_i/w)\,/\,H(\Lambda) = 1 \tag{34}$$

we get:

$$\begin{aligned} \frac{\partial \log \Lambda}{\partial \log p_j} &= -(p_j q_j/w)\, \frac{1+\varepsilon_{Dj}}{\varepsilon_F \bar{\varepsilon}_D - \varepsilon_H} \\ \frac{\partial \log \Lambda}{\partial \log w} &= \frac{1+\bar{\varepsilon}_D}{\varepsilon_F \bar{\varepsilon}_D - \varepsilon_H} \end{aligned}$$

Hence we obtain:

$$\begin{aligned} s_{ij} &= \frac{q_i(\varepsilon_F \varepsilon_{Di} - \varepsilon_H)}{\Lambda} \,.\, \Lambda \left[-(q_j/w)\, \frac{1+\varepsilon_{Dj}}{\varepsilon_F \bar{\varepsilon} - \varepsilon_H} + (q_j/w)\, \frac{1+\bar{\varepsilon}_D}{\varepsilon_F \bar{\varepsilon}_D - \varepsilon_H}\right] - q_i q_j \frac{\varepsilon_{Di}}{w} \\ &= \frac{q_i q_j}{w} \,.\, \frac{(\varepsilon_F \varepsilon_{Di} - \varepsilon_H)(\bar{\varepsilon} - \varepsilon_{Dj})}{\varepsilon_F \bar{\varepsilon}_D - \varepsilon_H} - q_i q_j \frac{\varepsilon_{Di}}{w} \\ &= -\frac{q_i q_j}{w} \,.\, \frac{\varepsilon_F(\varepsilon_{Di} - \bar{\varepsilon})(\varepsilon_{Dj} - \bar{\varepsilon}_D)}{\varepsilon_F \bar{\varepsilon}_D - \varepsilon_H} - q_i q_j \frac{\varepsilon_{Dj} + \varepsilon_{Di} - \bar{\varepsilon}}{w} \end{aligned}$$

which is symmetric in i and j.

When $\varepsilon_F \neq 0$, denote $\theta_i = \frac{\varepsilon_H - \varepsilon_F \varepsilon_{Di}}{\varepsilon_F}$, the elasticity of q_i in $1/F(\Lambda)$, and $\bar{\theta} = \sum_i (p_i q_i/w)\theta_i$ its weighted average across goods. It is more convenient to write the Slutsky substitution coefficient as:

$$\begin{aligned} s_{ij} &= \frac{q_i q_j}{w} \cdot \frac{(\theta_i - \bar{\theta})(\theta_j - \bar{\theta})}{\bar{\theta}} + \frac{q_i q_j}{w} \left(\theta_j + \theta_i - \bar{\theta} - \frac{\varepsilon_H}{\varepsilon_F} \right) \\ &= \frac{q_i q_j}{w} \cdot \frac{\theta_i \, \theta_j}{\bar{\theta}} - \frac{q_i q_j}{w} \frac{\varepsilon_H}{\varepsilon_F} \end{aligned}$$

In turn, the diagonal coefficients of the Slutsky matrix are:

$$\begin{aligned} s_{ii} &= \frac{q_i \varepsilon_{Di}}{p_i} - \frac{q_i q_j}{w} \cdot \frac{\theta_i \, \theta_j}{\bar{\theta}} - \frac{q_i q_j}{w} \frac{\varepsilon_H}{\varepsilon_F} \\ &= \frac{q_i \theta_i}{p_i} + \frac{q_i}{p_i} \frac{\varepsilon_H}{\varepsilon_F} - \frac{q_i q_j}{w} \cdot \frac{\theta_i \, \theta_j}{\bar{\theta}} - \frac{q_i q_j}{w} \frac{\varepsilon_H}{\varepsilon_F} \end{aligned}$$

What happens when $\varepsilon_F = 0$? When $\varepsilon_F = 0$, then $\varepsilon_H > 0$ by assumption, and we have instead:

$$\frac{\partial q_i}{\partial \Lambda} = -\frac{q_i \varepsilon_H}{\Lambda} \quad ; \quad \frac{\partial \log \Lambda}{\partial \log p_j} = (p_j q_j / w) \frac{1 + \varepsilon_{Dj}}{\varepsilon_H} \quad ; \quad \frac{\partial \log \Lambda}{\partial \log w} = -\frac{1 + \bar{\varepsilon}}{\varepsilon_H}$$

Hence we obtain:

$$s_{ij} = \frac{q_i \varepsilon_H}{\Lambda} \cdot \Lambda \left[-(q_j/w) \frac{1 + \varepsilon_{Dj}}{\varepsilon_H} + (q_j/w) \frac{1 + \bar{\varepsilon}}{\varepsilon_H} \right] - q_i q_j \frac{\varepsilon_{Di}}{w}$$

and thus the Slutsky substitution coefficiens are:

$$\begin{aligned} s_{ij} &= -\, q_i q_j \frac{\varepsilon_{Dj} + \varepsilon_{Di} - \bar{\varepsilon}}{w} \\ s_{ii} &= \frac{q_i \varepsilon_{Di}}{p_i} - q_i^2 \frac{2\varepsilon_{Di} - \bar{\varepsilon}}{w} \end{aligned}$$

Is the Slutsky matrix semi-definite negative? Denote $a_i = p_i q_i / w$ expenditure shares, we obtain:

$$\begin{aligned} s_{ij} \cdot p_i p_j / w &= \frac{a_i a_j \theta_i \theta_j}{\sum_k a_k \theta_k} - a_i a_j \frac{\varepsilon_H}{\varepsilon_F} \\ s_{ii} \cdot p_i^2 / w &= -a_i \theta_i + a_i \frac{\varepsilon_H}{\varepsilon_F} + \frac{a_i^2 \theta_i^2}{\sum_k a_k \theta_k} - a_i^2 \frac{\varepsilon_H}{\varepsilon_F} \end{aligned}$$

In order to show that the Slutsky matrix is definite negative, we need to show that for any vector x we have:

$$\sum_{ij} s_{ij} x_i x_j \leq 0$$

Normalizing the x's by $p_i/\sqrt{w}$, this is equivalent to showing the following inequality:

$$-\sum_i a_i \left(\theta_i - \frac{\varepsilon_H}{\varepsilon_F}\right) x_i^2 \ + \ \sum_{i,j} \frac{a_i x_i a_j x_j \theta_i \theta_j}{\sum_i a_i \theta_i} \ - \ \sum_{i,j} a_i a_j x_i x_j \left(\frac{\varepsilon_H}{\varepsilon_F}\right) \ \leq \ 0$$

$$\iff \ -\Big(\sum_i a_i\theta_i\Big)\left[\frac{\sum_i a_i\theta_i x_i^2}{\sum_i a_i\theta_i} \ - \ \frac{\left(\sum_i a_i\theta_i x_i\right)^2}{\left(\sum_i a_i\theta_i\right)^2}\right] \ + \ \frac{\varepsilon_H}{\varepsilon_F}\left[\sum_i a_i x_i^2 \ - \ \Big(\sum_i a_i x_i\Big)^2\right] \ \leq \ 0$$

$$\iff \ -\left[\frac{\sum_i a_i\theta_i x_i^2}{\sum_i a_i\theta_i} \ - \ \frac{\left(\sum_i a_i\theta_i x_i\right)^2}{\left(\sum_i a_i\theta_i\right)^2}\right] \ + \ \frac{\varepsilon_H}{\varepsilon_F \sum_i a_i\theta_i}\left[\sum_i a_i x_i^2 \ - \ \Big(\sum_i a_i x_i\Big)^2\right] \ \leq \ 0$$

The terms in brackets can be interpreted as variances and are positive. We need to distinguish two cases depending on the sign of $\frac{\varepsilon_H}{\varepsilon_F}$:
- First, if $\frac{\varepsilon_H}{\varepsilon_F} \leq 0$ (and $\theta_i > 0$), the sum is negative and the Slutsky matrix is semi-definite negative.
- Otherwise, if $\frac{\varepsilon_H}{\varepsilon_F} \geq 0$, the proof relies on Lemma 6 from (auxiliary result from Matsuyama and Uchshev, 2017). Recall that $\theta_i = \frac{\varepsilon_H}{\varepsilon_F} - \varepsilon_{Di}$ and that $\varepsilon_{Di} < 0$ is negative by assumption, hence $\theta_i > \frac{\varepsilon_H}{\varepsilon_F}$. Denote $b_i = \frac{\theta_i a_i}{\sum_j a_j \theta_j}$, and $\gamma = \frac{\varepsilon_H}{\varepsilon_F \sum_i a_i \theta_i}$. Since $\theta_i > \frac{\varepsilon_H}{\varepsilon_F}$, we have $b_i > \gamma a_i$ for all i. We obtain the result that we want by applying Lemma 6.

Is there a problem in the special case where $\varepsilon_F = 0$? In that case, we have:

$$\begin{aligned} s_{ij} \ . \ p_i p_j / w \ &= \ \ -\, a_i a_j \left(\varepsilon_{Dj} + \varepsilon_{Di} - \bar{\varepsilon}\right) \\ s_{ii} \ . \ p_i^2 / w \ &= \ a_i \varepsilon_{Di} \ - \ a_i^2 \left(2\varepsilon_{Di} - \bar{\varepsilon}\right) \end{aligned}$$

So we need:

$$\sum_i a_i x_i^2 \varepsilon_{Di} \ - \ \sum_{i,j} a_i a_j x_i x_j (\varepsilon_{Dj} + \varepsilon_{Di} - \bar{\varepsilon}) \ \leq \ 0$$

This statement is successively equivalent to:

$$\iff \sum_i a_i x_i^2 \varepsilon_{Di} \ - \ 2\sum_{i,j} a_i a_j x_i x_j \varepsilon_{Dj} + \bar{\varepsilon}\big(\sum_i a_i x_i\big)^2 \ \leq \ 0$$

$$\iff \sum_i a_i x_i^2 \varepsilon_{Di} \ - \ 2\big(\sum_i a_i x_i\big)\big(\sum_i a_i x_i \varepsilon_{Dj}\big) + \big(\sum_i a_i \varepsilon_{Dj}\big)\big(\sum_i a_i x_i\big)^2 \ \leq \ 0$$

$$\iff \frac{\sum_i a_i x_i^2 \varepsilon_{Di}}{\sum_i a_i \varepsilon_{Dj}} \ - \ \frac{2\big(\sum_i a_i x_i\big)\big(\sum_i a_i x_i \varepsilon_{Dj}\big)}{\sum_i a_i \varepsilon_{Dj}} + \big(\sum_i a_i x_i\big)^2 \ \geq \ 0$$

$$\iff \frac{\sum_i a_i \varepsilon_{Di}\Big(x_i^2 + \big(\sum_j a_j x_j\big)^2\Big)}{\sum_i a_i \varepsilon_{Dj}} \ \ \geq \ \frac{2\big(\sum_i a_i x_i\big)\big(\sum_i a_i x_i \varepsilon_{Dj}\big)}{\sum_i a_i \varepsilon_{Dj}}$$

$$\iff \frac{\sum_i a_i \varepsilon_{Di}\Big(x_i - \big(\sum_j a_j x_j\big)\Big)^2}{\sum_i a_i \varepsilon_{Dj}} \ + \ \frac{2\sum_i a_i \varepsilon_{Di} x_i \big(\sum_j a_j x_j\big)}{\sum_i a_i \varepsilon_{Dj}} \ \ \geq \ \frac{2\big(\sum_i a_i x_i\big)\big(\sum_i a_i x_i \varepsilon_{Dj}\big)}{\sum_i a_i \varepsilon_{Dj}}$$

$$\Longleftrightarrow \frac{\sum_i a_i \varepsilon_{Di}\Big(x_i - (\sum_j a_j x_j)\Big)^2}{\sum_i a_i \varepsilon_{Dj}} \;\geq\; 0$$

which is satisfied if all the ε_{Dj} are negative.

Lemma 6 *(Matsuyama Ushchev, 2017) Suppose that* $b_i > \gamma a_i > 0$ *for all* $i \in \{1, ..., N\}$ *where* γ *is a positive scalar and* $\sum_i a_i = \sum_i b_i = 1$. *Define* $\boldsymbol{A} = diag\{a_1, ..., a_n\} - \boldsymbol{a}\boldsymbol{a}^T$ *and* $\boldsymbol{B} = diag\{b_1, ..., b_n\} - \boldsymbol{b}\boldsymbol{b}^T$, *two (definite-negative) matrices. We obtain that the matrix* $\boldsymbol{M}$*:*

$$\boldsymbol{M} = \boldsymbol{B} - \gamma \boldsymbol{A}$$

is definite negative.

Proof of Lemma 6: This lemma is from Matsuyama and Ushchev (2017) which I report here again for convenience (see last part of the proof of Proposition 1 of Matsuyama and Ushchev 2017), switching the A and B notation.

One needs to show that, for each vector t of the $\mathbb{R}_{N+}$ with components $t_i \geq 0, \;\; i \in \{1, ..., N\}$:

$$t^T B\, t - \gamma \,.\, t^T A\, t \leq 0$$

Denote T_a and T_b the random variables such that: $Prob\{T_a = t_i\} = a_i$ and $Prob\{T_b = t_i\} = b_i$. Since $\sum_i a_i = \sum_i b_i = 1$, one can write each term above as a variance:

$$\begin{aligned} t^T A\, t &= \sum_i a_i t_i^2 - \left(\sum_i a_i t_i\right)^2 = Var(T_a) \\ t^T B\, t &= \sum_i b_i t_i^2 - \left(\sum_i b_i t_i\right)^2 = Var(T_b) \end{aligned}$$

Note that $Var(T_a) > 0$ unless $t_i = t_j$ for all i, j, in which case we also have $Var(T_b) = 0$ and thus $t^T B\, t - \gamma \,.\, t^T A\, t = 0$. Since $Var(T_a)$ is homogeneous of degree two and strictly positive aside from the case above, we can focus on t's such that $t^T A\, t = 1$. Under this assumption, we need to show that:

$$t^T B\, t \leq \gamma$$

Consider the maximization:

$$\max_t t^T B\, t \qquad s.t. \qquad t^T A\, t = 1$$

The maximum is attained when $Bt_i^* = \Lambda^* A t^*$ where Λ^* is the minimum value of the objective function, which can also be written:

$$b_i\Big(t_i^* - E[T_b^*]\Big) = \Lambda^* a_i\Big(t_i^* - E[T_a^*]\Big) \tag{35}$$

where $E[T_x]$ refers to the expectation of T_x.

The goal is to show that $\Lambda^* > \gamma$. To prove this claim by contradiction, suppose that $\Lambda^* < \gamma$. Given that $b_i > \gamma a_i$, we also have $b_i > \Lambda^* a_i$.

If $E[T_b] \leq E[T_a]$, we can see that:

$$\max t_i - E[T_b] \geq \max t_i - E[T_a] > 0$$

hence equation (35) cannot hold for $i = \arg\max t_j$.

If $E[T_b] > E[T_a]$, we can see that:

$$\min t_i - E[T_b] < \min t_i - E[T_a] < 0$$

hence, again, equation (35) cannot hold for $i = \arg\min t_j$. This yields a contradiction and proves Lemma 6.

Proof of Proposition 4

Suppose that demand can be written:

$$q_i = G_i(\Lambda)^{1-\sigma(\Lambda)} \, (p_i/w)^{-\sigma(\Lambda)}$$

with Λ implicitly defined by $\sum_i [G_i(\Lambda) p_i/w]^{1-\sigma(\Lambda)} = 1$.

The goal is to to show that these equations:

$$\left[\sum_i (G_i(\Lambda)\, p_i/w)^{1-\sigma(\Lambda)}\right]^{\frac{1}{1-\sigma(\Lambda)}} = 1 \qquad (36)$$

$$\left[\sum_i (G_i(U)/q_i)^{\frac{1-\sigma(U)}{\sigma(U)}}\right]^{\frac{\sigma(U)}{1-\sigma(U)}} = 1 \qquad (37)$$

have a unique solution in Λ and U respectively. To do so, we show that the left-hand side of each of these equations strictly increase in Λ and U around the solution, showing that the left-hand side can be equal to unity only once.

We distinguish two cases, depending on whether elasticity $\sigma(\Lambda)$ increases with Λ. If the first case we assume that $G_i(\Lambda)$ strictly increases with Λ. In the second case, we impose condition ii).

1) In the first case, suppose that $\sigma(\Lambda)$ increases with Λ and that $G_i(\Lambda)$ strictly increases with Λ. The equation above in Λ is equivalent to:

$$\sum_i (G_i(\Lambda)\, p_i/w)^{1-\sigma(\Lambda)} = 1$$

If $\sigma(\Lambda) \in (0, 1)$, each term $G_i(\Lambda)\, p_i/w$ in the summation increases in Λ and has to be smaller than unity. Hence, if $1 - \sigma(\Lambda)$ decreases with Λ, the left-hand side of this expression is strictly increasing with Λ. The same holds if we raise the whole expression on the left-hand side to the power $\frac{1}{1-\sigma(\Lambda)}$.

If $\sigma(\Lambda) > 1$, each term $G_i(\Lambda)\, p_i/w$ in the summation increases in Λ and has to be larger than unity. Hence, if $1 - \sigma(\Lambda)$ decreases with Λ (i.e. becomes more positive), the left-hand side of this expression is strictly decreasing in Λ. The inverse holds if we raise the whole expression on the left-hand side to the power $\frac{1}{1-\sigma(\Lambda)} < 0$.

Now consider the equation:

$$\sum_i \Big(G_i(U)/q_i\Big)^{\frac{1-\sigma(U)}{\sigma(U)}} = 1$$

If $\sigma(\Lambda) \in (0, 1)$, the exponent $\frac{1-\sigma(U)}{\sigma(U)}$ is positive and decreases with U. The term within parenthesis increases in U. Moreover, each summation term has to be smaller than unity. Hence, as U increases, each summation term increases (strictly) with U. The same holds if we raise the whole expression on the left-hand side to the power $\frac{\sigma(U)}{1-\sigma(U)}$.

If $\sigma(\Lambda) > 1$, the exponent $\frac{1-\sigma(U)}{\sigma(U)}$ is negative and decreases with U. The term within parenthesis increases in U. Moreover, each summation term has to be larger than unity. Hence, as U increases, each summation term decreases (strictly) with U. If we raise the whole expression on the left-hand side to the power $\frac{\sigma(U)}{1-\sigma(U)}$, we obtain a strictly increasing function of U.

2) In the second case, we assume that $\sigma(\Lambda)$ decreases with Λ and that, around each solution Λ_0 of equation (36), there exists a set of α_i such that $\sum_i \alpha_i = 1$ and such that $G_i(\Lambda)\alpha_i^{-\frac{1}{1-\sigma(\Lambda)}}$ increases in Λ.

Define $K_i(\Lambda) = G_i(\Lambda)\alpha_i^{-\frac{1}{1-\sigma(\Lambda)}}$ The left-hand side of equation (36) can then be rewritten:

$$\left[\sum_i \alpha_i (K_i(\Lambda)\, p_i/w)^{1-\sigma(\Lambda)}\right]^{\frac{1}{1-\sigma(\Lambda)}}$$

To show that it strictly increases in Λ, we use Lemma 7 discussed in the next appendix section. We obtain that the left-hand side of the above equation decreases with σ, which itself decreases with Λ. Moreover, the term $K_i(\Lambda)$ strictly increases in Λ, by assumption, hence the whole left term strictly increases with Λ.

We can again use the same approach to show that the left-hand side of (37) increases strictly with U. This is equivalent to showing that the following expression strictly increases in U:

$$\left[\sum_i \alpha_i \Big(K_i(U)/q_i\Big)^{\frac{1-\sigma(U)}{\sigma(U)}}\right]^{\frac{\sigma(U)}{1-\sigma(U)}}$$

Each exponent $\frac{1-\sigma(U)}{\sigma(U)}$ increases in U and each term $K_i(U)$ strictly increases with U. With Lemma 7 again, we obtain that the whole term strictly increases with U.

Hence, in both cases, Λ and U are well defined by equations (36) and (37) which admit no more than one solution. This implicitly defines utility U as a function of q_i. It is straightforward to see that such utility function is quasi-concave in q: indifference curves have the same shape as CES indifference curves, holding $\sigma = \sigma(U)$ constant.

Consumption quantities q chosen to maximize U would satisfy the following first-order conditions:

$$\frac{(\sigma(U)-1)}{q_i\sigma(U)}\left(\frac{q_i}{G_i(U)}\right)^{\frac{\sigma(U)-1}{\sigma(U)}} = \mu p_i$$

where μ is a constant term (combination of the Lagrange multiplier associated with the equation in U and the budget constraint multiplier). To satisfy the budget constraint, $\frac{(\sigma(U)-1)\mu}{\sigma(U)}$ has to equal $1/w$. In other words, $\left(\frac{q_i}{G_i(U)}\right)^{\frac{\sigma(U)-1}{\sigma(U)}}$ corresponds to the budget share of good i in consumption baskets:

$$\left(\frac{q_i}{G_i(U)}\right)^{\frac{\sigma(U)-1}{\sigma(U)}} = \frac{(\sigma(U)-1)\mu}{\sigma(U)}p_i q_i = \frac{p_i q_i}{w}$$

This leads to the demand q_i:

$$q_i = G_i(U)^{1-\sigma(U)}\,(p_i/w)^{-\sigma(U)}$$

which is the same expression as above, with Λ corresponding to utility. Moreover, we can see that utility U is such that $\sum_i \left(\frac{q_i}{G_i(U)}\right)^{\frac{\sigma(U)-1}{\sigma(U)}} = 1$ which, using the demand for q_i just above, can be written as:

$$\sum_i [G_i(U)p_i/w]^{1-\sigma(U)} = 1$$

which is the same equation as the one determining Λ, which proves that $\Lambda = U$.

Proof of equivalence between condition ii) and inequality (14)

We mention in the text that condition ii) of Proposition 4 is equivalent to inequality (14) when both σ and G_i are differentiable.

Taking the derivative of the log of $G_i(\Lambda)\alpha_i^{-\frac{1}{1-\sigma(\Lambda)}}$ with respect to Λ, we find that it is positive if and only if:

$$\frac{G_i'(\Lambda)}{G_i(\Lambda)} \;-\; (\log \alpha_i)\,.\,\frac{\partial}{\partial \Lambda}\left(\frac{1}{1-\sigma(\Lambda)}\right) > 0$$

Hence, for each good i, the minimum α_i such that it is positive is:

$$\alpha_i^* = \exp\left(\frac{(\sigma(\Lambda)-1)^2 G_i'(\Lambda)}{\sigma'(\Lambda)G_i(\Lambda)}\right)$$

One can see that inequality $\sum_i \alpha_i^* < 1$ corresponds to inequality (14) in the text.

Note: one can also verify that this condition is equivalent to imposing that $G_i(\Lambda)$ and $\sigma(\Lambda)$ are such that:

$$\left[\sum_i (G_i(\Lambda)\,p_i/w)^{1-\sigma(\Lambda)}\right]^{\frac{1}{1-\sigma(\Lambda)}}$$

increases for any set of y_i.

Lemma 7 *For any given set of $x_i \geq 0$ and $\alpha_i \geq 0$ such that $\sum_i \alpha_i$, the following expression is monotonically increasing in $\rho \in (-\infty, +\infty)$:*

$$\Big[\sum_i \alpha_i x_i^{\rho}\Big]^{\frac{1}{\rho}}$$

Proof of Lemma 7: First, consider two values $\rho < \rho' < 0$ and consider the mapping $m(x) = x^{\frac{\rho'}{\rho}}$ which is convex in x. Jensen's inequality implies that:

$$m\Big(\sum_i \alpha_i y_i\Big) \leq \sum_i \alpha_i m(y_i)$$

and thus:

$$\Big(\sum_i \alpha_i y_i\Big)^{\frac{1}{\rho}} \leq \Big(\sum_i \alpha_i y_i^{\frac{\rho'}{\rho}}\Big)^{\frac{1}{\rho'}}$$

Choosing $y_i = [x_i]^{\rho}$, we obtain:

$$\Big[\sum_i \alpha_i x_i^{\rho}\Big]^{\frac{1}{\rho}} \leq \Big[\sum_i \alpha_i x_i^{\rho'}\Big]^{\frac{1}{\rho'}}$$

Second, consider two values $\rho' > \rho > 0$ and consider again the mapping $m(x) = x^{\frac{\rho'}{\rho}}$ which is now concave in x. Jensen's inequality for concave functions implies:

$$m\Big(\sum_i \alpha_i y_i\Big) \geq \sum_i \alpha_i m(y_i)$$

and thus, taking to the exponent $1/\rho < 0$, we have:

$$\Big(\sum_i \alpha_i y_i\Big)^{\frac{1}{\rho}} \leq \Big(\sum_i \alpha_i y_i^{\frac{\rho'}{\rho}}\Big)^{\frac{1}{\rho'}}$$

Choosing $y_i = [x_i]^{\rho}$, we obtain:

$$\Big[\sum_i \alpha_i x_i^{\rho}\Big]^{\frac{1}{\rho}} \leq \Big[\sum_i \alpha_i x_i^{\rho'}\Big]^{\frac{1}{\rho'}}$$

Note that these terms are well defined when ρ converges to zero (on both sides):

$$\lim_{\rho \to 0} \Big[\sum_i \alpha_i x_i^{\rho}\Big]^{\frac{1}{\rho}} = \prod_i x_i^{\alpha_i}$$

hence the findings above also apply to $\rho = 0$. This proves Lemma 7.

Appendix B – Counter-examples

First case with homogeneous demand shifters

Here I show that we can find a case where conditions ii) fails and where the Slutsky substitution matrix is not semi-definite negative, thus proving that condition ii) cannot be entirely waived.

Suppose that $F(\Lambda) = \Lambda$ (no problem arises when F is locally constant) and that we have two goods 1 and 2, where $\varepsilon_{D1} < \varepsilon_H$ while $\varepsilon_{D2} > \varepsilon_H$ for the other good, i.e. $\varepsilon_H \in (\varepsilon_{D1}, \varepsilon_{D2})$. In particular, to fix ideas, supposed that all elasticities are constant, with $\varepsilon_H = \frac{\varepsilon_{D2}+\varepsilon_{D1}}{2} \equiv -\kappa < 0$ and denote $\delta \equiv \varepsilon_{D2} - \varepsilon_H = \varepsilon_H - \varepsilon_{D1} > 0$. Denote by the expenditure share of product 1 as $\frac{1-\epsilon}{2}$ and the expenditure share of good 2 as $\frac{1+\epsilon}{2}$ such that $\bar{\varepsilon}_D - \varepsilon_H = \epsilon\delta$. While elasticities are constant, we can still adjust the demand shifter for each good to obtain the desired market shares (hence ϵ can be chosen independently from the elasticities).

The off-diagonal coefficients of the Slutsky substitution matrix are then:

$$s_{12}p_1p_2/w \;=\; -\frac{a_1a_2(\varepsilon_{D1}-\varepsilon_H)(\varepsilon_{D2}-\varepsilon_H)}{\bar{\varepsilon}_D - \varepsilon_H} + a_1a_2\varepsilon_H \;=\; -\frac{(1-\epsilon^2)\delta^2}{4\epsilon\delta} - \frac{(1-\epsilon^2)\kappa}{4}$$

where a_i denotes the expenditure share of good i. The diagonal coefficients are:

$$s_{11}p_1^2/w \;=\; a_1\varepsilon_{D1} - \frac{a_1^2(\varepsilon_{D1}-\varepsilon_H)^2}{\bar{\varepsilon}_D - \varepsilon_H} + a_1^2\varepsilon_H \;=\; -\frac{(1-\epsilon)(\kappa+\delta)}{2} + \frac{(1-\epsilon)^2\delta^2}{4\epsilon\delta} - \frac{(1-\epsilon)^2\kappa}{4}$$

$$s_{22}p_2^2/w \;=\; a_2\varepsilon_{D2} - \frac{a_2^2(\varepsilon_{D2}-\varepsilon_H)^2}{\bar{\varepsilon}_D - \varepsilon_H} + a_2^2\varepsilon_H \;=\; -\frac{(1+\epsilon)(\kappa-\delta)}{2} + \frac{(1+\epsilon)^2\delta^2}{4\epsilon\delta} - \frac{(1+\epsilon)^2\kappa}{4}$$

One can see that the substitution coefficients become very large as ϵ approach zero (because some of the terms have ϵ in the denominator). Moreover, if we denote by Σ the matrix with coefficients $s_{ij}p_ip_j/w$, we obtain:

$$\lim_{\epsilon\to 0+} 4\epsilon\Sigma = \begin{pmatrix} +\delta & -\delta \\ -\delta & +\delta \end{pmatrix}$$

This matrix is semi-definite positive: $x^T 4\epsilon\Sigma x = \delta^2(x_1 - x_2)^2 \geq 0$. By continuity, when ϵ is small enough, the substitution matrix with coefficient s_{ij} is semi-definite positive, which is not consistent with a rational demand system.

Second case with iso-elastic substitution

In this case, I provide counter-examples to show that neither Λ or U are well defined if the assumptions of Proposition 4 are not satisfied.

- First, suppose that $\sigma(\Lambda)$ increases in Λ. In this case, the elasticity of substitution increases with income and issues are more likely to arise when consumption is concentrated in one or few goods.

When $G_i(\Lambda)$ is not monotonic in Λ for a good i, the budget constraint can be written:

$$G_i(\Lambda)p_i/w = 1$$

when the consumption of all other goods become negligible, i.e. when $(p_j/w)^{1-\sigma(\Lambda)} = 0$. If there exists $\Lambda_1 \neq \Lambda_2$ such that $G_i(\Lambda_1) = G_i(\Lambda_2)$, one can see that the equation above has at least two solutions when $p_i/w = 1/G_i(\Lambda_1)$.

Conversely, utility is not well defined by the implicit equation provided in Proposition 4 when G_i is not monotonic for a good. Suppose that $q_j^{\frac{\sigma(U)-1}{\sigma(U)}}$ is zero (or close to zero) for other goods j. In that case, we can see that $\left(\frac{q_i}{G_i(U)}\right)^{\frac{\sigma(U)-1}{\sigma(U)}} = 1 \Leftrightarrow G_i(U) = q_i$ has several solutions in U for some q_i if G_i is not monotonic, potentially violating the monotonicity of U w.r.t quantities.

We also need G'_i to have the same sign for all goods. If it is not the case, we can obtain situations where Λ and U are not well defined, or where U would decrease with quantities q_i for some goods.

- Counter-examples for the second case are more difficult to construct. Here we will assume here that $\sigma(\Lambda)$ and $G_i(\Lambda)$ are differentiable. Let us examine what happens when inequality (14) is not satisfied, i.e. when:

$$\sum_i \exp\left(\frac{(\sigma(\Lambda)-1)^2 G'_i(\Lambda)}{\sigma'(\Lambda)G_i(\Lambda)}\right) > 1$$

for a given $\Lambda = U_0$. In that case, we can show that it is possible to find a set of quantities q_i such that U_0 is the solution of equation (13) but where implicit utility would depend negatively on quantities. This amounts to showing that the following expression:

$$\left[\sum_i \left(G_i(U)/q_i\right)^{\frac{1-\sigma(U)}{\sigma(U)}}\right]^{\frac{\sigma(U)}{1-\sigma(U)}}$$

decreases with U for some q_i's (while it also decreases in each q_i).

Suppose that U_0 is the solution of equation (13) for a given set of q_i. We can always rearrange the q_i to match a given set of consumption shares while still having U_0 as the solution of equation (13) . In particular, choose q_i^* such that U_0 is still the solution of (13) and such that:

$$\left(G_i(U_0)/q_i^*\right)^{\frac{1-\sigma(U_0)}{\sigma(U_0)}} = \frac{1}{A}\exp\left(\frac{(\sigma(U_0)-1)^2 G'_i(U_0)}{\sigma'(U_0)G_i(U_0)}\right)$$

where $A \equiv \sum_i \exp\left(\frac{(\sigma(U_0)-1)^2 G'_i(U_0)}{\sigma'(U_0)G_i(U_0)}\right) > 1$, strictly larger than unity if condition ii) is not satisfied. Consider the function:

$$f(U,q) = \left[\sum_i \Big(G_i(U)/q_i\Big)^{\frac{1-\sigma(U)}{\sigma(U)}}\right]^{\frac{\sigma(U)}{1-\sigma(U)}}$$

which corresponds to the left-hand side of equation (13). One can see that the derivative in U at $U = U_0$ and $q = q^*$ is negative:

$$\begin{aligned} f_U(U_0,q^*) &= \sum_i \frac{G_i'(U_0)}{G_i(U_0)}\Big(\frac{G_i(U_0)}{q_i^*}\Big)^{\frac{1-\sigma(U_0)}{\sigma(U_0)}} + \frac{\sigma'(U_0)}{(1-\sigma(U_0))^2}\sum_i \Big(\frac{G_i(U_0)}{q_i^*}\Big)^{\frac{1-\sigma(U_0)}{\sigma(U_0)}} \log\Big(\frac{G_i(U_0)}{q_i^*}\Big)^{\frac{1-\sigma(U_0)}{\sigma(U_0)}} \\ &= \frac{\sigma'(U_0)}{(1-\sigma(U_0))^2}\log A \;<\; 0 \end{aligned}$$

while the derivative $f_{q_i}(U_0,q^*)$ in each q_i is also negative. This leads to an implicit utility function U of q that decreases with quantities.

Appendix C – Application to Trade, Arkolakis et al (2015)

In this application, we use the demand system described in Section 2.2.1 and in applications 3 and 9 of Section 4. In particular, we consider a symmetric demand $D(\cdot)$ across a continuum of product varieties. We adopt the notation of Arkolakis et al (henceforth ACDR) and refer to it for more details on the general setting. Instead of referring to goods, subscripts i and j refer to countries. We pick a specific country j and normalize its income to unity: $w_j = 1$ (wages may differ from unity for other countries).

Denote $P_j = 1/\Lambda_j$ the price aggregator in destination j and normalize the demand curve such that: $D(1) = 0$. To borrow from the ACDR notation, we denote the demand shifter by Q instead of $1/H$ (country specific but common across product varieties).

Let us assume that demand for a specific variety is a function:

$$q = Q(w_j/P_j)D(p/P_j)$$

with: $Q(w/P_j) \int pD(p/P_j) = w_j$, summing across a continuum of varieties. As in the discussion on choke prices in example 9, let us further assume that Q_j is iso-elastic, i.e.:

$$Q_j(w_j/P_j) = (w_j/P_j)^{\kappa}$$

(where $\kappa = 1$ for homothetic demand). With the normalization $D(1) = 0$ and wages $w_j = 1$ normalized to unity in country j, we obtain a productivity cutoff: $z^*_{ij} = \tau_{ij}/P_j$ where τ_{ij} denotes the bilateral iceberg trade costs: firms with marginal cost above $1/z^*_{ij}$ would face zero demand in country j if they also incur bilateral iceberg trade costs τ_{ij}.

As in ACDR, we obtain a gravity equation if we assume that productivity (inverse of marginal cost) is drawn from Pareto distributions. This leads to aggregate trade between countries i and j:

$$X_{ij} = \chi N_i b_i^{\theta}(w_i\tau_{ij})^{-\theta} L_j Q_j P_j^{1+\theta} \tag{38}$$

where χ_1 is a constant term, N_i is the mass of firms in country i, b_i parameterizes average productivity in country i, and L_j is the mass of consumers in the destination country. As in ACDR, we also find that aggregate profits are proportional to aggregate trade: $\Pi_{ij} = \zeta X_{ij}$ with $\zeta < 1$ (proportionality holds even if we allow for variable markups).

Note that, in equilibrium, the budget constraint (equivalent to trade balance) imposes: $\sum_i X_{ij} = L_j w_j = L_j$. As in ACDR, summing equation (38) across source countries, the budget constraint yields a condition that takes the form:

$$\chi Q_j P_j^{\theta+1}\Big(\sum_i N_i b_i^{\theta}(w_i\tau_{ij})^{-\theta}\Big) = 1$$

If Q is iso-elastic (i.e. $Q_j = P_j^{-\kappa}$), we obtain:

$$\chi P_j^{\theta+1-\kappa}\Big(\sum_i N_i b_i^{\theta}(w_i\tau_{ij})^{-\theta}\Big) = 1$$

where the special cases $\kappa \in \{0, 1\}$ correspond to the cases $\beta \in \{0, 1\}$ in ACDR. Hence:

$$d \log P_j = -\frac{\theta}{\theta + 1 - \kappa} \sum_i \lambda_{ij} \, d \log(w_i \tau_{ij})$$

(where λ_{ij} refers to import shares) and gains from trade satisfying the same formula as in ACDR, with κ playing the same role as parameter $\beta \in \{0, 1\}$ in ACDR (equal to 1 if preferences are homothetic):

$$\eta = \frac{(1 - \kappa) \, \rho}{\theta + 1 - \kappa}$$

now with a continuous representation (κ instead of β) that can take any value between 0 and 1. We can also have $\kappa > 1$ such that $\eta < 0$ even if $\rho > 0$.

References

Antonelli, G. B. (1886). *Sulla teoria matematica della economia politica.*

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2015). The elusive pro-competitive effects of trade.

Bertoletti, P. and F. Etro (2017a). Monopolistic competition, as you like it. *University Ca' Foscari of Venice, Dept. of Economics Research Paper Series, 08/2017.*

Bertoletti, P. and F. Etro (2017b). Monopolistic competition when income matters. *The Economic Journal 127*(603), 1217–1243.

Bertoletti, P., F. Etro, and I. Simonovska (2016). International trade with indirect additivity. *National Bureau of Economic Research WP No. 21984*.

Blackorby, C., D. Primont, and R. R. Russell (1978). *Duality, separability, and functional structure: Theory and economic applications*, Volume 2. Elsevier Science Ltd.

Caron, J., T. Fally, and J. Markusen (2017). Per capita income and the demand for skills. *National Bureau of Economic Research WP No. 23482*.

Caron, J., T. Fally, and J. R. Markusen (2014). International trade puzzles: A solution linking production and preferences. *The Quarterly Journal of Economics 129*(3), 1501–1552.

Deaton, A. (1974). A reconsideration of the empirical implications of additive preferences. *The Economic Journal 84*(334), 338–348.

Faber, B. and T. Fally (2017). Firm heterogeneity in consumption baskets: Evidence from home and store scanner data.

Fieler, A. C. (2011). Nonhomotheticity and bilateral trade: Evidence and a quantitative explanation. *Econometrica 79*(4), 1069–1101.

Gorman, W. M. (1981). Some engel curves. *Essays in the theory and measurement of consumer behaviour in honor of sir Richard Stone*.

Gorman, W. M. (1987). Separability. *The New Palgrave: A Dictionary of Economics*, London: Macmillan Press, 4, 305–11.

Gorman, W. M. (1995). *Collected Works of WM Gorman: Separability and Aggregation*, Volume 1. Oxford University Press.

Handbury, J. (2013). Are poor cities cheap for everyone? non-homotheticity and the cost of living across us cities. *Mimeograph, Wharton University*.

Houthakker, H. S. (1965). A note on self-dual preferences. *Econometrica: Journal of the Econometric Society*, 797–801.

Hurwicz, L. and H. Uzawa (1971). On the integrability of demand functions. *Preferences, utility, and demand*, 114–148.

LaFrance, J. T. and R. D. Pope (2006). Full rank rational demand systems. *Department of Agricultural & Resource Economics, UCB*.

Lewbel, A. (1991). The rank of demand systems: theory and nonparametric estimation. *Econometrica: Journal of the Econometric Society*, 711–730.

Lewbel, A. (2010). Shape-invariant demand functions. *The Review of Economics and Statistics 92*(3), 549–556.

Lewbel, A. and K. Pendakur (2009). Tricks with hicks: The easi demand system. *The American Economic Review 99*(3), 827–863.

Ligon, E. (2016). All λ-separable demands and rationalizing utility functions. *Economics Letters 147*, 16–18.

Matsuyama, K. and P. Ushchev (2017). "Beyond CES: Three alternative classes of flexible homothetic demand systems". *Working paper*.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2014). Market size, competition, and the product mix of exporters. *The American Economic Review 104*(2), 495–536.

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *The Review of Economic Studies 75*(1), 295–316.

Mrázová, M. and J. P. Neary (2013). Not so demanding: Preference structure, firm behavior, and welfare. *The American Economic Review (forthcoming)*.

Mrazova, M., J. P. Neary, and M. Parenti (2017). Sales and markup dispersion: Theory and empirics. *CEPR Discussion Paper No. DP12044*.

Nocke, V. and N. Schutz (2017). Quasi-linear integrability. *Journal of Economic Theory 169*, 603–628.

Ottaviano, G., T. Tabuchi, and J.-F. Thisse (2002). Agglomeration and trade revisited. *International Economic Review*, 409–435.

Pollak, R. A. (1972). Generalized separability. *Econometrica: Journal of the Econometric Society*, 431–453.

Preckel, P. V., J. Cranfield, T. W. Hertel, et al. (2005). Implicit additive preferences: A further generalization of the ces. In *Annual Meetings of the American Agricultural Economics Association*, pp. 24–27.

Weyl, E. G. and M. Fabinger (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy 121*(3), 528–583.