DISCUSSION PAPER SERIES

DP17249

Generalized Separability and Integrability: Consumer Demand with a Price Aggregator

Thibault Fally

INDUSTRIAL ORGANIZATION

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

ISSN 0265-8003

Generalized Separability and Integrability: Consumer Demand with a Price Aggregator

Thibault Fally

Discussion Paper DP17249
Published 24 April 2022
Submitted 21 April 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Generalized Separability and Integrability: Consumer Demand with a Price Aggregator

Abstract

This paper examines demand systems where the demand for a good depends on other prices only through a common price aggregator (a scalar function of all prices). We refer to this property as ``generalized separability'' and provide the functional forms of demand that this property implies when demand is rational, i.e., derived from utility maximization. Generalized separability imposes restrictions on either income or price effects, and greater flexibility is obtained by adding indirect utility as an additional aggregator. We provide examples and applications which encompass a large variety of examples from the literature. In particular, generalized separability can be used in simple general-equilibrium models to obtain a more tractable framework and yet generate a wider range of effects of market size and productivity on firm size, entry, and prices.

JEL Classification: D11, D40, L13

Keywords: Consumer demand, Separability, Price aggregator, integrability, Rationalization, non-homothetic preferences

Thibault Fally - fally@berkeley.edu
UC Berkeley and CEPR

Generalized Separability and Integrability: Consumer Demand with a Price Aggregator

Thibault Fally*
UC Berkeley ARE

April 2022

Abstract

This paper examines demand systems where the demand for a good depends on other prices only through a common price aggregator (a scalar function of all prices). We refer to this property as "generalized separability" and provide the functional forms of demand that this property implies when demand is rational, i.e., derived from utility maximization. Generalized separability imposes restrictions on either income or price effects, and greater flexibility is obtained by adding indirect utility as an additional aggregator. We provide examples and applications which encompass a large variety of examples from the literature. In particular, generalized separability can be used in simple general-equilibrium models to obtain a more tractable framework and yet generate a wider range of effects of market size and productivity on firm size, entry, and prices.

Keywords: Consumer Demand, Separability, Price aggregator, Integrability, Rationalization, Non-homothetic preferences.
JEL Classification: D11, D40, L13

*University of California 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, Peter Neary, Andres Rodriguez-Clare, Philip Ushchev, Brian Wright, Xavier Vives (Editor), and two anonymous referees for very helpful comments and suggestions. Contact: fally@berkeley.edu

1 Introduction

The integrability problem, which consists of characterizing demand systems that can be rationalized and derived from utility functions, has long been a central issue in economic theory. The earliest contributions date from Antonelli (1886), with 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 negative semi-definite for all prices and income levels.

While very general, the Hurwicz and Uzawa (1971) integrability conditions lack practicality. Perhaps a consequence is that applied theorists and practitioners have often focused on less general cases to ensure both tractability and rationality. In particular, one often focuses on directly or indirectly additive preferences. An attractive feature of these preferences is that demand 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 an aggregator can be, for instance, a price index (e.g. with constant elasticity of substitution preferences) or the marginal utility of income (with directly-separable preferences).[1] These preferences, however, have properties that may be undesirable and too restrictive in terms of income and price effects. For instance, direct separability implies that income elasticities and price elasticities are proportional across goods ("Pigou's law"), a testable prediction that has been empirically rejected, e.g., by Deaton (1974).

This paper characterizes demand systems that are more general but retain a key practical property of the widely-used demand systems mentioned above: the existence of a price aggregator that is common for all goods, a feature that is useful for demand estimation, welfare analysis, applied models of monopolistic and oligopolistic competition, and many other applications.

The paper aims to make three contributions. A first objective is to provide functional forms of demand (i.e. necessary conditions) to satisfy Slutsky symmetry when demand for a good depends only on its own price, income, and a common price aggregator. We also consider cases where demand depends on utility in addition to the price aggregator. A second objective is to provide sufficient conditions for such functional forms of demand to be rational, i.e. such that they can be derived from a well-defined quasi-concave utility function. A third contribution is to

[1] In models with symmetric demand across product varieties with an upper bound in marginal utility for each variety, there exists a finite reservation price (or choke price) that can also be used as a common price aggregator (see e.g. Arkolakis et al., 2019).

provide various examples of such demand systems, including some that have not been previously discussed in the literature, and to illustrate how functional forms of demand determine market size effects on firm size and prices in a simple general-equilibrium model.

Following Pollak (1972), we first consider "generalized separable" demand systems as those that satisfy:

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

for each good i, where p_i refers to its price, w to consumer income (total outlays), and Λ is a scalar (aggregator) that is a function of all prices and income. A key property of such demand is that all cross-price effects operate through Λ, a practical property for modeling and estimation, because the rank of the cross-price substitution matrix is then just one.

In fact, such a demand system can only take some specific functional forms in order to be integrable. Providing the sketch of a proof that we complete here, Gorman (1972, 1995) indicates that such a demand system can take either of two main forms[2] if we impose symmetry on the Slutsky substitution matrix:

$$q_i = \frac{D_i(F(\Lambda)p_i/w)}{H(\Lambda)} \tag{2}$$

$$q_i = A_i(\Lambda)(p_i/w)^{-\sigma(\Lambda)} \tag{3}$$

where D_i, F and H are positive real functions and where, in both cases, Λ is a scalar variable that adjusts so that the budget constraint is satisfied, and can thus be defined as an implicit function of prices and income (under additional assumptions on differentiability and invertibility). The general forms of these demand systems have rarely appeared in the applied literature so far in spite of their usefulness.[3]

In the first case (equation 2), sufficient conditions for integrability are expressed as conditions on elasticities of functions H, F and D_i, and ensure that demand q_i is decreasing in the aggregator Λ for any good i. We will refer to this case as a "Gorman-Pollak" demand system. It corresponds to directly-additive utility (used e.g. in Krugman, 1979) when the quantity shifter $H(\Lambda)$ is constant; it corresponds to indirectly-additive utility when the price shifter $F(\Lambda)$ is constant. This also generalizes the results of Matsuyama and Ushchev (2017) on homothetic single-aggregator demand ("HSA", corresponding to $F(\Lambda) = 1/H(\Lambda) = \Lambda$). This type of preferences can be used to rationalize many examples drawn from Mrázová and Neary (2013), e.g. bi-power and inverse bi-power demand functions, as well as Bulow-Pfleiderer

[2] There are other cases that can be ruled out under additional restrictions on price sensitivity.

[3] There are a few recent exceptions, including Bertoletti and Etro (2022, 2021) for the first case, Comin et al. (2021) and Matsuyama (2019) for the second case with homogeneous $\sigma(\Lambda) = \sigma$.

demand (Weyl and Fabinger, 2013). For instance, with iso-elastic functions H and D_i, the demand system is "self-dual addilog" as described in Houthakker (1965). In this more general demand system, income and price elasticities both depend on the functional form chosen for D_i, which can be very flexible; demand and price shifters $H(\Lambda)$ and $F(\Lambda)$ also influence income effects and depend flexibly on the price aggregator. However, this formulation still imposes some constraints on price and income effects, as it implies an affine relationship between price and income elasticities of demand across goods for a given consumer.

In the second case (equation 3), with common price elasticities across goods, the aggregator Λ coincides with indirect utility V (up to a one-to-one mapping). In that case, integrability requires that the demand shifters $A_i(\Lambda)$ increase quickly enough in Λ. While quasi-concavity is easy to obtain in this case, conditions for rationalization need to ensure that indifference curves do not cross and that utility is monotonically increasing in the consumption of each good. Notice that the price elasticity $\sigma(\Lambda)$ does not have to remain constant or monotonic across indifference curves, i.e. indifference curves can become flatter or more convex as income goes up. We will refer to that case as "generalized non-homothetic CES". This second case features Allen-Uzawa substitution elasticities that do not vary across goods but may vary with utility. Relative to Gorman-Pollak demand, this case allows for more flexible income patterns, but requires rigid price effects. It generalizes the implicitly-additive utility functions used by (Comin et al., 2021) (who impose a constant elasticity of substitution $\sigma(\Lambda) = \sigma$) to model structural change and sector-specific Engel curves across agriculture, manufacturing, and services. A similar demand structure is used in Atkin et al. (2020) to estimate welfare and price indices from shifts in Engel curves. With an elasticity of substitution that depends on utility, these preferences remain very tractable and empirically relevant. Several studies (such as Handbury, 2021, Faber and Fally, 2020 and Auer et al., 2021) based on expenditure surveys and scanner data have shown that price elasticities vary significantly with income.[4]

These two types of demand systems are appealing for their simplicity and tractability if we focus on either price or income effects. However, greater flexibility can be obtained by allowing demand to depend on utility in addition to the common aggregator Λ. This can be seen as a combination of the two cases discussed above, and encompasses many other examples of demand systems commonly used in the literature. We further extend the previous results to show that such demand $\widetilde{q}_i(p_i/w, \Lambda, V)$ must take the following functional form:

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

where real-valued functions D_i, F and H now also depend on indirect utility V as a second

[4] Auer et al. (2021) in particular provide and estimate a convenient parameterization of such preferences.

argument. Conversely, mild sufficient conditions on these functions ensure that such a demand system is rational, and we show how to characterize direct and indirect utility functions as implicit functions. While the first aggregator Λ must still affect demand only through common price and quantity shifters H and F, as in the first case described above, demand D_i for each good i can be a very flexible function of both its own price p_i and indirect utility V. Thus, the shape of Engel curves (through utility V) can be very different from the shape of the demand curve as a function of the price p_i.

The form of demand in (4) generalizes the previous forms (2) and (3) based on either aggregator Λ or V, and can be used to generate a variety of new and more general demand systems. This includes, for instance, directly and indirectly implicitly-additive separable preferences. Conversely, it is easy to construct preferences with desired properties that could be useful in specific settings. For instance, one can impose homotheticity while retaining very flexible price and substitution effects. Again, this encompasses various examples of homothetic preferences used in the literature, e.g. QMOR when D_i is quadratic (Diewert, 1976; Feenstra, 2018), HDIA when H is constant (Kimball, 1995) and HIIA when F is constant (Matsuyama and Ushchev, 2017).[5]

The single and double-aggregator demand systems can generate choke prices (as demand D_i for a good i equals zero at a finite price) that can be expressed as a simple function of income and the price aggregator (as well as utility in the more general case), with a functional form that is again more flexible than commonly used in macroeconomics and international trade. In particular, these forms of demand can be used to generalize the results of Bertoletti and Etro (2017) and Bertoletti et al. (2018) in which the choke price is proportional to income (see Fally, 2019). It can also rationalize the two-aggregator demand considered in Arkolakis et al. (2019), which is particularly appealing for its tractability and its applications to international trade models with heterogeneous firms. As shown by Thisse and Ushchev (2016), the latter can be generated by aggregating over many rational consumers with random utility; here, we show that such demand can be rationalized with a single representative consumer, which allows us to use standard tools in consumer theory to make welfare statements (such as revealed preferences, compensating variations, and other measures of welfare).

Demand systems with an aggregator are perhaps most useful in the case of monopolistic competition. In the limit where each firm has a negligible market share, it chooses its price by taking such an aggregator as given.[6] With the first type of demand system (Gorman-

[5]QMOR refers to: quadratic mean of order r expenditure function; HDIA: homothetic directly implicitly additive preferences; HIIA: homothetic indirectly implicitly additive.

[6]Recent work by Bertoletti and Etro (2022) formalizes this insight with asymmetric demand and covers the Gorman-Pollak demand system as an example. See also Anderson et al. (2018) on aggregative games where Λ could be used as an "aggregate".

Pollak form), the price aggregator Λ can be interpreted as an index of tightness of the budget constraint, or alternatively as an index of the toughness of competition in a model with firms. A change in the aggregator Λ can lead to a vertical and a horizontal shift of each demand curve, with different implications for markups depending on the shape of these demand curves.

With a simple general-equilibrium model of homogeneous firms under monopolistic competition and free entry, a wide range of comparative statics can be qualitatively obtained with the Gorman-Pollak single-aggregator demand system; all combinations of the signs of the effects of population and income on firm size (and thus prices) appear both in the subconvex case (where markups decrease with firm size) and in the superconvex case. On the contrary, some other forms of separability based on two aggregators can be used to restrict the range of comparative statics. For instance, with directly or indirectly "semi-separable" preferences (defined as a weaker form of additive separability), firm output and prices are invariant to either income or population. In many such cases, simple conditions on the functional form of demand can ensure uniqueness of equilibrium; with Gorman-Pollak preferences in particular, the second-order condition for profit maximization is sufficient.[7] This analysis fits within the framework of Parenti et al. (2017) based on general symmetric demand systems with a continuum of goods, and also complements the earlier results of Bertoletti and Etro (2016) for general symmetric preferences, Zhelobodko et al. (2012) based on directly-additive separable preferences, Bertoletti and Etro (2017) on indirectly-additive separable preferences, and Bertoletti and Etro (2021) on Gorman-Pollak demand. This illustrates the importance of the choice of functional forms of demand in determining key outcomes in general-equilibrium models.

The paper further relates to many others studying functional forms of utility and demand systems, with applications to demand estimation. In particular, Ligon (2016) focuses on cases where the aggregator corresponds to the Lagrange multiplier λ associated with the budget constraint, and shows that a form of separability in λ implies specific functional forms as well as direct additive separability. Nocke and Schutz (2017) study the ("quasi-") integrability of quasi-linear demand systems, i.e. without income effects. Fabinger and Weyl (2016) examine functional forms of demand and production functions that lead to closed-form solutions in models imposing relationships between marginal and average effects. The discussion of the existence of aggregators also mirrors the restrictions associated with the rank of a demand system (Gorman, 1981; Lewbel, 1991; LaFrance and Pope, 2006; Lewbel and Pendakur, 2009), which equals the number of price aggregators needed to recover Engel curves. Here, the number of aggregators corresponds to the rank of the cross-price substitution matrix. The two notions are thus distinct, and the demand systems studied here do not have restrictions in terms of the

[7] Uniqueness of equilibrium is however not guaranteed in the more general case where demand depend on utility in addition to the price aggregator.

rank of Engel curves.[8] Finally, Blackorby et al. (1978) study functional forms implied by various definitions of separability, and find that the same functional structure as with generalized non-homothetic CES is obtained when imposing stronger forms of separability that imply equality among Allen-Uzawa elasticities of substitution.

The remainder of the paper proceeds as follows. Section 2 examines the functional forms imposed by generalized separability. Section 3 provides sufficient conditions for each type of demand to be rationalized. Section 4 discusses various examples of these demand systems. Section 5 examines an application to monopolistic competition and studies market size effects in a simple general-equilibrium model.

2 Functional Forms under Generalized Separability

2.1 Single aggregator

Additively-separable utility yields demand as a simple function of a good's own price p_i and a single aggregator, the Lagrange multiplier. While practical, both direct and indirect separability put strong constraints on the structure of demand, such as a tight relationship between price elasticity and income elasticity, with for instance the adverse consequence that preferences with constant elasticity of substitution (CES) are the only directly-separable and indirectly-separable preferences that are 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:

$$q_i = \widetilde{q_i}(p_i/w, \Lambda) \tag{5}$$

where demand for each good i (in quantity) is a real function of its own normalized price and the aggregator Λ, i.e. a mapping $\widetilde{q_i}$ from $\mathbb{R}_+ \times \mathbb{R}_+$ to $\mathbb{R}_+$, and where $w > 0$ refers to total consumer expenditures and $p_i > 0$ refers to the price of good i. $\Lambda = \Lambda(p/w)$ a real function of the vector of normalized prices $p/w = (p_1/w, ..., p_N/w) \in \mathbb{R}_+^N$, and $N \in \mathbb{N}$ denote the number of goods. Without loss of generality, we assume that Λ is always positive.

We assume that the budget constraint holds for any vector of normalized prices p/w, which implies that the aggregator $\Lambda(p/w)$ must satisfy:

$$\sum p_i q_i = p_i \widetilde{q_i}(p_i/w, \Lambda(p/w)) = w.$$

[8]Demand systems such as PIGL, PIGLOG and AIDS aim to simplify income effects. Here the goal is rather to simplify cross-price effects.

Under the regularity assumption [A1]-i) made below, the solution to this equation in Λ is unique and we can use the budget constraint to obtain the derivatives of Λ w.r.t. prices. 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).

We say that the system of demand given by $\widetilde{q}_i$ and Λ is integrable if there exists a differentiable utility function $U(q)$ such that marginal utility $\frac{\partial U}{\partial q_i}$ evaluated at $\widetilde{q}_i$ (for a given vector of prices and income) is proportional to prices p_i across goods i.[9] We further assume that utility U is twice continuously differentiable, so that its cross-derivatives are symmetric. For the sake of simplicity and exposition, we focus on demand that can be inverted and assume that for each vector $q \in \mathbb{R}_+^N$, there exists a vector $p/w \in \mathbb{R}_+^N$ such that $q_i = \widetilde{q}_i(p_i/w, \Lambda(p/w))$.[10]

In an unpublished note 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:[11]

Regularity assumptions [A1] on functions $\widetilde{q}_i$:

i) $\widetilde{q}_i(p_i/w, \Lambda)$ is positive and twice continuously differentiable, with strictly negative derivatives in both arguments;

ii) Holding Λ constant, $p_i\widetilde{q}_i(p_i/w, \Lambda)$ has a non-zero derivative in p_i

iii) There are at least four goods ($N \geq 4$);

iv) Invertibility: for each $q \in \mathbb{R}_+^N$, $\exists p/w \in \mathbb{R}_+^N$ such that $q_i = \widetilde{q}_i(p_i/w, \Lambda(p/w))$ for all i.

Proposition 1 *If demand is integrable and satisfies conditions [A1], it can be written as either:*

$$\begin{array}{llcll} \textit{case 1:} & \widetilde{q}_i(p_i/w, \Lambda) & = & \dfrac{D_i(F(\Lambda)p_i/w)}{H(\Lambda)} & \textit{for all goods } i \textit{ and all } p_i, w, \Lambda \\ \textit{case 2:} & \widetilde{q}_i(p_i/w, \Lambda) & = & A_i(\Lambda)(p_i/w)^{-\sigma(\Lambda)} & \textit{for all goods } i \textit{ and all } p_i, w, \Lambda \\ \textit{+ case 2':} & \widetilde{q}_i(p_i/w, \Lambda) & = & a_i A(\Lambda)(p_i/w)^{-\sigma_0} & \textit{for } \underline{\textit{all but one}} \textit{ good } i \end{array}$$

[9] A distinction is often made between integrability and rationalization, whereby the latter further requires U to be quasi-concave. In other words, integrability imposes the Slutsky substitution matrix to be symmetric while rationalization also requires that it be semi-definite negative.

[10] This assumption is made for convenience as the proof mostly focuses on the inverse demand. Most of the arguments are local and would apply to subsets of prices and quantities where we have invertibility.

[11] 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 inelegant 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".

or a combination of cases 2 and 2' (depending on Λ), where a_i, σ_0 and ρ_0 are positive constant terms, and D_i, F, H, A and A_i are differentiable real functions with a single argument.

To prove Proposition 1, it is actually easier to work with the inverse demand, i.e. expressing normalized prices as a function of quantities, as such objects are more directly related to marginal utility. A key integrability condition comes from the symmetry of the Hessian of the utility function. As inverse demand is proportional to marginal utility, its derivatives also need to feature some symmetry, a condition equivalent to Slutsky symmetry for Marshallian demand.[12] When cross-price effects are captured by a single aggregator, these symmetry conditions impose conditions on price elasticities that can then be integrated to provide the functional forms in Proposition 1. As part of the proof of Proposition 1, we find that inverse demand takes a very similar functional form in both cases. In the case 1, we can express inverse demand as:

$$\frac{p_i}{w} = \frac{D_i^{-1}(H(\Lambda)q_i)}{F(\Lambda)} \tag{6}$$

where Λ is now seen as a function of the vector of consumption and can be implicitly defined as a solution to the budget constraint using inverse demand: $\sum_i q_i D_i^{-1}(H(\Lambda)q_i)/F(\Lambda) = 1$.[13]

Since the third case 2' is relatively less interesting and elegant (CES for all but one good), the remainder of the paper focuses on cases 1 and 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 exclude such cases. Also note that functional forms are unique up to a constant term and a monotonic transformation of Λ, both in cases 1 and 2. Moreover, as we will see later in Section 3.2 under additional restrictions, aggregator Λ coincides in case 2 with indirect utility V (up to a monotonic transformation).

Before turning to more general demand systems and sufficient conditions for rationalization, it is useful at this point to summarize some key properties implied by these two types of demand, especially in terms of price and income effects.

Price and income elasticities in case 1. Let us denote by $\varepsilon_{Di} = \frac{\partial \log D_i}{\partial \log p_i}$, $\varepsilon_H = \frac{\partial \log H}{\partial \log \Lambda}$ and $\varepsilon_F = \frac{\partial \log F}{\partial \log \Lambda}$ the elasticity of D_i, H and F in their argument. In case 1, the price elasticity of

[12] A working paper version provides a proof based on direct demand, examining price effects and Slutsky symmetry, which more closely follows the steps proposed by Gorman (1972) in his unpublished notes. Actually, earlier literature on rationalization (e.g. Samuelson, 1950) puts more attention onto properties of inverse demand than relatively more recent works that have placed a greater emphasis on the Slutsky substitution matrix (following Hurwicz and Uzawa, 1971).

[13] In the Marshallian demand formulation, Λ is a function of prices and income. In the inverse demand formulation, we redefine Λ as a function of quantities, where Λ can again be implicitly characterized by the budget constraint. As an abuse of notation, we use the same notation in both approaches.

Marshallian demand is:

$$\frac{\partial \log q_i}{\partial \log p_j} = \varepsilon_{Di} \cdot \mathbb{1}_{(i=j)} - \frac{W_j(1+\varepsilon_{Dj})(\varepsilon_H - \varepsilon_F \varepsilon_{Di})}{\varepsilon_H - \varepsilon_F \bar{\varepsilon}_D} \tag{7}$$

where W_j is the expenditure share of good j, $\mathbb{1}_{(i=j)}$ is a dummy equal to one when $i = j$, and $\bar{\varepsilon}_D = \sum_i W_i \varepsilon_{Di}$. When that good has a negligible market share, the own price elasticity is determined by the shape of function D_i: $\frac{\partial \log q_i}{\partial \log p_i} \approx \varepsilon_{Di}$. Since we impose few constraints on ε_{Di}, the shape of each demand curve and the patterns of price elasticities can be very flexible.

In turn, the income elasticity of demand is:

$$\frac{\partial \log q_i}{\partial \log w} = 1 + \frac{(\varepsilon_H + \varepsilon_F)(\bar{\varepsilon}_D - \varepsilon_{Di})}{\varepsilon_H - \varepsilon_F \bar{\varepsilon}_D}. \tag{8}$$

Using this expression, one can see that homotheticity implies that either $\varepsilon_H = -\varepsilon_F$ or $\varepsilon_{Di} = \bar{\varepsilon}_D$ for all goods i (see Section 4.2).

As pointed out by Pigou (1910) and Deaton (1974), own-price elasticities and income elasticities are colinear (across goods) when demand is derived from directly-additive utility ($\varepsilon_H = 0$) when each good i's expenditure share is small: $\frac{\partial \log q_i}{\partial \log w} = \frac{\varepsilon_{Di}}{\bar{\varepsilon}_D}$. With $\varepsilon_H \neq 0$, the relationship between income elasticity and price elasticity is affine. The relative rankings can even be flipped if $\varepsilon_F + \varepsilon_H > 0$, with price-elastic goods being relatively less income elastic.[14]

Price and income elasticities in case 2. In the second case, price effects are simpler: the own-price elasticity is constant for a given level of aggregator Λ, and we will see in Proposition 4 that in this case we can interpret aggregator Λ as indirect utility: $\Lambda = V$.

This demand system is most interesting and useful for its very flexible income effects. Comparing goods, first we can see that changes in $A_i(\Lambda)$ in Λ need not be related to $\sigma(\Lambda)$, thus breaking away from the link between price and income elasticities discussed for the first case above. Starting with the special case where $\sigma(\Lambda) = \sigma$ is constant, we find that income elasticities are: determined by the elasticity of each A_i w.r.t Λ:

$$\frac{\partial \log q_i}{\partial \log w} = \frac{\varepsilon_{Ai}}{\bar{\varepsilon}_A}. \tag{9}$$

where $\bar{\varepsilon}_A$ is the average of elasticities $\varepsilon_{Ai} = \frac{\Lambda A_i'(\Lambda)}{A_i(\Lambda)}$ weighted by expenditures shares. Hence, good i is income-elastic if and only if $\varepsilon_{Ai}/\bar{\varepsilon}_A > 1$.

In the more general case where $\sigma(\Lambda)$ is not constant, function A_i plays a similar role and dictates income effects, while $\sigma(\Lambda)$ determines how the price elasticity varies with income.

[14]Note that we can obtain inferior goods (as it is already the case with indirectly-additive preferences).

2.2 Generalization with indirect utility as an additional aggregator

The single-aggregator cases impose either a tight constraint on price elasticities (case 2 above) or an affine relationship between price and income elasticities (case 1), and exclude several demand systems examined in the literature that more generally require two aggregators. The next objective is to examine a combination of cases 1 and 2, by studying demand systems that depend on a common aggregator Λ as well as on utility V as a second aggregator. The goal is to generate more flexible price and income effects that are not as tightly linked, and provide a more general formulation that includes most demand systems used in practice. Hence, we now suppose that demand takes the form:

$$q_i = \widetilde{q}_i(p_i/w, \Lambda, V) \tag{10}$$

where $\widetilde{q}_i$ is now a mapping from $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$ to $\mathbb{R}_+$, $V = V(p/w)$ refers to the indirect utility function evaluated at p/w, and Λ, as earlier, satisfies the budget constraint:

$$\sum_i \frac{p_i q_i}{w} = \sum_i \frac{p_i \widetilde{q}_i(p_i/w, \Lambda, V)}{w} = 1.$$

Under conditions [A2] imposed for Proposition 2, inverse demand $\widetilde{q}_i^{-1}$ is well defined and can be expressed as a function of quantity q_i, the direct utility function U, and the aggregator Λ, which we can alternatively express as a function of quantities q such that the budget constraint holds (see Appendix for details). This gives:

$$p_i/w = \widetilde{q}_i^{-1}(q_i, \Lambda, U) \tag{11}$$

where $\widetilde{q}_i^{-1}$ denotes the inverse demand w.r.t. normalized prices, and where $\Lambda = \Lambda(q)$ is now implicitly defined as a function of quantities. We use the fact that the derivatives of the indirect utility function and the direct utility function are proportional to demand q_i across goods i.

We can generalize Proposition 1 under a set of similar regularity restrictions on differentiability, minimum number of goods and price effects:

Regularity assumptions [A2] on functions $\widetilde{q}_i$:

i) $\widetilde{q}_i(p_i/w, \Lambda, V)$ is positive and twice continuously differentiable, with a strictly negative derivative in p_i and Λ;

ii) Holding Λ and V constant, $p_i\widetilde{q}_i(p_i/w, \Lambda, V)$ is not constant over the range of prices p_i;

iii) There are at least four goods and, for any vector of normalized prices p/w (except for a set of prices of measure zero), the price elasticity takes at least three values across goods.

iv) Invertibility: for each $q \in \mathbb{R}_+^N$, $\exists p/w \in \mathbb{R}_+^N$ s.t. $q_i = \widetilde{q}_i(p_i/w, \Lambda(p/w), V(p/w))$ for all i.

Proposition 2 *If demand $\widetilde{q}_i$ is integrable, depends on two aggregators as in equation (10) and satisfies regularity conditions [A2], it can be written as:*

$$\widetilde{q}_i(p_i/w, \Lambda, V) \; = \; \frac{1}{H(\Lambda, V)} D_i \left(\frac{p_i F(\Lambda, V)}{w} \, , \, V \right) \tag{12}$$

where D_i, F and H are mappings from $\mathbb{R}_+ \times \mathbb{R}$ to $\mathbb{R}$, with indirect utility V as second argument.

This functional form is again imposed by symmetry conditions, and the proof of Proposition 2 follows very similar steps as for Proposition 1. Surprisingly, these symmetry conditions do not impose strong constraints on functional forms in terms of how indirect utility, used as a second aggregator, influences demand patterns. We can also verify that Proposition 1 is a special case where demand depends on either V or Λ.[15]

This leads to considerably greater flexibility: indirect utility can influence partial demand functions D_i as well as functions F and H. In the single-aggregator case, we have seen that the own-price elasticity is given by the elasticity of D_i, and thus the shape of D_i influences how price elasticities (and thus markups in models of imperfect competition) vary along the demand curve depending on the level of demand for a particular good i. With utility V as an additional aggregator, the shape of demand curves can itself vary with utility, and allows for flexible good-specific Engel curves unrelated to the effect of its own price. Moreover, interpreting aggregator Λ as capturing the tightness of the budget constraint, its effect on the price shifter F and quantity shifter H can now also depend on the level of utility of consumers (and indirectly on their income).

Note also that inverse demand has a similar form, now as a function of direct utility $U(q)$:

$$\widetilde{q}_i^{-1}(q_i, \Lambda, U) \; = \; \frac{1}{F(\Lambda, U)} D_i^{-1} \left(q_i H(\Lambda, U) \, , \, U \right)$$

where D_i^{-1} denotes the inverse of D_i with respect to its first argument, and where Λ can instead be seen as a function of quantities q (again implicitly defined by the budget constraint).

At this point, we only impose restrictions on Slutsky symmetry (integrability), but we will see in the next section that fairly mild additional restrictions are sufficient to ensure that such systems are rational, so that these considerations on price and income effects will remain valid.

[15] In order to avoid a taxonomy of cases, condition [A2]-iii) above imposes enough heterogeneity in price elasticities across goods. This excludes special cases that would resemble cases 2 and 2' in Proposition 1. This restriction does not lead to an important loss of generality given that a key motivation for Proposition 2 is to examine more flexible demand systems in both income and price effects.

3 Rationalization

Let us now examine the reciprocals of Proposition 1 and 2. Under which conditions these demand systems can be rationalized, i.e. can be derived from maximizing a well-behaved quasi-concave and monotone utility function? These functional forms, imposed by the symmetry of the Slutsky matrix, do not necessarily lead to quasi-concavity or monotonicity of utility functions (see Appendix for counter-examples in the single-aggregator case). We now explore sufficient conditions to guarantee that the demand systems described in Proposition 1 and 2 are rational.

3.1 Rationalization of Gorman-Pollak Demand

Suppose that demand is given by:

$$q_i = \frac{D_i(F(\Lambda)p_i/w)}{H(\Lambda)} \tag{13}$$

where D_i, F and H are mappings from $\mathbb{R}_+$ to $\mathbb{R}_+$, and 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:

$$H(\Lambda) \ = \ \sum_i (p_i/w)\, D_i(F(\Lambda)p_i/w). \tag{14}$$

As before, we denote by $\varepsilon_{Di} = \frac{\partial \log D_i}{\partial \log p_i}$ the elasticity of D_i in its argument, and $\varepsilon_F = \frac{\partial \log F}{\partial \log \Lambda}$ and $\varepsilon_H = \frac{\partial \log H}{\partial \log \Lambda}$ the elasticity of F and H in Λ. To ensure that (14) has a unique solution in Λ and that this demand system is well-defined and rational, we impose the following regularity restrictions on D_i, F and H:

Regularity assumptions [**A3**] on functions D_i, F and H:

i) D_i is continuously differentiable, $\varepsilon_{Di} < 0$;

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

iii) For any good i and $y_i > 0$, there exists $\Lambda \in \mathbb{R}_+$ such that: $y_i D_i \left(y_i F(\Lambda)\right)/H(\Lambda) = 1/N$

Note that instead of condition [A3]-ii) we could assume that $\varepsilon_F \, \varepsilon_{Di} - \varepsilon_H$ has the same sign for all i, Λ and p_i/w. If it is positive instead of negative (for all goods and prices), condition ii) is satisfied if we consider the change in variable $\Lambda' = 1/\Lambda$. Assumptions i) and ii) imply that the solution in Λ to equation (14) is always unique, but they are also needed to show that utility is quasi-concave and that the Slutsky substitution matrix is negative semi-definite. Condition iii) ensures that equation (14) has a solution in Λ: in other words, the aggregator Λ can always adjust in order to satisfy the budget constraint. It is automatically satisfied, for instance, if we assume that the image of the mapping $\Lambda \mapsto \frac{D_i(F(\Lambda)p_i/w)}{H(\Lambda)}$ is $(0, +\infty)$, conditional on p_i/w.[16]

As with Proposition 1, it is useful to consider the inverse demand, which shares a similar functional form. We can redefine Λ as an implicit function of q using the budget constraint:

$$\sum_i q_i \, D_i^{-1}(H(\Lambda)q_i) \;=\; F(\Lambda) \tag{15}$$

which, under conditions ii) and iii) has a unique solution in Λ for any q (see Appendix).

Under these conditions, we obtain:

Proposition 3 *If H and D_i satisfy the regularity conditions [A3], the demand described in equations (13) and (14) can be rationalized and obtained from a continuous quasi-concave utility:*

$$U(q) = \sum_i \int_{q'=q_{i0}}^{x=H(\Lambda(q))q_i} D_i^{-1}(q')dq' \;-\; \int_{l=\Lambda_0}^{\Lambda(q)} H'(l)F(l)dl \tag{16}$$

where $\Lambda(q)$ satisfies (15) for each q, and Λ_0, $q_{0i} \geq 0$ are constant terms.

Proposition 3 rationalizes such demand in a constructive way, by directly providing a utility function (see Appendix for details).[17] Note that this utility function is unique, up to a monotonic transformation.

The least obvious part of the proof is to show that it is quasi-concave, accounting for how the aggregator Λ responds to changes in q. Note that equation (15) can be seen as a first-order condition such that the expression above for U has a zero derivative in Λ. As such, marginal utility takes a simple form:

$$\frac{\partial U}{\partial q_i} = HD_i^{-1}(Hq_i). \tag{17}$$

An alternative is to build on the proof provided by Matsuyama and Ushchev (2017) for the homothetic case (HSA), and examine the Slutsky substitution matrix. Thanks to the functional form obtained in Proposition 1, the Slutsky substitution matrix is symmetric, but conditions in

[16] Alternatively, in condition iii), one could replace the term $1/N$ on the left hand side by a term that varies across goods i as long as this term sums up to unity across goods.

[17] 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: $U = \max_\Lambda \{\sum_i u_i(\Lambda q_i) - \Phi(\Lambda)\}$ but this approach is equivalent to assuming $H'(\Lambda) > 0$ in the formulation provided here, and omits useful cases (such a continuum of cases providing a bridge between directly-additive and homothetic-single-aggregator preferences) where the second order condition of this maximization is not satisfied yet the utility function constructed above remains quasi-concave.

Proposition 1 do not guarantee its semi-definite negativity. As one could expect, the conditions needed for semi-definite negativity are the same as conditions [A3] above providing the quasi-concavity of the utility function.

One can ask whether the set of conditions [A3] can be relaxed, but I argue here that all are needed. First, the demand system would clearly not be well defined if it does not have a solution in equation (14), so condition iii) is unavoidable. It is possible to impose simpler conditions to ensure existence, but such conditions would be less general or practical. Second, restriction ii) is the simplest and more direct way to ensure that the equation defining the price aggregator has a unique solution. It is required for good i for a given level of prices when a good i has a sufficiently large expenditure share. In the Appendix, I provide an example with two goods where restrictions i) and iii) are met but the Slutsky matrix is no longer negative semi-definite when $\varepsilon_F\,\varepsilon_{Di} - \varepsilon_H$ does not have the same sign for the two goods. Finally, restriction i) ensures that we have a negative effect of prices on demand when the expenditure share of a good is small (a positive price effect would not be rational for small expenditure shares). Inverting D_i is also needed in equations (15) and (16) to retrieve utility.

Drawing from Pollak (1972), indirect utility can be expressed as:

$$V(p,w) \;=\; -\sum_i \int_{y_{i0}}^{(p_i/w)F(\Lambda)} D_i(y)dy \;+\; \int_{\Lambda_0}^{\Lambda} F'(l)H(l)dl \;+\; g_0 \tag{18}$$

where y_{i0}, g_0 and Λ_0 are constant terms (see details in the Appendix). $\Lambda = \Lambda(p/w)$ can either be implicitly defined by the budget constraint as above, or by taking the derivative of expression (18) w.r.t. Λ. This expression can also be useful to compute equivalent and compensating variations, implicitly defined such that $V(p', w - CV) = V(p,w)$ and $V(p, w + EV) = V(p', w')$. Taking the derivative w.r.t. income, one can interpret the product of the two shifters as the marginal utility of income (in log):

$$\frac{\partial V(p,w)}{\partial \log w} \;=\; F(\Lambda)H(\Lambda).$$

In terms of price and income effects, already discussed in Section 2.1 (expressions 7 and 8), assumptions [A3] do not impose stark additional restrictions. Given that we assume $\varepsilon_H > \varepsilon_F \varepsilon_{Di}$, note however that the own-price elasticity is always negative, which rules out Giffen goods (but not inferior goods). Given that restriction, we can also see that the cross-price elasticity ($i \neq j$) is positive if and only if $\varepsilon_{Dj} < -1$.

Such demand is slightly more general than the one used in Pollak (1972) and more recently in Bertoletti and Etro (2022) as it does not require either $F(\Lambda)$ and $H(\Lambda)$ 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 $H(\Lambda)$. If $F(\Lambda)$ is strictly monotonic (which is satisfied in practice for most applications, see e.g. Fally, 2019), then without loss of generality we can assume $F(\Lambda) = \Lambda^\beta$ with $\beta \in \{-1, 1\}$.

3.2 Rationalization of Generalized Non-Homothetic CES

Now, consider the case 2 of Proposition 1. Let us assume that expenditure shares are given by:

$$p_i q_i / w \;=\; (G_i(\Lambda)\, p_i / w)^{1-\sigma(\Lambda)} \tag{19}$$

where σ and each G_i is a continuous mapping from $\mathbb{R}_+$ to $\mathbb{R}_+$, where $\Lambda(p/w)$ is itself a function of the vector of normalized prices p/w. We assume that the budget constraint is satisfied, i.e.:

$$\sum_i (G_i(\Lambda) p_i / w)^{1-\sigma(\Lambda)} = 1. \tag{20}$$

To ensure integrability, we impose the following sufficient regularity restriction [A4]:

Regularity assumptions [A4] For each Λ, we have $\sigma(\Lambda) \neq 1$ and *either* one of the following two conditions:

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

Continuity is sufficient for the main statement. However, 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) is formally equivalent to imposing:[18]

$$\sum_i \exp\left(\frac{(\sigma(\Lambda)-1)^2 G_i'(\Lambda)}{\sigma'(\Lambda) G_i(\Lambda)}\right) < 1 \tag{21}$$

(see Appendix for the proof of equivalence). Under these conditions, we obtain Proposition 4:

[18] 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.

Proposition 4 *Suppose that demand can be written as in equation (19) where G_i and σ are continuous and where Λ is implicitly defined by (20). This demand system is integrable if conditions [A4] are satisfied. Under [A4], demand can be derived from a utility function that is implicitly defined by:*

$$\sum_i (q_i/G_i(U))^{\frac{\sigma(U)-1}{\sigma(U)}} = 1 \tag{22}$$

which has a unique solution in U, with $\Lambda = U$ for the demand q_i described above.

The constant elasticity case $\sigma(\Lambda) = \sigma$ corresponds to implicitly additive utility as in Comin et al. (2021). This is not equivalent to the standard CES since, even in that case, non-trivial income effects through the demand shifter $G_i(\Lambda)$ allow for very flexible Engel curves. The main contribution of this proposition is to generalize to variable elasticity of substitution.

Notice that the expenditure function takes a simple form, as in Comin et al. (2021):

$$e(p, U) \ = \ \left[\sum_i (G_i(U)p_i)^{1-\sigma(U)}\right]^{\frac{1}{1-\sigma(U)}}.$$

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 (72). As the more general case allows for varying curvature of indifference curves, one needs to ensure in particular that these indifference curves do not cross.

The proof proceeds as follows. First, notice that $\left[\sum_i \alpha_i x_i^\rho\right]^{\frac{1}{\rho}}$ is monotonically increasing in ρ if $\sum_i \alpha_i = 1$ ("Generalized Mean Inequality", Lemma 1 in Appendix). This allows us to obtain comparative statics in the exponent in equations (72) and (20). We can then show that the solutions to these equations are unique, for a given set of income and prices, or quantities. Once we have uniqueness, it is easy to verify the quasi-concavity of the utility function (as in Comin et al., 2021). The last step is to check that this utility maximum problem does yield the demand system described above.

Again, as for Proposition 3, a potential concern is whether restrictions [A2] are necessary. When neither condition i) nor ii) is satisfied, neither the demand system described above nor the utility in Proposition 4 is well defined. Counter-examples in the Appendix further illustrate the role of each condition, showing that equations (20) and (72) admit multiple solutions in Λ and U if conditions i) and ii) are not satisfied. Incidentally, this shows that monotonicity in demand shifters $G_i(\Lambda)$ is not sufficient.[19]

One should also point out why we need different conditions depending on whether $\sigma(\Lambda)$ decreases or increases with Λ. In the first case, where $\sigma(\Lambda)$ increases with Λ, indifference

[19] We can also have $\sigma(\Lambda) = 1$ for a discrete number of values of Λ.

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 around their midpoint.

3.3 Rationalization with two aggregators Λ and V

Suppose that demand takes the form:

$$q_i(p_i/w, \Lambda, V) \; = \; \frac{1}{H(\Lambda, V)} D_i \left(\frac{p_i F(\Lambda, V)}{w} \, , \, V \right) \tag{23}$$

where D_i, F and H all are positive continuously-differentiable mappings from $\mathbb{R}_+ \times \mathbb{R}$ to $\mathbb{R}_+$, with aggregator V as a second argument (which must coincide with indirect utility if such demand is rational). Denote by ε_{Di} the elasticity of D_i with respect to price p_i (holding Λ and V constant) and by ε_H and ε_F the elasticities of H and F in terms of Λ (holding V constant).

We then impose the following sufficient regularity restrictions:

Regularity assumptions [**A5**] on functions D_i, F and H:
i) D_i is continuously differentiable, with $\varepsilon_{Di} < 0$;
ii) H and F are continuously differentiable, with $\varepsilon_F \, \varepsilon_{Di} < \varepsilon_H$ for all i, Λ, V and p_i/w;
iii) For any good i, $y_i > 0$, $V \in \mathbb{R}$, $\exists \Lambda \in \mathbb{R}_+$ such that: $y_i D_i \left(y_i F(\Lambda, V), V \right) / H(\Lambda, V) = 1/N$.

These conditions ensure that, for each V and p/w, there is a unique Λ such that the budget constraint is satisfied, i.e. such that $\sum_i (p_i/w) q_i(p_i/w, \Lambda, V) = 1$ with demand defined in equation (23) above. A similar result is obtained for the inverse demand. For any given vector of quantities q and utility U, the following budget condition for inverse demand:

$$\sum_i q_i D_i^{-1}(q_i H(\Lambda, U), U)/F(\Lambda, U) = 1 \tag{24}$$

has a unique solution in Λ.

These conditions are very similar to those used in the single-aggregator case for Gorman-Pollak demand in Proposition 3.[20] Under these conditions, we obtain the following proposition characterizing utility for more general demand systems with two aggregators including utility:

[20] Again, as in Proposition 3, in condition iii) one could replace the term $1/N$ by a series of good-specific terms that sum up to unity across goods.

Proposition 5 *Suppose that demand can be written as in equation (23) satisfying regularity assumptions [A5] above, where V is indirect utility and Λ is an aggregator such that the budget constraint (24) holds. Then:*

i) Utility U must satisfy:

$$\sum_i \int_{q=q_{i0}}^{q_i H(\Lambda,U)} D_i^{-1}(q,U)\,dq \;-\; G(\Lambda,U) \;=\; 0 \tag{25}$$

for some constant terms $q_{i0} \geq 0$ and a continuously differentiable real-valued function $G(\Lambda,U)$ such that $\frac{\partial G}{\partial \Lambda}(\Lambda,U) = \frac{\partial H}{\partial \Lambda}(\Lambda,U)F(\Lambda,U)$.

ii) Conversely, if the left-hand-side of equation (25) is decreasing in U (with a strictly negative partial derivative in U), equations (25) and (24) uniquely characterize a well-behaved utility (monotonic, continuous and quasi-concave) that yields demand as in equation (23).

Taken together, under conditions [A2] and [A5], Propositions 2 and 5 provide a characterization of rational demand functions with two aggregators Λ and V capturing cross-price effects, and a characterization of their associated utility functions.

The proof of Proposition 5 (see Appendix) combines elements of Propositions 3 and 4. First, the implicit solution for utility U must be monotonically increasing in q_i for each good i. Here, this property is obtained by assuming that the left-hand side of equation (25) is decreasing in U (conditional on q and Λ), given that the left-hand side has a strictly positive derivative in each q_i and has a zero derivative in Λ.

Next, the proof that utility U is quasi-concave in q is similar to the one in Proposition 3 for the single-aggregator case. Considering the left-hand-side of equation (25) as a function of q and U, it suffices to show that it is quasi-concave in q (holding U constant) in order to obtain that the implicit function for U is quasi-concave in q. Holding U constant, we can see that the left-hand side of equation (25) has the same structure w.r.t. q and Λ as the right-hand side of equation (16) for utility in the single-aggregator case in Proposition 3.

One must also ensure that Λ is well defined (implicitly defined such that the budget constraint holds). Condition [A5]-iii) leads to the existence of Λ, while condition ii) provides uniqueness. As shown in the Appendix, the same two conditions also ensure the existence and uniqueness of Λ as a function of quantities instead of normalized prices.

Proposition 5 does not provide precise criteria, e.g. as in Proposition 4, to determine when the left-hand side of equation (25) is decreasing in U, but in practical cases this condition is easy to check. For instance, if neither F nor H, nor G depend on U, as in several of the examples provided below in Section 4, a sufficient condition for the existence and monotonicity in U is that $D_i(q_i, U)$ is strictly decreasing in U (holding q_i constant) and varies from $+\infty$ to zero in the limit over the range of U. Conversely, interesting cases also arise when only H and F depend on U (e.g. semi-separable preferences, as discussed in Section 4.1), in which case monotonicity is again not difficult to characterize.

Proposition 5 highlights how to characterize direct utility as a function of quantities q. As in the single-aggregator case, we obtain a similar characterization of indirect utility as a function of normalized prices p/w. Integrating by part, we show in the Appendix that the indirect utility satisfies the following equation:

$$\sum_i \int_{y=y_{i0}}^{\frac{p_i}{w}F(\Lambda,V)} D_i(y,V)dy \;\;=\;\; K(\Lambda, V) \tag{26}$$

where K is such that $\frac{\partial K}{\partial \Lambda}(\Lambda, V) = \frac{\partial F}{\partial \Lambda}(\Lambda, V)H(\Lambda, V)$, and Λ can again be implicitly defined such that the partial derivatives in Λ are equalized, or equivalently can be implicitly defined such that the budget constraint holds (here as a function of normalized prices p/w). Using Roy's identity, we can obtain Marshallian demand directly from this expression, which is sometimes simpler than using expression (25) in Proposition 5 (e.g. as in cases of indirect separability).

4 Special cases and examples

This section discusses additional examples where these results can be applied, including a discussion of different forms of separability and several examples of homothetic preferences. The remainder of the section examines demand systems with two aggregators as in Thisse and Ushchev (2016), and shows that one of the two aggregators can be set equal to indirect utility without loss of generalization. We also discuss extensions to demand with choke prices.

4.1 Forms of separability and non-homothetic examples

Direct and indirect additive separability Let us recall here the functional form taken in one of the simplest cases discussed earlier —direct additive separability— as it will serve as a reference for other generalizations. Preferences are directly separable if there is only a single aggregator and function H is constant. In that case, we can write utility as:

$$U(q) \;=\; \sum_i \int_{q=q_{i0}}^{q_i} D_i^{-1}(q)dq$$

which also leads to a simple demand function: $q_i = D_i(\Lambda p_i/w)$. Directly-separable preferences have been used extensively in the literature, across many fields in economics. The main reason for their wide use is their tractability, and they already offer flexible price effects along each demand curve for each good. However, as pointed out for instance by Deaton (1974), assuming direct separability comes at the cost of imposing strong restrictions on price and income elasticities.

A first step away from directly separable preferences is to consider indirectly separable preferences, for which indirect utility can be written as

$$V(p/w) \ = \ \sum_i \int_{y=y_{i0}}^{p_i/w} D_i(y)dy,$$

which leads to a demand function even simpler than in the previous case: $q_i = D_i(p_i/w)/\Lambda$ with $\Lambda = \sum_j (p_j/w)D_j(p_j/w)$. However, these preferences still impose strong restrictions on demand patterns and also a tight link between income and price elasticities. In particular, CES demand is the only form of either directly or indirectly-additive preferences that is homothetic.

A parameterized version of non-homothetic CES Auer et al. (2021) propose a useful parameterization of the generalized CES from Proposition 4 with demand specified as:

$$q_i = \alpha_i U^{\gamma_i}(p_i/w)^{-\sigma(U)} \qquad \text{with} \qquad \sigma(U) = \bar{\sigma} + \sigma_1 \log U.$$

This can be derived from indirect and direct utility implicitly defined by:

$$\sum_i \alpha_i V^{\gamma_i}(p_i/w)^{1-\sigma(V)} = 1 \qquad \text{and} \qquad \sum_i (\alpha_i U^{\gamma_i})^{\frac{1}{\sigma(U)}} q_i^{\frac{\sigma(U)-1}{\sigma(U)}} = 1.$$

When σ_1 is negative, i.e. when the price elasticity decreases with utility, inequality (21) provides a condition for rationalization that conveniently simplifies into the following:

$$\sum_i \alpha_i \exp\left[\gamma_i \frac{1-\bar{\sigma}}{\sigma_1}\right] \ < \ 1.$$

A set of sufficient conditions for rationalization is then: $\sum_i \alpha_i = 1$, $\bar{\sigma} > 1$, $\sigma_1 < 0$ and $\gamma_i < 0$.

This parameterization allows for good-specific income elasticities, price elasticities that vary with income, while keeping common price elasticities across goods. This provides a practical framework for estimation, as shown by Auer et al (2020).

Implicit additive separability A type of separability which has recently seen a gain in interest is implicit (additive) separability, which again can be distinguished into direct and indirect implicit separability. Preferences are directly implicitly separable if utility can be characterized as the solution of an equation of the type:[21]

$$\sum_i \int_{q=q_{i0}}^{q_i} D_i^{-1}(q, U)dq \ = \ 1 \tag{27}$$

where D_i is a function of two arguments.

Such preferences are a special case of Proposition 5 but not Proposition 3. In fact, implicitly-additive preferences depend on a single aggregator only when they are also directly separable or when price elasticities are uniform (non-homothetic CES case). With two aggregators as in Proposition 5, preferences are implicitly additively separable if and only if H does not depend on Λ, and in this case it is without loss of generality to assume $H = 1$.

Similar results are obtained for the implicitly-indirectly-additive case, defined as when indirect utility can be characterized as the solution of:

$$\sum_i \int_{y=y_{i0}}^{p_i/w} D_i(y, V)dq \ = \ 1. \tag{28}$$

It is a special case of Proposition 5 when $F = 1$. Implicit separability (direct or indirect) can prove useful in order to generate price and income effects that are less tightly related as with direct and indirect separability. In particular, for a given consumer, the ranking in price elasticities across goods can be totally uncorrelated with the ranking of income elasticities.[22] When such demand features a choke price (see Section 4.4), note that the choke price does not depend on aggregator Λ and solely depends on income and utility.

Direct semi-separability Let us introduce a new class of preference, which we refer to as "semi-separable" (a weaker form of additive separability), where we can express either direct or indirect utility as a more simple function of quantities or prices as well as the aggregator. First, let us define preferences as directly semi-separable if we can write utility as:

$$U(q) = \frac{1}{G(\Lambda)} \sum_i \int_{q=0}^{H(\Lambda)q_i} D_i^{-1}(q)dq \tag{29}$$

[21] Blackorby et al. (1991) provide yet another generalization of implicit separability.

[22] Implicit separability, resp. direct and indirect, offers less flexibility than the general form of Proposition 5 for how demand can shift, resp. horizontally and vertically (with shifts that can only depend on utility).

where H, G and D_i are twice continuously-differentiable, with $G' > 0$, $H' > 0$, $D_i > 0$ and $D_i' < 0$. As with Gorman-Pollak demand, we define Λ such that the derivative w.r.t. Λ of the expression above is null, i.e. such that:

$$\frac{\sum_i q_i D_i^{-1}(H(\Lambda)q_i)}{\sum_i \int_{q=0}^{H(\Lambda)q_i} D_i^{-1}(q)dq} = \frac{F(\Lambda)}{G(\Lambda)} \tag{30}$$

where $F(\Lambda) \equiv G'(\Lambda)/H'(\Lambda)$ is assumed to be a positive and continuously differentiable.[23]

This demand system is a special case of Proposition 5. Conditions [A5] required by Proposition 5 are met if $D_i(F(\Lambda)y_i)/H(\Lambda)$ has a strictly negative derivative in Λ and goes from $+\infty$ to 0 (in the limit) as Λ increases, holding y_i fixed. In this case, the system of equations (29) and (30) has a unique solution in the aggregator Λ and utility U, and define a well-behaved utility for any q. Demand for good i is then:

$$q_i = \frac{D_i(VF(\Lambda)p_i/w)}{H(\Lambda)} \tag{31}$$

where $V = V(p/w)$ refers to indirect utility and Λ satisfies equation (30).

These preferences provide a generalization of directly-additive separability, and also retain some of the properties associated with direct separability. Directly-separable preferences correspond to the limit case where both H and G are constant and $F(\Lambda) = \Lambda$. These preferences offer a similar degree of flexibility as Gorman-Pollak preferences with a single aggregator (the multiplicative specification of utility, equation (29) mirrors the additive specification in Proposition 3). For a given consumer, there is again an affine relationship between income elasticities and price elasticities across goods. Moreover, as will be discussed in Section 5.2, another reason to introduce this new type separability is to highlight a more general class of preferences with similar implications for market size effects as additively-separable preferences.

Indirect semi-separability We can obtain a similar functional form for indirect utility if we make the same functional form assumptions as above for D_i^{-1} instead of D_i. Suppose that indirect utility can be expressed as:

$$V(p/w) = \frac{1}{L(\Lambda)} \sum_i \int_{y=F(\Lambda)p_i/w}^{\infty} D_i(q)dq \tag{32}$$

where F, L and D_i are twice continuously differentiable, with $F' > 0$, $L' < 0$, $D_i > 0$ and $D_i' < 0$. We define Λ such that the derivative w.r.t. Λ of the expression above is null, i.e. such

[23] Since both G and H are strictly monotonic functions of Λ, it is without loss of generality to impose either $H = 1$ or $G = 1$, whichever is more practical.

that:

$$\frac{\sum_i (p_i/w) D_i(F(\Lambda)p_i/w)}{\sum_i \int_{y=F(\Lambda)p_i/w}^{\infty} D_i(q)dq} = \frac{H(\Lambda)}{L(\Lambda)} \tag{33}$$

where we denote $H(\Lambda) = -L'(\Lambda)/F'(\Lambda)$, a positive and continuously-differentiable function of Λ. Note that this equation in Λ does not involve indirect utility V. Again, this is a special case of Proposition 5. The conditions for integrability are the same as above (in terms of D_i, F and H) for directly semi-separable preferences. In this case, Marshallian demand takes the form:

$$q_i = \frac{D_i(F(\Lambda)p_i/w)}{VH(\Lambda)}. \tag{34}$$

As the name suggests, such preferences provide a generalization of indirectly-additive separability, which corresponds to the limit case where F and L are constant. Such preferences yield similar properties as indirectly-additive preferences in terms of market size effects in general-equilibrium models with economies of scale, as we will discuss in Section 5.2.

Bi-power demand. A prominent type of demand studied in Mrázová and Neary (2013) is the bi-power form, where demand for good i takes the form: $q_i = \gamma_i p_i^{-\nu_i} + \delta_i p_i^{-\sigma_i}$ in partial equilibrium, i.e. holding other prices and income constant.[24] This example includes not only iso-elastic demand curves as special cases, but also a variety of other demand curves used in the literature, such as the PIGL family, the Pollak family, and QMOR.

In general equilibrium, other prices and income may potentially affect all four determinants of the demand curve: γ_i, ν_i, δ_i and σ_i.[25] Allowing for Λ and indirect utility as aggregators, Proposition 2 indicate that bi-power demand must then take the form:

$$q_i = \frac{\alpha_i(V)[F(\Lambda,V)p_i/w]^{-\nu_i(V)} + \beta_i(V)[F(\Lambda,V)p_i/w]^{-\sigma_i(V)}}{H(\Lambda,V)}$$

where α_i, β_i, ν_i and σ_i are now functions of utility. As utility (or income) increases, different goods i may be associated with smaller or larger demand, and may be associated with higher or smaller price elasticities.

In particular, it may be convenient to restrict to iso-elastic demand shifters: $F(\Lambda) = \Lambda$ and $H(\Lambda) = \Lambda^{-\eta}$. For the "demand manifold" to remain invariant to utility (see Mrázová and Neary, 2013), one must also assume that the exponents ν_i and σ_i are constant, which yields:

[24] We can also examine bi-power inverse demand in a similar fashion.

[25] A property highlighted by Mrázová and Neary (2013) is that the relationship between the price elasticity and the curvature of demand (the "demand manifold") depends only on the exponents ν_i and σ_i, and is invariant to shocks in the demand shifters, γ_i and δ_i.

$$q_i \;=\; \alpha_i(V)\Lambda^{\eta}[\Lambda p_i/w]^{-\nu_i} + \beta_i(V)\Lambda^{\eta}[\Lambda p_i/w]^{-\sigma_i}. \tag{35}$$

Applying Proposition 5, such demand system can be rationalized if either $\min\{\nu_i, \sigma_i\} > \eta$ or $\max\{\nu_i, \sigma_i\} < \eta$, and if the expression above is strictly decreasing in V.[26]

A non-homothetic generalization of QMOR. An interesting special case of (35) is when the coefficient η is equal to one of the two exponents for prices. This happens to provide a generalization of symmetric QMOR studied in Section (4.2). A convenient feature is that we can solve explicitly for the aggregator Λ as a function of indirect utility. Borrowing a similar functional form as homothetic QMOR, we can obtain a more general specification where price effects are very similar to QMOR, yet allow for more flexible Engel curves. Such generalization remains a special case of the two-aggregator demand systems described in Proposition 5.

Suppose that $\nu_i = \nu > 1$ and $\sigma_i = \sigma > 1$ are identical across all goods and that $\sigma < \nu$, and suppose that $\eta = \nu$, we can obtain an explicit solution for the aggregator Λ as a function of prices, utility and income:

$$\Lambda^{\sigma-1} \;=\; \sum_i \beta_i(V)\left(\frac{p_i}{w}\right)^{1-\sigma}.$$

Indirect utility can then be seen as an implicit solution of an equation that no longer involves Λ:

$$\sum_i \alpha_i(V)\left(\frac{p_i}{w}\right)^{1-\nu} + \left(\sum_i \beta_i(V)\left(\frac{p_i}{w}\right)^{1-\sigma}\right)^{\frac{1-\nu}{1-\sigma}} \;=\; 1.$$

If $\alpha_i(V)$ and $\beta_i(V)$ are positive, assuming that they strictly decrease with V provides a sufficient condition for this indirect utility function to coincide with rational consumer preferences.

Such demand system then yields a demand that features substitution and price effects that are very similar to homothetic QMOR:

$$q_i \;=\; \alpha_i(V)\left(\frac{p_i}{w}\right)^{-\nu} + \beta_i(V)\left(\frac{p_i}{w}\right)^{-\sigma}\left(\sum_j \beta_j(V)\left(\frac{p_j}{w}\right)^{1-\sigma}\right)^{\frac{\sigma-\nu}{1-\sigma}}$$

and now allows for richer income effects through the functions α_i and β_i which can both flexibly demand on indirect utility. This demand system also provides a generalization of non-homothetic CES preferences described in Proposition 4 in the limit case where $\nu = \sigma$. As noted previously, we could even allow ν and σ to be functions of indirect utility V, but the combination of $\alpha_i(V)$ and $\beta_i(V)$ already provide a way to parameterize how income affects the curvature of indifference curves.

[26] If $\alpha_i(V)$ and $\beta_i(V)$ are positive, a sufficient condition is that they both decrease with V. We can also allow $\beta_i(V)$ to be negative, which leads to choke prices as discussed in Section 4.4.

Modeling richer income effects. More generally, suppose that demand q_i for product i is provided by a demand curve $\widetilde{D}_i(p_i)$ in partial equilibrium, holding utility and other aggregates constant. Mrázová and Neary (2013) indicate that any of such demand curve can be obtained from a directly-additive utility function, in which case demand can be specified as $q_i = \widetilde{D}_i(\Lambda p_i)$ where Λ captures the response to all other changes in prices and income. However, our results indicate many other ways to rationalize such demand curves with more flexible Engel curves and richer income effects. First, using Proposition 3, we can derive such demand curve from a Gorman-Pollak demand system $q_i = \widetilde{D}_i(F(\Lambda)p_i)/H(\Lambda)$ where changes in other prices and income influence both the price shifter F and the quantity shifter H. Going one step further, Proposition 5 shows that we can make such demand system even more flexible by specifying $q_i = \widetilde{D}_i(F(\Lambda, V)p_i, V)/H(\Lambda, V)$. The non-homothetic versions of QMOR and bi-power demand (above) provide two examples. Additional examples are described in the Appendix, e.g. based on linear demand.

4.2 Homotheticity

There are many reasons for which one may want to impose homotheticity, e.g. to allow for simple aggregation properties across consumers with heterogeneous income levels, to provide a straightforward interpretation of price indices, or to model balanced growth paths with multiple sectors. The homothetic double-aggregator specification described in this section offers a parsimonious yet flexible framework that encompasses various examples of homothetic preferences that have been used in the literature.

In the double-aggregator homothetic case, the demand shifters F and H can be expressed as a function of the aggregator Λ only, while demand depend on both Λ and the ideal price index P:

$$q_i \;=\; \frac{w}{H(\Lambda)P} D_i\left(\frac{F(\Lambda)p_i}{P}\right) \tag{36}$$

where aggregator Λ can be implicitly defined by the budget constraint as in equation (24).

The ideal price index P is then implicitly defined by the following equation:

$$\sum_i \int_{y=y_{i0}}^{\frac{p_i F(\Lambda)}{P}} D_i(y)dy \;-\; \int_{\lambda=\Lambda_0}^{\Lambda} F'(\lambda)H(\lambda)d\lambda \;=\; c_0. \tag{37}$$

Similarly, utility U can be implicitly defined as the solution of:

$$\sum_i \int_{q=q_{i0}}^{\frac{q_i H(\Lambda)}{U}} D_i^{-1}(q)dq \;-\; \int_{\lambda=\Lambda_0}^{\Lambda} H'(\lambda)F(\lambda)d\lambda \;=\; c_1 \tag{38}$$

where c_0, c_1 and Λ_0 are constant terms. Note that Λ is such that the partial derivative of the left-hand side w.r.t Λ is null for both (37) and (38). It is also straightforward to check that the implicit solution for utility in (38) is homogeneous of degree one in quantities q.

In spite of imposing homotheticity, this specification offers rich price effects, especially if we compare them to CES preferences: it allows for a flexible specification of each demand curve thanks to D_i, and allows for competition (through the aggregator Λ) to shift demand curves vertically (through the price shifter F) or horizontally (through the quantity shifter H).

Below we review several special cases of the homothetic demand described in (36), including the three cases presented in Matsuyama and Ushchev (2017) and Feenstra (2018)'s QMOR:

Homothetic Single Aggregator. This Gorman-Pollak demand system is homothetic if and only if $H(\Lambda)F(\Lambda)$ is constant or if it is CES. In the former case, without loss of generality we assume that $F(\Lambda) = 1/H(\Lambda) = \Lambda$. A homogeneous utility representation is then given by:

$$\log U(q) \;=\; \log(\Lambda) \;+\; \sum_i \int_{x=x_{i0}}^{q_i/\Lambda} D_i^{-1}(x)dx$$

where Λ is such that $\sum_i (q_i/\Lambda) D_i^{-1}(q_i/\Lambda) = 1$, and x_{i0} are constant terms. In this case, the single aggregator Λ is homogeneous of degree one in quantities q_i in the primal version. We can also express Λ as a function of prices p_i, and write expenditure shares as:

$$p_i q_i / w \;=\; \Lambda p_i \; D_i\left(\Lambda p_i\right). \tag{39}$$

This specification is particularly attractive for empirical purposes, as it allows for flexible demand curves D_i and yet a single aggregator Λ to capture income as well as all other prices.

Homothetic Direct Implicit Additivity. When $H(\Lambda) = 1$ is constant, utility can be defined implicitly with a simple expression that does not involve aggregator Λ. In the homothetic case, this yields:

$$\sum_i \int_{q=q_{i0}}^{\frac{q_i}{U}} D_i^{-1}(q)dq \;=\; 1. \tag{40}$$

This case is described in Matsuyama and Ushchev (2017), and also corresponds to Kimball (1995) when D_i^{-1} is identical across goods. Demand for good i corresponds to:

$$q_i \;=\; (w/P) D_i\left(\Lambda p_i / P\right) \tag{41}$$

where Λ can again be defined implicitly by the budget constraint.

Homothetic Indirect Implicit Additivity. Symmetrically, when $F(\Lambda) = 1$, indirect utility and the price index can be defined implicitly without involving aggregator Λ. For the ideal price index, we obtain:

$$\sum_i \int_{y=y_{i0}}^{\frac{p_i}{P}} D_i(y)dy \;=\; 1. \tag{42}$$

In this case, demand corresponds to:

$$q_i \;=\; \frac{wD_i\left(p_i/P\right)}{\sum_j p_j D_j\left(p_j/P\right)}. \tag{43}$$

Homothetic semi-separability. In the case of semi-separable preferences, we obtain explicit expressions for both the price index and utility. Suppose that the price index is defined by:

$$P^{-\eta} \;=\; \eta\Lambda^\eta \sum_i S_i(\Lambda p_i) \;=\; \Lambda^\eta \sum_i (\Lambda p_i) D_i(\Lambda p_i)$$

with $D_i = -S_i'$ and $\eta > 0$, and with Λ implicitly defined by the second equality. The integrability condition on elasticities imposes that price elasticities ε_{Di} are either always greater or always smaller than $-(\eta + 1)$ across all goods and price levels.[27] Based on this specification, demand for good i is then:

$$q_i \;=\; wP^\eta \Lambda^{\eta+1} D_i(\Lambda p_i).$$

In this homothetic case, an interesting property is that direct semi-separability is equivalent to indirect semi-separability, and utility can alternatively be defined by:

$$U^{\frac{\eta}{1+\eta}} \;=\; \Lambda^{\frac{\eta}{1+\eta}} \sum_i \int_0^{q_i/\Lambda} D_i^{-1}(q)dq$$

where Λ can be defined as a function of q such that the derivative of the expression above is null. Note also that the three special cases of homothetic preferences discussed above (HSA, HDIA and HIIA) are distinct from this type of demand. The invariance properties of such a demand system can be useful for tractability (see Proposition 8 in Section 5.2).

Symmetric QMOR. QMOR preferences have first been studied by Diewert (1976) and more recently studied by Feenstra (2018) imposing some symmetry in the price effects. Take $D_i(y) = \alpha_i y^{r-1} + \beta_i y^{\kappa r-1}$ and $F(\Lambda) = \Lambda$ and $H(\Lambda) = \Lambda^{r-1}$ with $r < 0$ and $\kappa \in (0, 1)$. In this

[27] In the later case, apply Proposition 5 with a change in variable $\Lambda' = 1/\Lambda$.

case, we can obtain an explicit expression both for the price index and the aggregator Λ:

$$P^r \;=\; \sum_i \alpha_i p_i^r \;+\; \Big(\sum_i \beta_i p_i^{\kappa r}\Big)^{\frac{1}{\kappa}} \qquad ; \qquad \Lambda^{-\kappa r} \;=\; \sum_i \beta_i \left(\frac{p_i}{P}\right)^{\kappa r}.$$

Demand is then:

$$q_i \;=\; \frac{w}{P}\left(\frac{p_i}{P}\right)^{r-1}\left[\alpha_i + \beta_i\left(\frac{\Lambda p_i}{P}\right)^{-r(1-\kappa)}\right].$$

Note that this is a special case of the non-homothetic QMOR presented earlier. With $\kappa = 1/2$, symmetric $\alpha_i = \alpha$ and $\beta_i = \beta$, we obtain the symmetric QMOR specification as in Feenstra (2018). When $\alpha > 0$ and $\beta < 0$, note that we get a finite reservation price (choke price). We discuss such possibility below in Section 4.4.

Other examples. Yet other examples of homothetic demand with two aggregators are the homothetic translog and linear demand, discussed below in Section 4.4 and in the Appendix.

4.3 Double-shifter demand system

Thisse and Ushchev (2016) show that the following demand system can be obtained by aggregating over many consumers who make indivisible consumption choices among horizontally-differentiated product varieties:

$$q_i \;=\; Q(p/w)D_i(F(p/w)p_i/w) \tag{44}$$

where Q and F are two aggregators, i.e. two continuously-differentiable mappings from $\mathbb{R}_+^N$ to $\mathbb{R}_+^N$, and D_i is a continuously-differentiable mappings from $\mathbb{R}_+$ to $\mathbb{R}_+$. Note that the budget constraint implies: $Q(p/w) = 1/\sum_j (p_j/w)D_j(F(p/w)p_j/w)$.

For instance, aggregate consumption that mimics indirectly-additive preferences can be obtained by aggregating over consumers with multinomial logit idiosyncratic utility terms. As discussed in Thisse and Ushchev (2016), more general aggregate consumption patterns can be obtained with alternative distributions of random utility terms across consumers.

The demand system specified in equation (44) is easy to manipulate, estimate, and provides a natural extension of directly-additive and indirectly-additive preferences which only include one of the two demand shifters. For instance, this specification of demand is used in Arkolakis et al. (2019) to obtain a so-called "gravity equation" for aggregate trade between countries, in a model with heterogeneous firms and asymmetric countries. Arkolakis et al. (2019) give examples of demand with such structure but do not provide a micro-foundation for the more general functional form.

At first sight, this does not appear to be a special case of the demand systems used in Proposition 2 and 5, which assume that the indirect utility V to be one of the two aggregators. However, with symmetry and rank arguments, we can show that the triplet of gradients of F, Q and V cannot have a rank higher than two if the demand system is integrable. Hence, such demand can actually be re-expressed using shifters that are functions of just utility V and another aggregator Λ.

Proposition 6 *Suppose that the demand system takes the form given by (44), that it is integrable, and that the pair of gradients* $\left\{\frac{\partial Q}{\partial \log p}, \frac{\partial F}{\partial \log p}\right\}$ *has rank two for all* (p, w). *Demand can then be written as:*

$$q_i p_i / w \;=\; \widetilde{Q}(\Lambda, V) D_i(\widetilde{F}(\Lambda, V) p_i / w) \tag{45}$$

for some functions $\widetilde{Q}$ *and* $\widetilde{F}$ *of indirect utility* V *and a common aggregator* Λ.

Hence, when such demand is integrable (with a representative consumer) it can be seen as a special case of the demand systems examined in Proposition 5.

Such a demand system correspond to Gorman-Pollak demand with a single aggregator when Q and F can be written as a function of a single aggregator Λ instead of two aggregators (Λ, V), but in this case the gradients of the two aggregators, $\frac{\partial Q}{\partial \log p}$ and $\frac{\partial F}{\partial \log p}$, must be colinear.

4.4 Demand with choke prices

In various applications, demand from a consumer may be equal to zero if the price of a good is too high. Such upper bound is called a choke price or reservation price. This is an often desired feature for estimation, as zeros are prevalent in microdata at the finest level, and for applied modeling, e.g. to generate non-trivial extensive margins and explain selection across markets.[28] One would want such choke price to be an equilibrium outcome, and depend on consumer income and the toughness of competition. For instance, there is substantial evidence (see e.g. Hummels and Klenow, 2005) that richer consumers buy a larger variety of products.

The Gorman-Pollak and double-aggregator demand system studied above can be accommodated to yield such choke prices. With a demand structure as in Proposition 5, suppose that $D_i(y_i, V) = 0$ for all $y_i \geq a_i(V)$ in the double-aggregator case —this becomes $D_i(y_i) = 0$ for

[28] Choke prices are particularly 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. (2019) among others.

all $y_i \geq a_i$ in the single-aggregator case (Gorman-Pollak demand, Proposition 3). Most of the results shown previously hold if, with a slight abuse of notation, we define $D_i^{-1}(0, V) = a_i(V)$.

In this framework, the choke price p_i^* depends on income, utility and the aggregator Λ. For a consumer with income w, aggregator Λ and utility V, demand for good i is null if and only if:

$$p_i \;\geq\; p_i^* \;=\; \frac{a_i(V)w}{F(\Lambda, V)}.$$

With a single aggregator, the choke price has a more restrictive functional form: $p_i^* = \frac{a_i w}{F(\Lambda)}$.

The choke price is proportional to income when preferences are indirectly additive since the terms a_i and F are constant in this case. Bertoletti et al. (2018) exploit this property to obtain a tractable model of trade and argue that it fits key patterns of how prices vary with income and population across markets. [29] A similar property can be obtained with implicitly indirectly additive preferences (see Section 4.1) as the choke price would then just depend on income and utility.

The most simple case of demand with choke prices is one that is linear in its own price. As shown in the Appendix, there are various ways to generate such demand with one or two aggregators that influence how other prices and income shift demand vertically and horizontally.

Another tractable example used in the macroeconomic and trade literature is the Translog expenditure function (Feenstra, 2003; Novy, 2013). A typical assumption is that the cross-price elasticities are symmetrical. Demand associated with Translog can then be expressed as a function of a single aggregator Λ even when some varieties are not consumed (see Appendix), with expenditure shares taking the form:

$$p_i q_i / w \;=\; \alpha_i - \gamma \log(\Lambda p_i / w)$$

with a choke price $p_i^* = \exp(\alpha_i/\gamma_i) w/\Lambda$.

Yet another example of preferences with two aggregators generating choke prices is QMOR and its non-homothetic extension described earlier, with choke prices arising with $\beta_i(V) < 0$.

5 An application to monopolistic competition

Summarizing other prices by a one or two aggregators is particularly useful for applications to imperfect competition, because such aggregators synthetise all relevant information on a firm's competitors. Under monopolistic competition, assuming that each firm has a negligible

[29] See Fally (2019) for a generalization of Arkolakis et al. (2019) and Bertoletti et al. (2018) on the gains from trade, using demand with a single aggregator and a choke price.

market share (as in Dixit and Stiglitz, 1977), this aggregator can be taken as given by a specific firm.[30] This facilitates theoretical analysis of the equilibrium as well as empirical estimation, while allowing for flexible equilibrium outcomes and comparative statics. This section discusses additional restrictions needed for such applications with a continuum of goods, then examines a simple general-equilibrium model with free entry under monopolistic competition to illustrate the role of modeling choices on the demand side.

5.1 With a continuum of goods

Models of monopolistic competition typically assume a continuum of product varieties,[31] where each variety accounts for a measure zero of aggregate expenditures. Here we discuss additional assumptions that should be imposed on the structure of demand such that it is well defined and well behaved on a continuum.

The discussion provided here fits within the framework of Parenti et al. (2017). A first assumption is that the set of potential varieties is compact and is included in $[0, \bar{N}]$; such assumption is typically not restrictive and this upper bound $\bar{N}$ not binding in equilibrium if there is a fixed cost of producing a new variety and if $\bar{N}$ is large enough. A consumption profile q is now defined as a mapping from $[0, \bar{N}]$ to $\mathbb{R}_{\geq 0}$ that belongs to $L^2([0, \bar{N}])$, i.e. such that its square has a finite integral sum.[32] In this framework, utility and the aggregator Λ are two functionals, i.e. real valued functions defined over $L^2([0, \bar{N}])$. They are assumed to be symmetric over $[0, \bar{N}]$, i.e. that consumers are indifferent to switching labels across products i; here, this implies that function $D_i = D$ is identical across all goods i.

While strict quasi-concavity implies that consumers exhibit love for variety, we need to assume that utility does not drop too much when the quantity consumed $q_i = 0$ is zero for a non-trivial measure of goods. To be more precise, here we assume $\int_{0^+}^{a} D^{-1}(x)dx < \infty$ (a finite integral sum around zero). This implies that the expenditure share on a range of goods is zero in the limit if the quantity for these goods goes to zero (i.e. no good is essential): $\lim_{q_i \to 0+} q_i D^{-1}(q_i H(\Lambda, U), U) = 0$. A sufficient condition for these properties to hold is that

[30] The tools developed by Anderson et al. (2018) could be used in this case, using Λ as an aggregate. Under Bertrand competition, a firm with non-negligible market share would account for the effect of its own price on Λ, holding other prices as given. Under Cournot competition, a firm would account for the effect of its own production quantity on Λ, holding other quantities as given, using the inverse demand formulation and specifying Λ as a function of quantities instead of prices.

[31] The use of a continuum mostly originated from Industrial Organization (e.g. Vives, 1990) , and is popular in macroeconomics and international trade (e.g. Romer (1990), Grossman and Helpman, 1991, Melitz, 2003).

[32] $L^2([0, \bar{N}])$ is a natural space on which to define consumption profiles as it is a Hilbert space and includes all bounded consumption profiles. Parenti et al. (2017) use this property to prove existence of an equilibrium. A less elegant alternative to obtain completeness would be to assume an uniform upper bound on the consumption profile q within a consumer's budget set if such upper bound is not binding in equilibrium.

the elasticity of D is strictly larger than unity (or infinite) in the limit where the quantity of a good goes to zero.

Extending Proposition 5 to a continuum, utility $U(q)$ needs to satisfy:

$$\int_{i=0}^{\bar{N}} \int_{q=0}^{q_i H(\Lambda,U)} D^{-1}(q,U)\, dq\, di \;-\; G(\Lambda,U) \;=\; 0 \tag{46}$$

where aggregator Λ is itself an implicit solution to:

$$\int_{i=0}^{\bar{N}} q_i D^{-1}(q_i H(\Lambda,U),U) di \;=\; F(\Lambda,U) \tag{47}$$

and where D^{-1}, H, F and G are continuously differentiable real functions with $\frac{\partial G}{\partial \Lambda} = \frac{\partial H}{\partial \Lambda} F$.

Uniqueness of (Λ, U) is ensured by assuming that $\varepsilon_D \varepsilon_F < \varepsilon_H$ and that the left-hand-side of (46) has a negative partial derivative in U. A sufficient condition for existence of Λ (conditional on U) is that $\frac{D^{-1}(q_i H(\Lambda,U),U)}{F(\Lambda,U)}$ takes on values from $+\infty$ to 0 over the range of Λ. Existence of utility is then guaranteed if we combine the following two conditions: i) we assume that $\int_{q=0}^{q_i H(\Lambda,U)} D^{-1}(q,U)/G(\Lambda,U) dq$ spans from $+\infty$ to 0 as utility decreases (holding Λ and q_i constant); ii) we assume that it goes to zero as Λ tends to zero, for a any given U and q_i.

Finally, a key assumption imposed by Parenti et al. (2017) is that utility is Frechet-differentiable in any $q \in L^2[0, \bar{N}]$, which provides a rigorous definition of marginal utility in this context with a continuum of goods. Conditions to ensure Frechet-differentiability of U and Λ are discussed in the Appendix.[33]

While we focus here on symmetric demand, we refer to Bertoletti and Etro (2022) for a discussion of the assumptions and approximations required under monopolistic competition when preferences are asymmetric across product varieties.[34]

5.2 Market size effects

To illustrate the role of the demand side and in particular how assumptions and modeling choices influence key outcomes, the remainder of this section examines a simple general-equilibrium model with free entry under monopolistic competition with homogeneous firms. In particular, the goal is to examine how changes in market size (either from changes in population or income) affect firm size, prices, and the number of firms, depending on functional form assumptions on the demand side. A more elaborate study with heterogeneous firms, several markets, and richer

[33]Here one might be able to relax the requirement of Frechet differentiability given the existence of one or two aggregators summarizing all cross-price effects.

[34]Alternatively, one could also examine competition over a discrete number of goods, and take the limit to infinity, as in Vives (1987). Income effects would decline at a rate $1/\sqrt{N}$, or faster, with the number of goods.

interactions, is however beyond the scope of the present paper.

Model setup. Consider a single economy with a population L of identical consumers. There is a continuum of products, each of them produced by a single firm, where N denotes the measure of active firms. There is free entry of firms, who compete under monopolistic competition. Consumer preferences are described by those in the previous sub-section, with utility U and aggregator Λ satisfying equations (46) and (47).

There is only one factor of production, labor. We assume that w is the efficiency of each worker, and L is the number of workers, so that Lw is the supply of labor in efficiency units. We normalize the return of a unit of labor to unity, so that individual income is w.

All firms have access to the same technology and cost structure, so firms are homogeneous. Q denotes total production by firm, while $q = Q/L$ is the quantity consumed by variety and by worker. For each firm, the cost of producing Q is given by a constant marginal cost c and a fixed cost f, hence total costs equal $C(Q) = cQ + f$ in terms of efficiency units of labor. With a continuum of firms under monopolistic competition, each firm takes aggregates as given and unaffected by its decisions, including utility U and the aggregator Λ.

In all cases below, we assume that the price elasticity of demand is strictly larger than unity (in absolute terms), to ensure finite markups, and that the second order condition in profit maximization is satisfied. In terms of inverse demand, this implies that $QD^{-1}(Q,U)$ is concave and has a negative second derivative in Q.

Equilibrium conditions. Two equilibrium conditions describe the supply side. First, firms maximize profits. Sales for each firm are equal to production $Q = Lq$ times the price $p = wD^{-1}(HQ/L,U)/F$ where F and H themselves depend on aggregator Λ and utility U. Profits are thus: $\pi = \max_Q\{Qw\ D^{-1}(HQ/L,U)/F - cQ - f\}$. Maximizing over Q (taking Λ and U as constant under monopolistic competition) leads to the usual first order condition relating markups and the inverse of the price elasticity of demand:

$$\frac{p-c}{p} \;=\; -\frac{(HQ/L)\ (D^{-1})'(HQ/L,U)}{D^{-1}(HQ/L,U)} \;\equiv\; 1/\sigma \tag{48}$$

with $p/w = D^{-1}(HQ/L,U)/F$. The right-hand side is the inverse of the price elasticity of demand, $\sigma(HQ/L,U)$, which can be expressed as a function of utility U as well as consumption quantity Q/L multiplied by the quantity shifter $H(\Lambda,U)$.

Next, free entry implies that firms make zero profits in equilibrium: $\pi = 0$. Hence, the price

p is equal to the average cost for each firm:

$$p = wD^{-1}(HQ/L, U)/F = (cQ + f)/Q. \tag{49}$$

Two equilibrium conditions describe the demand side: equations (46) and (47) described above. With symmetry across product varieties, utility U is such that:

$$N \int_{q=0}^{H(\Lambda,U)Q/L} D^{-1}(q, U) dq \;=\; G(\Lambda, U) \tag{50}$$

while the budget constraint can be written:

$$(NQ/L)\; D^{-1}(H(\Lambda, U)\, Q/L\,,\, U) \;=\; F(\Lambda, U). \tag{51}$$

Note that combining the budget constraint (51) and the free entry condition (49) leads to $N(cQ + f) = Lw$ (regardless of the demand system), which we will refer to as the "resource constraint".

We define an equilibrium as a set (Q, N, U, Λ) satisfying conditions (48), (49), (50) and (51).

Market size effects across preferences specifications. A central question, with implications for various fields in economics, is how prices and firm size depend on market size, where market size itself can be thought of as the product of population and per capita income. As shown in e.g. Parenti et al. (2017), in such a model we already know that price p and firm size Q are independent of income w when preferences are directly additively separable; independent of population L when preferences are indirectly additively separable; and fully determined by total GDP when preferences are homothetic.

Here, first we show that single-aggregator demand can generate a wide range of comparative statics in terms of key outcomes. Conversely, we can construct various demand systems that maintain independence of firm size and prices with respect to population L or income w. Finally, even with the most general form of demand in aggregator Λ and utility V, we still obtain sharp welfare comparisons to the first-best allocation.

In all these cases, comparative statics depend crucially on whether demand is "superconvex" or "subconvex" (see e.g. Mrázová and Neary, 2013). We say that demand is superconvex if the price elasticity of demand σ increases with sales (i.e. if $\varepsilon_\sigma > 0$), and subconvex if σ decreases with sales (if $\varepsilon_\sigma < 0$). As earlier, ε_F and ε_H denote the elasticity of F and H in Λ. We start with Gorman-Pollak demand with a single aggregator Λ:

Proposition 7 *Suppose that demand is Gorman-Pollak with a single aggregator and that the second-order condition for profits maximization is satisfied:*

i) For any value of fixed and marginal costs, an equilibrium exists and is unique.

ii) Comparative statics in firm size Q (with opposite effects on prices p) are described in Table 1, with all possible combinations of signs depending on the sign of ε_σ, ε_H and ε_F.

Table 1: Comparative statics: market size effects on production Q

Sign of effect on Q	Subconvex case ($\varepsilon_\sigma < 0$)		Superconvex case ($\varepsilon_\sigma > 0$)	
	Effect of L	Effect of w	Effect of L	Effect of w
$\varepsilon_H = 0$ & $\varepsilon_F > 0$ (DA)	(+)	(0)	(–)	(0)
$\varepsilon_H > 0$ & $\varepsilon_F = 0$ (IA)	(0)	(–)	(0)	(+)
$\varepsilon_H < 0$ & $\varepsilon_F < 0$	(–)	(+)	(+)	(–)
$\varepsilon_H < 0$ & $\varepsilon_F > 0$	(+)	(+)	(–)	(–)
$\varepsilon_H > 0$ & $\varepsilon_F < 0$	(–)	(–)	(+)	(+)
$\varepsilon_H > 0$ & $\varepsilon_F > 0$	(+)	(–)	(–)	(+)

Notes: Sign of the effect of an increase in population L and income w on production Q in the single-aggregator case (Gorman-Pollak demand); effects on prices are opposite; effects of increases in L and w on the aggregator Λ are always positive. DA: directly-additive; IA: indirectly-additive separable preferences

This proposition (see proof in the Appendix) highlights the usefulness of demand with a single aggregator. First, having demand depend on just a single aggregator Λ, aside from the price or quantity, greatly simplifies the analysis of the market equilibrium and automatically implies uniqueness. Thanks to an envelope theorem argument, the free entry condition implies a tight positive relationship between market size proxies (L and w) and the aggregator Λ, which can be used as an indicator for the toughness of competition.

In the meantime, such demand is sufficiently flexible to generate a wide range of comparative statics, thanks to the different effects of the aggregator on the price shifter F and the demand shifter H. Those respectively determine the signs of the effects of population and income on firm size, with all combinations possible. These signs are switched in subconvex ($\varepsilon_\sigma < 0$) vs. superconvex cases ($\varepsilon_\sigma > 0$). In particular, this offers more flexibility than direct and indirect additive separability (first two rows of Table 1).

With increasing returns to scale, the effects on prices are always opposite to those on firm size. Moreover, given the resource constraint $cQ + f = Lw/N$, changes in firm size dictate whether entry increases or decreases relative to total income Lw. In the case where an increase in market size leads to larger firm size, it is actually possible that the number of firms N decreases (with an increase in either population L or income w).

In a more detailed analysis, Bertoletti and Etro (2021) also study market size effects within a similar model and demand structure, imposing either $\varepsilon_H = 1$ or $\varepsilon_F = 1$, and examine extensions with heterogeneous firms.

Analysis with demand that depends on both aggregator Λ and utility V is more difficult, because both can be influenced by the toughness of competition. In particular, the equilibrium is not always unique with more general demand patterns. We discuss several cases below where simple sufficient conditions for uniqueness can still be obtained.

For instance, a variety of comparative statics can also be achieved with (directly or indirectly) implicitly-additive preferences. Based on the specification of equation (27), the price elasticity of substitution can be a flexible function of both quantities Q/L and the level of utility U, which itself depends on the number of firms N. As shown by Parenti et al. (2017), flexibility with respect to these two arguments allows generating a wide gallery of comparative statics (see Appendix). In the empirically-relevant case where the price elasticity is decreasing with firm size (subconvex demand) and decreasing with utility (as estimated in Faber and Fally, 2020 and Auer et al., 2021), the equilibrium is unique, and firm size Q decreases with income while prices increase with income. Qualitatively, a larger population L can lead to either larger or smaller firms in equilibrium, depending on the shape of demand and income effects.

Conversely, for tractability, one may ask whether firm size and prices can remain independent of either population or income with preferences that are neither directly nor indirectly-additive. The following proposition extends these convenient properties to the semi-separable preferences described in Section 4.1:

Proposition 8 *Suppose that preferences are semi-separable, as defined in Section 4.1, and that demand is either subconvex or superconvex (i.e. the sign of ε_σ does not flip):*

*i) If preferences are **directly** semi-separable, firm size Q and price p do not depend on income w. If $\varepsilon_H < 0$, the equilibrium is unique (sufficient condition). If, in addition, demand is subconvex ($\varepsilon_\sigma < 0$), firm size increases with population L.*

*ii) If preferences are **indirectly** semi-separable, firm size Q and price p do not depend on population L. If $\varepsilon_F > 0$, the equilibrium is unique (sufficient condition). If, in addition, demand is subconvex ($\varepsilon_\sigma < 0$), firm size decreases with income w.*

With both direct and indirect semi-separability, we can also easily show (see Appendix) that existence of an equilibrium is guaranteed when the demand for a good (conditional on its own normalized price) spans from zero to infinity over the range of the aggregator Λ, as we assume here.

In the case of directly semi-separable preferences (i), uniqueness can be shown by expressing firm size Q as a function of Λ using the markup equation and expressing aggregator Λ as a function of Q by combining equations (50) and (51); the latter yields a condition such as equation (30) that is independent of utility and the number of firms N. As shown in the Appendix, a simple condition such as $\varepsilon_H < 0$ ensures that these two relationships have a unique fixed point.[35] Moreover, income (or utility) does not appear in these relationships, which implies that equilibrium firm size does not depend on income w. If firm size does not change as income increases, the number of firms must then increase proportionally with income, given the resource constraint $N(cQ + f) = Lw$.

The case of indirectly semi-separable preferences (ii) is similar but relies on the dual. Using the markup equation for prices instead of quantities, one can uniquely express prices as a function of aggregator Λ. In turn, equation (33) characterizes the aggregator as a function of prices and income. Neither equation depends on population L, and a simple condition such as $\varepsilon_F > 0$ ensures that these relationships have a unique fixed point. The equilibrium is then unique and firm size does not depend in population L. Again, the resource constraint implies that the number of firms N is proportional to population.

Finally, another type of independence is obtained for homothetic preferences. In this case, the results of Parenti et al. (2017) also apply: production Q, the number of firms N and prices p (relative to the unit cost of labor) depend only on total GDP (i.e. Lw) and do not depend on L and w individually, conditional on total GDP.

Proposition 9 *Suppose that preferences are homothetic, as defined in Section 4.1, and that the second-order condition for profits maximization is satisfied:*

i) If demand is subconvex ($\varepsilon_\sigma < 0$) and $\varepsilon_H < 0$, the equilibrium is unique and firm size increases with market size.

ii) If demand is superconvex ($\varepsilon_\sigma > 0$) and $\varepsilon_H + \varepsilon_F < 0$, the equilibrium is unique but an increase in market size does not necessarily lead to an increase in firm size.

iii) If preferences are homothetic and semi-separable, firm size and prices depend on neither population nor income.

The case of homothetic semi-separable preferences is noteworthy (iii), as it leads to firm size, markups and prices that are independent of market size (both income w and population L) in equilibrium. This property is the same as with CES demand, except that it allows for more flexible demand curves and variable markups. It can be practical in situations where one would want to shut down such adjustment channels on firm size for simplicity and tractability.

[35] See Appendix for weaker sufficient conditions and additional comparative statics.

Taken together, these examples illustrate how functional form assumptions made on the demand side influence key results on market size effects related to firm size, entry, and prices in general equilibrium models. These results highlight the need for flexible forms (unless we want to purposefully shut down some specific channels) and show how demand with one or two aggregators (Λ and U) can provide a rich and tractable framework.

Entry relative to the first-best allocation. Last but not least, a central question is whether entry is welfare maximizing, excessive, or insufficient. Following the insights from Matsuyama and Ushchev (2020), one can obtain a clear and simple answer even in the most general case with both aggregators Λ and V (as in Proposition 5), depending on whether demand is subconvex or superconvex.

Looking first at the demand side, one can see that the relative welfare gains from increased variety $d\log N$ vs. increased consumption $d\log Q$ (per variety) are determined by:

$$\frac{\frac{dU}{d\log N}}{\frac{dU}{d\log Q}} = v\Big(H(\Lambda)Q/L\,,\,U\Big)$$

where $v(q)$ is the inverse of the elasticity of $\int_{q'=0}^{q} D^{-1}(q',U)\,dq'$ w.r.t q. We can interpret $v(q)$ as capturing the relative gains from product variety. At the first-best allocation, i.e. optimal entry N and production Q given the resource constraint $cQ + f = Lw/N$ (total costs across all firms equal total GDP), the indifference curve in terms of N and Q must be tangent to the manifold in N and Q implicitly defined by the resource constraint, and thus the relative gains from variety must equal the ratio of average costs to marginal costs:[36]

$$v\Big(H(\Lambda)Q/L\,,\,U\Big) = \frac{cQ+f}{cQ}. \tag{52}$$

The market equilibrium, however, imposes that the ratio of average cost over marginal cost must be equal to the relative markup p/c (zero profit condition), which itself is determined by the price elasticity of demand:

$$\frac{cQ+f}{cQ} = \frac{\sigma}{\sigma-1} \tag{53}$$

where $\sigma = \sigma\Big(H(\Lambda)Q/L\,,\,U\Big)$ is a function of utility U, production Q and the price aggregator.

Hence, entry is excessive (resp. insufficient) when the gains from variety v is smaller (resp. larger) than the markup $\frac{\sigma}{\sigma-1}$. As discussed for instance in Vives (1999),[37] markups capture the

[36] Note that, given the resource constraint, inefficient entry N is equivalent to inefficient firm size Q.

[37] Vives (1999) uses Spence (1976) quasi-linear preferences in partial equilibrium but the comparison between markups and gains from variety plays a similar role in such setting.

gains from variety on the supply side and must equal the gains from variety on the demand side if we are in a welfare-maximizing allocation.

As Matsuyama and Ushchev (2020) pointed out, one cannot rank these two statistics in general without additional structure. In their analysis, they focus on three cases of homothetic preferences (discussed in Section 4.2). Here, we extend their insight by considering the preferences described in Proposition 5 (in U and Λ), where movements along indifference curves can still be expressed as changes in a single scalar variable Λ. As in Matsuyama and Ushchev (2020), this allows us to rank markups and the gains from variety depending on whether demand is subconvex or superconvex. We obtain the following proposition:

Proposition 10 *Suppose that demand is as in Proposition 5. Comparing the market equilibrium to the first-best allocation:*

i) If demand is subconvex ($\varepsilon_\sigma < 0$), there is excessive entry ($v < \frac{\sigma}{\sigma-1}$).

ii) If demand is superconvex ($\varepsilon_\sigma > 0$), there is insufficient entry ($v > \frac{\sigma}{\sigma-1}$).

This result crucially relies on having a synthetic price aggregator (conditional on utility) and would not be obtained if demand depended on two or more aggregators in addition to utility.

6 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, total expenditures and a good's own price ("generalized separability", following the terminology of Pollak 1972). Here I show that such a 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 has not been formally demonstrated and is not well known today in spite of its usefulness. Furthermore, I show that these two types of demand systems can be rationalized (i.e. can be derived from well-behaved utility functions) under fairly mild regularity restrictions that guarantee a well-behaved quasi-concave utility.

The first case of demand allows for flexible price effects but more restricted income effects. This case encompasses directly and indirectly additive preferences, and homothetic demand with a single aggregator described in Matsuyama and Ushchev (2017).

The second case of demand allows for flexible income effects (Engel curves) but more restricted price effects; Allen-Uzawa substitution elasticities have to be constant across goods to ensure the symmetry of the Slutsky matrix but they may increase or decrease with utility, and thus vary indirectly with income. In that second case, the aggregator actually coincides with indirect utility.

This paper further extends these results to demand systems that allow for a price aggregator as well as indirect utility (which can be interpreted as an additional aggregator), thus allowing for combinations of the two cases of demand mentioned above. We can again characterize the functional form that such demand must take, provide sufficient conditions to ensure that it can be rationalized, and characterize the utility function associated with such demand systems. This allows for greater flexibility and encompasses a wider set of demand systems frequently used in the literature, thus providing a unified general structure. Special cases include implicitly-additive preferences, all three types of homothetic demand described in Matsuyama and Ushchev (2017), QMOR preferences as in Feenstra (2018), and double-aggregator demand as in Thisse and Ushchev (2016) and Arkolakis et al. (2019).

There can be numerous applications and uses of such demand systems with a price aggregator. Recent research in macroeconomics, international trade, industrial organization, and development economics has highlighted in different contexts the crucial role of the demand side and its interactions with income disparities, fostered by an increased availability of precise micro-data on consumption baskets across households, such as scanner data. This paper aims to provide useful tools to model richer price and income effects in a tractable manner, for both theoretical and empirical applications. For instance, we show that having a single-aggregator helps in obtaining uniqueness of equilibrium, providing simple criteria for comparative statics, and drawing sharper conclusions in terms of excessive entry relative to a first-best allocation, in spite of allowing for richer demand patterns.

References

Anderson, S. P., N. Erkal, and D. Piccinin (2018). Aggregative games and oligopoly theory: Short-run and long-run analysis. *Working paper*.

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

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2019). The elusive pro-competitive effects of trade. *The Review of Economic Studies 86*(1), 46–80.

Atkin, D., B. Faber, T. Fally, and M. Gonzalez-Navarro (2020). Measuring welfare and inequality with incomplete price information. *Working paper*.

Auer, R., A. Burstein, S. Lein, and J. Vogel (2021). Unequal expenditure switching: Evidence from switzerland. *Working paper*.

Bertoletti, P. and F. Etro (2016). Preferences, entry, and market structure. *The RAND Journal of Economics 47*(4), 792–821.

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

Bertoletti, P. and F. Etro (2021). Monopolistic competition with generalized additively separable preferences. *Oxford Economic Papers 73*(2), 927–952.

Bertoletti, P. and F. Etro (2022). Monopolistic competition, as you like it. *Economic Inquiry 60*(1), 293–319.

Bertoletti, P., F. Etro, and I. Simonovska (2018). International trade with indirect additivity. *American Economic Journal: Microeconomics 10*(2), 1–57.

Blackorby, C., R. Davidson, and W. Schworm (1991). Implicit separability: Characterisation and implications for consumer demands. *Journal of Economic Theory 55*(2), 364–399.

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

Comin, D. A., D. Lashkari, and M. Mestieri (2021). Structural change with long-run income and price effects. *Econometrica 89*(1), 311–374.

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

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of econometrics 4*(2), 115–145.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *The American economic review 67*(3), 297–308.

Faber, B. and T. Fally (2020). Firm heterogeneity in consumption baskets: Evidence from home and store scanner data. *Review of Economic Studies, Forthcoming*.

Fabinger, M. and E. G. Weyl (2016). Functional forms for tractable economic models and the cost structure of international trade. *Working paper*.

Fally, T. (2019). Generalized separability and the gains from trade. *Economics Letters 178*.

Feenstra, R. C. (2003). A homothetic utility function for monopolistic competition models, without constant price elasticity. *Economics Letters 78*(1), 79–86.

Feenstra, R. C. (2018). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. *Journal of International Economics 110*.

Gorman, W. M. (1972). Conditions for generalized additive separability. Unpublished transcript, now printed in Gorman (1995).

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.

Grossman, G. M. and E. Helpman (1991). Quality ladders in the theory of growth. *The review of economic studies 58*(1), 43–61.

Handbury, J. (2021). Are poor cities cheap for everyone? Non-homotheticity and the cost of living across US cities. *Econometrica 89*(6), 2679–2715.

Houthakker, H. S. (1965). A note on self-dual preferences. *Econometrica*, 797–801.

Hummels, D. and P. J. Klenow (2005). The variety and quality of a nation's exports. *American Economic Review 95*(3), 704–723.

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

Kimball, M. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking 27*(4), 1241–77.

Krugman, P. R. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics 9*(4), 469–479.

LaFrance, J. T. and R. D. Pope (2006). Full rank rational demand systems. *CUDARE WP*.

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

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. (2019). Engel's law in the global economy: Demand-induced patterns of structural change, innovation, and trade. *Econometrica 87*(2), 497–528.

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

Matsuyama, K. and P. Ushchev (2020). When does procompetitive entry imply excessive entry?

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica 71*(6), 1695–1725.

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*.

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

Novy, D. (2013). International trade without ces: Estimating translog gravity. *Journal of International Economics 89*(2), 271–282.

Parenti, M., P. Ushchev, and J.-F. Thisse (2017). Toward a theory of monopolistic competition. *Journal of Economic Theory 167*, 86–115.

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

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy 98*(5).

Samuelson, P. A. (1950). The problem of integrability in utility theory. *Economica 17*(68), 355–385.

Spence, M. (1976). Product selection, fixed costs, and monopolistic competition. *The Review of economic studies 43*(2), 217–235.

Thisse, J.-F. and P. Ushchev (2016). When can a demand system be described by a multinomial logit with income effect? *Higher School of Economics Research Paper No. WP BRP 139.*

Vives, X. (1987). Small income effects: A Marshallian theory of consumer surplus and downward sloping demand. *The Review of Economic Studies 54*(1), 87–103.

Vives, X. (1990). Trade association disclosure rules, incentives to share information, and welfare. *the RAND Journal of Economics*, 409–430.

Vives, X. (1999). *Oligopoly pricing: old ideas and new tools.* MIT press.

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.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012). Monopolistic competition: Beyond the constant elasticity of substitution. *Econometrica 80*(6), 2765–2784.

Appendix: Proofs and additional derivations

Proposition 1

Preliminaries: Inverse demand Consider the demand system:

$$q_i = \widetilde{q}_i(p_i/w, \Lambda(p/w)).$$

Following condition [A1]-iv), for the sake of exposition we assume for the most part that for any $q \in \mathbb{R}_+^N$, there exists a vector of normalized prices $p/w \in \mathbb{R}_+^N$ that generates demand q, i.e. such that $q_i = \widetilde{q}_i(p_i/w, \Lambda(p/w))$.[38]

First, note that Λ can be seen an implicit function of normalized prices p_i/w such that the budget constraint holds, i.e. such that:

$$\sum_i (p_i/w)\, \widetilde{q}_i(p_i/w, \Lambda) = 1.$$

If we assume that each $q_i(p_i/w, \Lambda)$ is strictly decreasing in Λ (here we assume a strictly negative derivative), the solution in Λ is unique and continuously differentiable.

Since we assume that expenditure shares $(p_i/w)\, q_i(p_i/w, \Lambda)$ monotonically decreases or increases with prices (holding Λ constant), demand can be inverted such that expenditure shares can be obtained as a function of q_i and the aggregator Λ:

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

i.e. such that $(p_i/w)\widetilde{q}_i(p_i/w, \Lambda) = W_i(\widetilde{q}_i(p_i/w, \Lambda), \Lambda)$ for any $\Lambda = \Lambda(p/w)$ and $p/w \in \mathbb{R}_+^N$. As demand $\widetilde{q}_i(p_i/w, \Lambda)$ has a strictly negative derivative in Λ (by assumption), by the implicit theorem we can also conclude that W_i has a strictly negative derivative in Λ. Then we can also redefine Λ as an implicit differentiable function $\Lambda(q)$ of the vector of quantities such that the budget constraint holds, i.e. such that: $\sum_i W_i(q_i, \Lambda) = 1$. As an abuse of notation, Λ denotes the aggregator both as a function of normalized prices and as a function of quantities q given that they coincide when q is the demand associated with normalizes prices p/w.[39]

In the remainder of the proof, since we focus on inverse demand, Λ primarily refers to such a function of quantities q rather than normalized prices p/w.

Proof of Proposition 1

As described just above, the proof of Proposition 1 relies on the inverse demand function (using expenditures shares $W_i(q, \Lambda)$ as functions of quantities and the aggregator Λ) rather than direct demand, and Λ is defined as a function of quantities q), where $W_i(q_i, \Lambda)$ is twice differentiable with a negative derivative in Λ and non-zero derivative in q_i

Differentiating the budget constraint $\sum_i W_i(q_i, \Lambda) = 1$ w.r.t. q_i implies:

$$\varepsilon_j(q_j, \Lambda) \;=\; \frac{S(q)}{W_j}\, \frac{\partial \Lambda}{\partial \log q_j} \tag{54}$$

[38] Alternatively, Proposition 1 applies to the image $\mathcal{Q} = \left\{q\ ;\ q_i = \widetilde{q}_i(p_i/w, \Lambda(p/w)) > 0, p/w \in \mathbb{R}_+^N\right\} \subset \mathbb{R}_+^N$.

[39] As a side note, we can also show that iso-Λ curves are connected, which implies that if any differentiable function of q that have a gradient that is proportional to the gradient of Λ (w.r.t. q) can be expressed as a function of Λ.

where $\varepsilon_j(q_j, \Lambda) \equiv \left.\frac{\partial \log W_j}{\partial \log q_j}\right|_{\Lambda}$ denotes the elasticity w.r.t. own quantity q_j, holding aggregators constant, and where $S(q) \equiv \sum_i \frac{\partial W_i}{\partial \Lambda}(q, \Lambda(q))$ is strictly negative.

For such a demand system to be integrable and satisfy Slutsky symmetry, there must exist a utility function $U(q)$ and another real function λ such that $\lambda(q) > 0$ and:

$$\frac{\partial U}{\partial \log q_i} = \lambda(q) W_i(q_i, \Lambda(q))$$

for any q. As mentioned in the text, we further assume that U is twice continuously differentiable. Differentiating again, we obtain:

$$\frac{\partial U}{\partial \log q_i \partial \log q_j} = \frac{\partial \lambda}{\partial \log q_j} W_i + \lambda \frac{\partial W_i}{\partial \Lambda} \frac{\partial \Lambda}{\partial \log q_j}.$$

The existence and continuity of the derivatives imply that the cross derivative is symmetric, hence:

$$\left(\frac{1}{W_j} \frac{\partial \log \lambda}{\partial \log q_j}\right) + \frac{\partial \log W_i}{\partial \Lambda} \left(\frac{1}{W_j} \frac{\partial \Lambda}{\partial \log q_j}\right) = \left(\frac{1}{W_i} \frac{\partial \log \lambda}{\partial \log q_i}\right) + \frac{\partial \log W_j}{\partial \Lambda} \left(\frac{1}{W_i} \frac{\partial \Lambda}{\partial \log q_i}\right)$$

Incorporating the expression from 54, this is equivalent to:

$$\left(\frac{S}{W_j} \frac{\partial \log \lambda}{\partial \log q_j}\right) \; + \; \frac{\partial \log W_i}{\partial \Lambda} \; \varepsilon_j \; = \; \left(\frac{S}{W_i} \frac{\partial \log \lambda}{\partial \log q_i}\right) \; + \; \frac{\partial \log W_j}{\partial \Lambda} \; \varepsilon_i.$$

holds for any $i \neq j$. Define $A_i(q) = \frac{S(q)}{W_j} \frac{\partial \log \lambda}{\partial \log q_j}(q)$, we obtain a key symmetry requirement that we will exploit below:

$$A_j(q) \; + \; \frac{\partial \log W_i}{\partial \Lambda}(q_i, \Lambda) \; \varepsilon_j(q_j, \Lambda) \; = \; A_i(q) \; + \; \frac{\partial \log W_j}{\partial \Lambda}(q_j, \Lambda) \; \varepsilon_i(q_i, \Lambda). \tag{55}$$

Next, we we can see that we will be in either of these three cases (almost everywhere) in a neighborhood of any q:

- $\mathcal{Q}_1$ is the set of vectors of quantities q such that $\varepsilon_i(q_i, \Lambda)$ takes at least two different values across goods i *even if we exclude any one good.*
- $\mathcal{Q}_2$ is the set of vectors of quantities q such that $\varepsilon_i(q_i, \Lambda)$ are identical across goods i.
- $\mathcal{Q}_3$ is the set of vectors of quantities q such that all $\varepsilon_i(q_i, \Lambda)$ are identical for all but one good.

For a neighborhood around q, suppose that $\left.\frac{\partial \varepsilon_i}{\partial q_i}(q_i, \Lambda)\right|_{\Lambda} \neq 0$ and $\left.\frac{\partial \varepsilon_i}{\partial q_j}(q_j, \Lambda)\right|_{\Lambda} \neq 0$ for at least two goods i and j. In that case, we can see that ε_i, ε_j and ε_k will differ almost everywhere in a neighborhood of q for i, j and any third good k; hence we are in $\mathcal{Q}_1$ almost everywhere around q. This is the first case considered below.

Next, suppose that $\left.\frac{\partial \varepsilon_{i0}}{\partial q_{i0}}(q_{i0}, \Lambda)\right|_{\Lambda} \neq 0$ for just one good $i0$, i.e. $\varepsilon_j(q_j, \Lambda)$ does not depend on q_j for goods other than $i0$ in the neighborhood of q. If $\varepsilon_j(q_j, \Lambda)$ takes two different values across goods j, we are in case 1. If $\varepsilon_j(q_j, \Lambda)$ is identical across all goods $j \neq i0$, we are in case 3 below.

Finally, if $\frac{\partial \varepsilon_{i0}}{\partial q_{i0}}(q_{i0}, \Lambda)\Big|_{\Lambda} = 0$ in a neighborhood of q, we are either in case 1 below (if ε_i takes on at least two values even if we exclude a single good), in case 2 (if ε_i is identical across all goods), or in case 3 (if ε_i is identical across all but one good).

Case 1 In an open set of q, suppose that $\varepsilon_i(q_i, \Lambda)$ takes at least two different values across goods i, even if we exclude any one good.

In this case, even if we exclude a single good j, there exists a vector $x_i(q)$ such that $\sum_i x_i = 0$ and $\sum_i \varepsilon_i x_i \neq 0$. Multiplying Equation (55) by $x_i(q)$ and summing up across goods i (for a given j), we obtain:

$$\left(\sum_i x_i \frac{\partial \log W_i}{\partial \Lambda}\right) \varepsilon_j \;=\; \left(\sum_i x_i A_i\right) \;+\; \left(\sum_i x_i \varepsilon_i\right) \frac{\partial \log W_j}{\partial \Lambda}.$$

As $\sum_i \varepsilon_i x_i \neq 0$, we obtain that there exists two functions $h(q)$ and $m(q)$ such that:

$$\frac{\partial \log W_j}{\partial \Lambda} \;=\; h(q)\, \varepsilon_j(q_j, \Lambda) \;+\; m(q).$$

In particular, this holds also for any pair of goods i and j. Taking the difference, we get:

$$\frac{\partial \log W_j}{\partial \Lambda} - \frac{\partial \log W_i}{\partial \Lambda} \;=\; h(q) \left(\varepsilon_j(q_j, \Lambda) - \varepsilon_i(q_i, \Lambda)\right).$$

In particular, take two goods for which $\varepsilon_i \neq \varepsilon_j$. Note that the left-hand side only depends on q_j, q_i and Λ. This implies that $h(q)$ can be written as a function of q_j, q_i and Λ only.

If we're not in case 3, we can also find a third good i' such that $\varepsilon_{i'} \neq \varepsilon_i$ and $\varepsilon_{i'} \neq \varepsilon_j$. Applying the same argument, it must be that h can be written as just a function of Λ, so we now denote h as: $h = h(\Lambda)$.

Taking again a derivative in $\log q_j$, holding Λ constant, and noticing that the cross derivative is symmetric, $\frac{\partial \varepsilon_j}{\partial \Lambda} = \frac{\partial \log W_j}{\partial \log q_j \partial \Lambda} = \frac{\partial \log W_j}{\partial \Lambda \partial \log q_j}$, we obtain:

$$\frac{\partial \varepsilon_j}{\partial \Lambda} \;=\; h(\Lambda)\, \frac{\partial \varepsilon_j}{\partial \log q_j} \;=\; \frac{\partial \log H}{\partial \Lambda}\, \frac{\partial \varepsilon_j}{\partial \log q_j} \tag{56}$$

where we define $\log H$ as the integral of h:

$$H(\Lambda) = \exp\left(\int_{\Lambda^*}^{\Lambda} h(t)dt\right)$$

taking any fixed reference point Λ^*. We would have then $H(\Lambda^*) = 1$ by definition (it's also important to notice that H does not depend on j and q_j).

Using this, let's show that differential equation (66) implies:

$$\varepsilon_j(q_j, \Lambda) = \varepsilon_j(q_j H(\Lambda), \Lambda^*) \tag{57}$$

To show this result, consider the function

$$e_j(x) = \varepsilon_j(q_j H(\Lambda)/H(x), x).$$

Taking all other variables Λ and q_j as fixed, only varying x between Λ^* and Λ. We find that the derivative of $e_j(x)$ w.r.t. x is zero:

$$e'_j(x) \;=\; \frac{\partial \varepsilon_j}{\partial \Lambda}\,(q_j H(\Lambda)/H(x), x) \;\;-\;\; \frac{\partial \log H}{\partial \Lambda}(x)\;\frac{\partial \varepsilon_j}{\partial \log q_j}(q_j H(\Lambda)/H(x), x) \;=\; 0.$$

Hence e_j does not depend on x. Moreover, $e_j(\Lambda)$ corresponds to: $e_j(\Lambda) = \varepsilon_j(q_j, \Lambda)$, while $e_j(\Lambda^*)$ is such that:

$$e_j(\Lambda^*) \;=\; \varepsilon_j(q_j H(\Lambda)/H(\Lambda^*), \Lambda^*) \;=\; \varepsilon_j(q_j H(\Lambda), \Lambda^*)$$

given that $H(\Lambda^*) = 1$ by definition of H. Hence we get the equality between the last two expressions: $\varepsilon_j(q_j, \Lambda) = \varepsilon_j(q_j H(\Lambda), \Lambda^*)$, which holds for any q_j. Hence we have proven equation (57).

Integrating over q_j from a reference point q_j^* in the region where equality (57) holds, we obtain that demand can be written as:

$$\begin{aligned}
\frac{W_j(q_j,\Lambda)}{W_j(q_j^*,\Lambda)} &= \exp\left[\int_{q_j^*}^{q_j} \varepsilon_j(q,\Lambda)\frac{dq}{q}\right] \\
&= \exp\left[\int_{q_j^*}^{q_j} \varepsilon_j(qH(\Lambda),\Lambda^*)\frac{dq}{q}\right] \\
&= \exp\left[\int_{q_j^* H(\Lambda)}^{q_j H(\Lambda)} \varepsilon_j(q,\Lambda^*)\frac{dq}{q}\right] \\
&= \frac{W_j(q_j H(\Lambda),\Lambda^*)}{W_j(q_j^* H(\Lambda),\Lambda^*)}.
\end{aligned}$$

It shows that the effect of q_j on W_j is independent of Λ, provided that we adjust for the shifter $H(\Lambda)$.

Next, take a fixed reference q_j^* as given and define F_j as:

$$F_j(\Lambda) \;\equiv\; \frac{W_j(q_j^* H(\Lambda),\Lambda^*)}{W_j(q_j^*,\Lambda)}.$$

Taking any two goods i and j, we obtain:

$$\begin{aligned}
\frac{\log(F_j/F_i)}{\partial \Lambda} &= h(\Lambda)\;\big(\varepsilon_j(q_j^* H(\Lambda),\Lambda^*) - \varepsilon_i(q_i^* H(\Lambda),\Lambda^*)\big) \;-\; \frac{\partial \log W_j}{\partial \Lambda}(q_j^*,\Lambda) + \frac{\partial \log W_i}{\partial \Lambda}(q_i^*,\Lambda) \\
&= h(\Lambda)\;\big(\varepsilon_j(q_j^*,\Lambda) - \varepsilon_i(q_i^*,\Lambda)\big) \qquad - \frac{\partial \log W_j}{\partial \Lambda}(q_j^*,\Lambda) + \frac{\partial \log W_i}{\partial \Lambda}(q_i^*,\Lambda) \\
&= 0.
\end{aligned}$$

Since $F_j(\Lambda^*) = 1$ for all goods j, this implies that these functions $F_j = F_i = F(\Lambda)$ is identical across all goods.

Starting with Equation (68) and combining with the properties of F above, we finally obtain:

$$\begin{aligned}
W_j(q_j,\Lambda) &= \frac{W_j(q_j^*,\Lambda)}{W_j(q_j^* H(\Lambda),\Lambda^*)}\; W_j(q_j H(\Lambda),\Lambda^*) \\
&= \frac{1}{F(\Lambda)}\; W_j(q_j H(\Lambda),\Lambda^*)
\end{aligned}$$

Dividing by q_i, this implies that normalized price must equal:

$$\frac{p_i}{w} = \frac{1}{q_i F(\Lambda)} \ W_j(q_j H(\Lambda), \Lambda^*).$$

As we assume that demand is strictly monotonic in prices, holding Λ constant, it can be inverted such that we can express q_i as a function of p_i/w and Λ. Denoting D_i the inverse of $\frac{1}{q_i} \ W_j(q_j, \Lambda^*)$ (holding Λ^* fixed), we obtain:

$$q_i = \frac{1}{H(\Lambda)} \ D_j(F(\Lambda)p_j/w). \tag{58}$$

Case 2 is the simplest. Suppose that ε_i is the same across i's. Since each $\varepsilon_i(q_i, \Lambda)$ depends only on q_i and Λ, it must be that these elasticities only depend on Λ, i.e.:

$$\varepsilon_j(q_j, \Lambda) \ = \ 1 - 1/\sigma(\Lambda)$$

for some function $\sigma(\Lambda) \neq 1$.

Integrating, this implies that demand can be written as:

$$W_j(q_j, \Lambda) \ = \ A_j(\Lambda)^{-\frac{1}{\sigma(\Lambda)}} q_j^{1-\frac{1}{\sigma(\Lambda)}} \tag{59}$$

for some good-specific functions $A_j(\Lambda)$. This leads the demand function in the text:

$$\widetilde{q}_i(p_i/w, \Lambda) = A_i(\Lambda)(p_i/w)^{-\sigma(\Lambda)}.$$

Case 3 Suppose that ε_i is the same across i's except for a single good $i0$. Again, since each $\varepsilon_i(q_i, \Lambda)$ depends only on q_i and Λ (except good $i0$), it must be that these elasticities only depend on Λ, i.e.:

$$\varepsilon_i(q_i, \Lambda) \ = \ \bar{\varepsilon}(\Lambda)$$

for each good $i \neq i0$, for some function $\bar{\varepsilon}(\Lambda) \neq 0$. In that case, Equation (55) can be rewritten:

$$A_j \ + \ \frac{\partial \log W_{i0}}{\partial \Lambda} \ \bar{\varepsilon}(\Lambda) \ = \ A_{i0} \ + \ \frac{\partial \log W_j}{\partial \Lambda} \ \varepsilon_{i0}$$

for any good $j \neq i0$. For $i \neq j$ and $i \neq i0$, we have:

$$A_j \ + \ \frac{\partial \log W_i}{\partial \Lambda} \ \bar{\varepsilon}(\Lambda) \ = \ A_i \ + \ \frac{\partial \log W_j}{\partial \Lambda} \ \bar{\varepsilon}(\Lambda).$$

Taking the difference, we obtain for any two goods $j, i \neq i0$:

$$\left(\frac{\partial \log W_{i0}}{\partial \Lambda} - \frac{\partial \log W_i}{\partial \Lambda} \right) \bar{\varepsilon}(\Lambda) \ = \ (A_{i0} - A_i) \ + \frac{\partial \log W_j}{\partial \Lambda} \ (\varepsilon_{i0} - \bar{\varepsilon}(\Lambda)).$$

Taking again the difference with the same expression with a fourth good k instead of j, we obtain:

$$0 \ = \ \left(\frac{\partial \log W_j}{\partial \Lambda}(q_j, \Lambda) - \frac{\partial \log W_k}{\partial \Lambda}(q_k, \Lambda) \right) \ (\varepsilon_{i0}(q_{i0}, \Lambda) - \bar{\varepsilon}(\Lambda)).$$

Since $\varepsilon_{i0}(q_{i0}, \Lambda) \neq \bar{\varepsilon}(\Lambda)$, it implies that $\frac{\partial \log W_j}{\partial \Lambda}(q_j, \Lambda) = \frac{\partial \log W_k}{\partial \Lambda}(q_k, \Lambda)$, which must hold for any pair of good k and j except $i0$. This implies that there exist some functions $A(\Lambda)$ and $\widetilde{W}_j(q_j)$ such that $W_j(q_j, \Lambda) = \widetilde{W}_j(q_j)A(\Lambda)$ for all $j \neq i0$. Since $\frac{\partial \log W_j}{\partial \log q_j} = -\bar{\varepsilon}(\Lambda)$, we can also conclude that $\bar{\varepsilon}(\Lambda)$ is constant and does not depend on Λ. Thus, denoting $\bar{\varepsilon} = 1 - 1/\sigma$, we obtain:

$$W_j(q_j, \Lambda) \;=\; w_j q_i^{1-\frac{1}{\sigma}} A(\Lambda). \tag{60}$$

for some constant terms w_j for each $j \neq i0$. We obtain the functional form in the text by inverting and expressing q_i as a function of Λ and p_i/w.

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 case 1 with cases 2 and 3, hence the functional form of case 1 needs to hold globally across all Λ's.

Combination of cases 1+2 Here we show that we cannot have a combination of cases 1 and 2 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 occur along different Λ's. By contradiction, suppose that there exists Λ^* such that, at least locally,

$$\begin{aligned} W_i(q_i, \Lambda) &= W_j(q_j H(\Lambda), \Lambda^*)/\, F(\Lambda) && \text{if } \Lambda < \Lambda^* \\ W_i(q_i, \Lambda) &= A_i(\Lambda)^{-\frac{1}{\sigma(\Lambda)}} q_i^{1-\frac{1}{\sigma(\Lambda)}} && \text{if } \Lambda > \Lambda^* \end{aligned}$$

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

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

Since it must hold for any i and any y, it implies that $\frac{\partial \log W_i}{\partial \log y} = 0$, 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+3 Here we show that we cannot have a combination of cases 1 and 3 globally, using the same arguments as above. Note again that for a given Λ, case 1 and 3 are mutually exclusive by definition. Hence, if we have a mixture of cases 1 and 3, it must occur 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:

$$\begin{aligned} W_i(q_i, \Lambda) &= W_j(q_j H(\Lambda), \Lambda^*)/\, F(\Lambda) && \text{if } \Lambda < \Lambda^* \\ W_i(q_i, \Lambda) &= w_j H(\Lambda) q_i^{1-\frac{1}{\sigma}} && \text{if } \Lambda > \Lambda^*. \end{aligned}$$

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

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

Again, since it must hold for any i and any y, it implies that $\frac{\partial \log W_i}{\partial \log y} = 0$, which contradicts our assumption that $W_i(y_i, \Lambda)$ is not locally constant across y_i for any given Λ.

Proposition 2

Preliminaries. As for Proposition 1, it is easier to prove Proposition 2 by examining the inverse demand, i.e. normalized prices as a function of quantities q (here with two aggregators Λ and U).

Consider the demand system:

$$q_i = \widetilde{q}_i(p_i/w, \Lambda, V)$$

where $V = V(p/w)$ is indirect utility and Λ is an implicit function of normalized prices p_i/w such that the budget constraint holds, i.e. such that:

$$\sum_i (p_i/w)\, \widetilde{q}_i(p_i/w, \Lambda, V(p/w)) = 1.$$

If we assume that each $\widetilde{q}_i(p_i/w, \Lambda, V)$ is monotonically decreasing in Λ (here we assume a strictly negative derivative), the solution in Λ is unique.

Since we also assume that expenditure shares $(p_i/w)\, \widetilde{q}_i(p_i/w, \Lambda, V)$ monotonically decreases or increases with prices (holding Λ and V constant), and since we assume that for each q_i there exist a vector of normalized prices such that $\widetilde{q}_i(p_i/w, \Lambda(p/w), V(p/w))$, such demand can be inverted such that there exist functions W_i such that:

$$q_i p_i/w = W_i(q_i, \Lambda, U)$$

i.e. such that $(p_i/w)\widetilde{q}_i(p_i/w, \Lambda(p/w), V(p/w)) = W_i(\widetilde{q}_i(p_i/w, \Lambda(p/w), V(p/w)), \Lambda(p/w), V(p/w))$ for any p/w. By definition, note also that direct and indirect utility are equal, $V(p/w) = U(q)$, when demand q is evaluated at normalized prices p/w.

As demand $\widetilde{q}_i(p_i/w, \Lambda, V)$ has a strictly negative derivative in Λ (holding q_i and V constant), by the implicit theorem we can also conclude that W_i has a strictly negative derivative in Λ. As in the single-aggregator case (Proposition 1), we can thus redefine Λ as an implicit function of the vector of quantities such that the budget constraint holds, i.e. such that: $\sum_i W_i(q_i, \Lambda, U(q)) = 1$ when V coincides with utility $U(q)$.

Again, in the remainder of the proof of Proposition 2, Λ refers to a function of quantities q rather than normalized prices.

Proof of Proposition 2

For such a demand system to be integrable (and satisfy Slutsky symmetry), there must exist a utility function $U(q)$ and another scale function such that:

$$\frac{\partial U}{\partial \log q_i} = \lambda(q) W_i(q_i, \Lambda(q), U(q)). \tag{61}$$

We further assume that such utility function is twice continuously differentiable. Differentiating the budget constraint $\sum_i W_i(q_i, \Lambda(q), U(q)) = 1$ implies:

$$\left.\frac{\partial \log W_j}{\partial \log q_j}\right|_{\Lambda} = -\sum_i \frac{\partial W_i}{\partial \Lambda} \frac{\partial \Lambda}{\partial \log q_j} - \sum_i \frac{\partial W_i}{\partial U} \frac{\partial U}{\partial \log q_j}. \tag{62}$$

Using $\frac{\partial U}{\partial \log q_j} = \lambda W_j$, we obtain:

$$\frac{S_\Lambda(q)}{W_j}\,\frac{\partial \Lambda}{\partial \log q_j} \;=\; \varepsilon_j(q_j, \Lambda) \;-\; S_U(q)\lambda(q) \tag{63}$$

where $\varepsilon_j(q_j, \Lambda) \equiv \left.\frac{\partial \log W_j}{\partial \log q_j}\right|_\Lambda$ denotes the elasticity w.r.t. own quantity q_j, holding aggregators constant, where $S_\Lambda(q) \equiv -\sum_i \frac{\partial W_i}{\partial \Lambda}$ is different from zero by assumption, and where $S_U(q) \equiv -\sum_i \frac{\partial W_i}{\partial U}$.

Next, differentiating equation (61), we obtain:

$$\frac{\partial U}{\partial \log q_i \partial \log q_j} \;=\; \frac{\partial \lambda}{\partial \log q_j} W_i \;+\; \lambda \frac{\partial W_i}{\partial \Lambda}\frac{\partial \Lambda}{\partial \log q_j} \;+\; \lambda^2 \frac{\partial W_i}{\partial U} W_j.$$

The cross derivative is symmetric as we assume that U is twice continuously differentiable. Hence, dividing by $\lambda W_i W_j$ we obtain:

$$\left(\frac{1}{W_j}\frac{\partial \log \lambda}{\partial \log q_j}\right) \;+\; \frac{\partial \log W_i}{\partial \Lambda}\left(\frac{1}{W_j}\frac{\partial \Lambda}{\partial \log q_j}\right) \;+\; \lambda \frac{\partial \log W_i}{\partial U} \;=$$
$$\left(\frac{1}{W_i}\frac{\partial \log \lambda}{\partial \log q_i}\right) \;+\; \frac{\partial \log W_j}{\partial \Lambda}\left(\frac{1}{W_i}\frac{\partial \Lambda}{\partial \log q_i}\right) \;+\; \lambda \frac{\partial \log W_j}{\partial U}.$$

Incorporating the expression from (63), this is equivalent to:

$$\left(\frac{S_\Lambda}{W_j}\frac{\partial \log \lambda}{\partial \log q_j}\right) \;+\; \frac{\partial \log W_i}{\partial \Lambda}\left(\varepsilon_j \;-\; S_U \lambda\right) \;+\; \lambda \frac{\partial \log W_i}{\partial U} \;=$$
$$\left(\frac{S_\Lambda}{W_i}\frac{\partial \log \lambda}{\partial \log q_i}\right) \;+\; \frac{\partial \log W_j}{\partial \Lambda}\left(\varepsilon_i \;-\; S_U \lambda\right) \;+\; \lambda \frac{\partial \log W_j}{\partial U}. \tag{64}$$

Define $A_i(q) \;=\; \frac{S_\Lambda}{W_i}\frac{\partial \log \lambda}{\partial \log q_i} + \frac{\partial \log W_i}{\partial \Lambda} - \lambda \frac{\partial \log W_i}{\partial U}$ we get an expression that is very similar to the single-aggregator case:

$$A_j(q) \;+\; \frac{\partial \log W_i}{\partial \Lambda}(q_i, \Lambda, U)\; \varepsilon_j(q_j, \Lambda, U) \;=\; A_i(q) \;+\; \frac{\partial \log W_j}{\partial \Lambda}(q_j, \Lambda, U)\; \varepsilon_i(q_i, \Lambda, U) \tag{65}$$

and holds for any $i \neq j$.

Unlike the previous Proposition, here we directly assume that $\varepsilon_i(q_i, \Lambda, U)$ takes at least two different values across goods i, almost everywhere, even if we exclude any one good.

In this case, even if we exclude a single good j, there exists a vector $x_i(q)$ such that $\sum_i x_i = 0$ and $\sum_i \varepsilon_i x_i \neq 0$. Multiplying Equation (65) by $x_i(q)$ and summing up across goods i (for a given j), we obtain:

$$\left(\sum_i x_i \frac{\partial \log W_i}{\partial \Lambda}\right) \varepsilon_j \;=\; \left(\sum_i x_i A_i\right) \;+\; \left(\sum_i x_i \varepsilon_i\right)\; \frac{\partial \log W_j}{\partial \Lambda}.$$

As $\sum_i \varepsilon_i x_i \neq 0$, we obtain that there exists two functions $h(q)$ and $m(q)$ such that:

$$\frac{\partial \log W_j}{\partial \Lambda}(q_j, \Lambda, U) \;=\; h(q)\; \varepsilon_j(q_j, \Lambda, U) \;+\; m(q).$$

In particular, this holds also for any pair of goods i and j. Taking the difference, we get:

$$\frac{\partial \log W_j}{\partial \Lambda}(q_j, \Lambda, U) - \frac{\partial \log W_i}{\partial \Lambda}(q_i, \Lambda, U) \;=\; h(q)\, \Big(\varepsilon_j(q_j, \Lambda, U) - \varepsilon_i(q_i, \Lambda, U)\Big)$$

Take two goods for which $\varepsilon_i \neq \varepsilon_j$. Note that the left-hand side only depends on q_j, q_i and Λ. This implies that $h(q)$ can be written as a function of q_j, q_i, Λ and U only.

We can also find a third good i' such that $\varepsilon_{i'} \neq \varepsilon_i$ and $\varepsilon_{i'} \neq \varepsilon_j$. Applying the same argument, it must be that h can be written as just a function of Λ and U, so we now denote h as: $h = h(\Lambda, U)$.

Taking again a derivative in $\log q_j$, holding Λ and U constant, and noticing that the cross derivative is symmetric, $\frac{\partial \varepsilon_j}{\partial \Lambda} = \frac{\partial \log W_j}{\partial \log q_j \partial \Lambda} = \frac{\partial \log W_j}{\partial \Lambda \partial \log q_j}$, we obtain:

$$\frac{\partial \varepsilon_j}{\partial \Lambda} \;=\; h(\Lambda, U)\, \frac{\partial \varepsilon_j}{\partial \log q_j} \;=\; \frac{\partial \log H}{\partial \Lambda}\, \frac{\partial \varepsilon_j}{\partial \log q_j} \tag{66}$$

where we define $\log H$ as the integral of h, for a given U:

$$H(\Lambda, U) = \exp\left(\int_{\Lambda^*}^{\Lambda} h(t, U) dt\right)$$

taking any fixed reference point Λ^*. We would have then $H(\Lambda^*, U) = 1$ by definition (it's also important to notice that H does not depend on j and q_j).

Using this, let's show that differential equation (66) implies:

$$\varepsilon_j(q_j, \Lambda, U) = \varepsilon_j(q_j H(\Lambda, U), \Lambda^*, U) \tag{67}$$

To show this result, consider the function

$$e_j(x) = \varepsilon_j(q_j H(\Lambda, U)/H(x, U), x, U)$$

Taking all other variables Λ, U and q_j as fixed, only varying x between Λ^* and Λ. We find that the derivative of $e_j(x)$ w.r.t. x is zero:

$$e_j'(x) \;=\; \frac{\partial \varepsilon_j}{\partial \Lambda}\,(q_j H(\Lambda, U)/H(x, U), x, U) \;\;-\;\; \frac{\partial \log H}{\partial \Lambda}(x, U)\; \frac{\partial \varepsilon_j}{\partial \log q_j}(q_j H(\Lambda, U)/H(x, U), x, U) \;=\; 0$$

Hence e_j does not depend on x. Moreover, $e_j(\Lambda)$ corresponds to: $e_j(\Lambda) = \varepsilon_j(q_j, \Lambda, U)$, while $e_j(\Lambda^*)$ is such that:

$$e_j(\Lambda^*) \;=\; \varepsilon_j(q_j H(\Lambda, U)/H(\Lambda^*, U), \Lambda^*, U) \;=\; \varepsilon_j(q_j H(\Lambda, U), \Lambda^*, U)$$

given that $H(\Lambda^*, U) = 1$ by definition of H. Hence we get the equality between the last two expressions: $\varepsilon_j(q_j, \Lambda, U) = \varepsilon_j(q_j H(\Lambda, U), \Lambda^*, U)$, which holds for any q_j. Thus we have proven equation (67).

Integrating over q_j from a reference point q_j^* in the region where equality (67) holds, we obtain that demand can be written as:

$$\begin{aligned} \frac{W_j(q_j, \Lambda, U)}{W_j(q_j^*, \Lambda, U)} \;&=\; \exp\left[\int_{q_j^*}^{q_j} \varepsilon_j(q, \Lambda, U) \frac{dq}{q}\right] \\ &=\; \exp\left[\int_{q_j^*}^{q_j} \varepsilon_j(q H(\Lambda, U), \Lambda^*, U) \frac{dq}{q}\right] \end{aligned}$$

$$
\begin{aligned}
&= \exp\left[\int_{q_j^* H(\Lambda,U)}^{q_j H(\Lambda,U)} \varepsilon_j(q,\Lambda^*,U)\frac{dq}{q}\right] \\
&= \frac{W_j(q_j H(\Lambda,U),\Lambda^*,U)}{W_j(q_j^* H(\Lambda,U),\Lambda^*,U)}.
\end{aligned}
$$

It shows that the effect of q_j on W_j is independent of Λ and U, provided that we adjust for the shifter $H(\Lambda, U)$.

Next, take a fixed reference q_j^* as given and define F_j as:

$$
F_j(\Lambda,U) \equiv \frac{W_j(q_j^* H(\Lambda,U),\Lambda^*,U)}{W_j(q_j^*,\Lambda,U)}
$$

Taking any two goods i and j, we obtain:

$$
\begin{aligned}
&\frac{\log(F_j/F_i)}{\partial\Lambda} \\
&= h(\Lambda,U)\ \big(\varepsilon_j(q_j^* H(\Lambda,U),\Lambda^*,U) - \varepsilon_i(q_i^* H(\Lambda,U),\Lambda^*,U)\big)\ - \frac{\partial \log W_j}{\partial\Lambda}(q_j^*,\Lambda,U) + \frac{\partial \log W_i}{\partial\Lambda}(q_i^*,\Lambda,U) \\
&= h(\Lambda,U)\ \big(\varepsilon_j(q_j^*,\Lambda,U) - \varepsilon_i(q_i^*,\Lambda,U)\big) \qquad - \frac{\partial \log W_j}{\partial\Lambda}(q_j^*,\Lambda,U) + \frac{\partial \log W_i}{\partial\Lambda}(q_i^*,\Lambda,U) \\
&= 0.
\end{aligned}
$$

Since $F_j(\Lambda^*, U) = 1$ for all goods j, this implies that these functions $F_j = F_i = F(\Lambda, U)$ are identical across all goods.

Starting with Equation (68) and combining with the properties of F above, we finally obtain:

$$
\begin{aligned}
W_j(q_j,\Lambda,U) &= \frac{W_j(q_j^*,\Lambda,U)}{W_j(q_j^* H(\Lambda,U),\Lambda^*,U)}\ W_j(q_j H(\Lambda,U),\Lambda^*,U) \\
&= \frac{1}{F(\Lambda,U)}\ W_j(q_j H(\Lambda,U),\Lambda^*,U).
\end{aligned}
$$

Dividing by q_i, this implies that normalized price must equal:

$$
\frac{p_i}{w} = \frac{1}{q_i F(\Lambda,U)}\ W_j(q_j H(\Lambda,U),\Lambda^*,U).
$$

As we assume that demand is strictly monotonic in prices, holding Λ and U constant, it can be inverted such that we can express q_i as a function of p_i/w and Λ. Denoting D_i the inverse of $\frac{1}{q_i}\ W_j(q_j,\Lambda^*,U)$ w.r.t. q_i (holding U constant and holding Λ^* fixed), we obtain the expression in Proposition 2:

$$
q_i = \frac{1}{H(\Lambda,V)}\ D_j(F(\Lambda,V)p_j/w,V). \tag{68}
$$

Proof of Proposition 3

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

$$\widetilde{U}(q,\Lambda) = \sum_i u_i(H(\Lambda)q_i) \; - \; \int_{\Lambda_0}^{\Lambda} F(\Lambda)H'(\Lambda)d\Lambda$$

where:[40]

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

and $u_i' = D_i^{-1}$. Next, Λ can be defined as an implicit function of q such that:

$$\sum_i q_i u_i'(H(\Lambda)q_i)/F(\Lambda) \; = \; 1 \tag{69}$$

As in Propositions 1 and 2, the remainder of the proof refers to Λ as a function of q rather than normalized prices p/w.

We proceed in three steps. First we show that equation (69) admits a solution $\Lambda(q)$ for each q and that this solution is unique. Second we show that utility defined as $U(q) = \widetilde{U}(q, \Lambda(q))$ 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 and coincides with Λ for optimal consumption baskets.

Step 1: Implicit function $\Lambda(q)$. Here we show that for any vector q of consumption, there is a unique Λ such that equation (69) holds.

First, using part ii) of restrictions [A3], we can see that the elasticity of $D_i(F(\Lambda)y_i)/H(\Lambda)$ w.r.t. Λ is given by $\varepsilon_F\varepsilon_{Di} - \varepsilon_H$ which is assumed to be negative, hence it strictly decreases with Λ. Symmetrically, we obtain that $u_i'(H(\Lambda)q_i)/F(\Lambda)$ also strictly decreases with Λ: its elasticity w.r.t. Λ is $\varepsilon_H/\varepsilon_{Di} - \varepsilon_F$, which is also negative given that ε_{Di} is negative and $\varepsilon_H - \varepsilon_{Di}\varepsilon_F$ is positive. Adding up across goods, we obtain that the left-hand side of equation (69) decreases strictly with Λ. This implies that the solution to equation (69) is unique (if it exists).

Existence is then guaranteed using condition [A3]-iii), which we can symmetrically reformulate in terms of quantities. We assume that, for any good i and $y_i > 0$, there exists $\Lambda \in \mathbb{R}$ such that: $y_i D_i\left(y_i F(\Lambda)\right)/H(\Lambda) = 1/N$. Using $u_i' = D_i^{-1}$, note that this equality is equivalent to $1/N = 1/(Ny_i)u_i'\left(1/(Ny_i)H(\Lambda)\right)/F(\Lambda)$. Hence, denoting $q_i = 1/Ny_i$, we obtain that for any good i and $q_i > 0$, there exists $\Lambda \in \mathbb{R}$ such that:

$$q_i u_i'\left(q_i H(\Lambda)\right)/F(\Lambda) = 1/N.$$

For a given vector of quantities q, for each good we obtain a Λ such that the equality above holds. Taking the maximum of the Λ's obtained across the N goods (and using the monotonicity property in Λ described just above), we obtain a Λ^{max} such that

$$\sum_i q_i u_i'\left(q_i H(\Lambda^{max})\right)/F(\Lambda^{max}) \leq 1.$$

[40] 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 $D_i(0) = b_i$.

For the same vector q, by taking the minimum of such Λ's across goods, we obtain a Λ^{min} such that

$$\sum_i q_i u_i' \left(q_i H(\Lambda^{min}) \right) / F(\Lambda^{min}) \geq 1.$$

As the left hand side of this expression is continuous in Λ, the intermediate-value theorem ensures that a solution to equation (69) exists between Λ^{min} and Λ^{max}.

Finally, note that the derivative of the left-hand side of equation (69) is strictly negative. Using the implicit function theorem, we can thus obtain the derivatives of Λ w.r.t. q as described below.

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

Derivatives of the aggregator Λ. Here we consider the properties of $\Lambda(q)$, the solution of equation (69). Taking the derivative of equation (69), we get:

$$\sum_i q_i u_i'(H(\Lambda) q_i)/F(\Lambda) \; = \; 1$$

$$\frac{\partial \Lambda}{\partial q_i} \left[H' \sum_j q_j^2 u_j'' - F' \right] \; + \; [u_i' + H q_i u_i''] \; = \; 0$$

and thus:

$$\frac{\partial \Lambda}{\partial q_i} = \frac{u_i' + H q_i u_i''}{\Delta(q)}$$

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

We can verify that $\Delta(q)$ is positive. Note that $\frac{u_i'}{u_i'' H q_i} = \varepsilon_{Di}$, the elasticity of function D_i. Thus, we obtain:

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

Recall that $u_i' > 0$ and that assumption [A3]-ii) imposes: $\varepsilon_F \, \varepsilon_{Di} < \varepsilon_H$ for all i. Since we also assume downward slopping demand, $\varepsilon_{Di} < 0$, this implies $\varepsilon_F > \varepsilon_H / \varepsilon_{Di}$ for all i and therefore $\Delta > 0$. This implies that the derivatives of Λ are always well defined. Also, knowing that Δ is positive will be useful again below.

Derivatives of utility U. The first derivatives are:

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

where the term in brackets is null for any q, thanks to condition (69). Second derivatives are then:

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

and thus, incorporating the derivatives in Λ, we obtain:

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

Negative semi-definiteness. To show that utility is quasi-concave, 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 vector $t \in \mathbb{R}^N$ 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 thus normalize the sum $\sum_i t_i(u_i' + Hq_i u_i'')$ up to any constant without loss of generality.

The 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 $\sum_i t_i u_i' \;=\; 0$. The first-order condition is: $2u_i'' t_i = \mu_1 u_i' + \mu_2(u_i' + Hq_i u_i'')$ where μ_1 and μ_2 are the Lagrange multipliers for the two constraints. This leads to t_i being proportional to:

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

for some μ (note that the second-order conditions are satisfied as the objective function is concave: $u_i'' < 0$ for all goods 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 correspond to:

$$\mu \;=\; -\frac{\sum_i \frac{u_i'^2}{Hu_i''}}{\sum_i q_i u_i'}$$

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

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

Next, using the optimal $t_i = \frac{u_i'}{H u_i''} - \bar{\varepsilon}_D q_i = q_i \varepsilon_{Di} - q_i \bar{\varepsilon}_D$, 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:

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

The term in parentheses is the same on the left and on the right. This term is negative iff:

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

$$
\iff \frac{\sum_i q_i u_i'}{\sum_i q_i u_i' \frac{1}{(-\varepsilon_{Di})}} < \frac{\sum_i q_i u_i' (-\varepsilon_{Di})}{\sum_i q_i u_i'}
$$

This last inequality is satisfied as long as price elasticity are not equal across all goods: 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 inequality is equivalent to:

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

This holds, given that $\bar{\varepsilon}_D$ is a weighted average of ε_{Di}, and $\varepsilon_{Di}\varepsilon_F < \varepsilon_H$ is assumed in part ii) of restrictions [A3] 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:

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

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

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

Using $H(\Lambda) \, u_i'(H(\Lambda) q_i) \; = \; \mu p_i$, we obtain:

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

and thus, given the definition of u_i':

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

and:

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

The final step is to show that Λ can be implicitly defined as a function of all normalized prices p_i/w. To see this, notice that q_i must satisfy the budget constraint:

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

which can be rewritten:

$$\sum_i (p_i/w) D_i(F(\Lambda)p_i/w)/H(\Lambda) = 1.$$

The solution of this equation in Λ is unique, which shows that we can alternatively define Λ as a function of normalized prices p/w. To prove that there is a unique solution, we can follow the same approach and assumptions as in Step 1 above: condition [A3]-ii) ensures uniqueness while condition [A3]-iii) provides existence.

Alternative proof of Proposition 3 using the Slutsky Matrix Alternatively, it is possible to prove Proposition 3 by showing that the Slutsky matrix is symmetric and negative semi-definite, and then apply Hurwicz and Uzawa (1971) theorem. This is the approach taken by Matsuyama and Ushchev (2017) for the homothetic case. A similar approach can be extended here to the non-homothetic case (see a previous working paper version, Fally 2018).

From direct to indirect utility We start from the following geometric equality that applies to any strictly monotonic mapping T:

$$\int_{q_0}^{q_1} T^{-1}(q)dq \; + \; T^{-1}(q_0)q_0 \; = \; -\int_{y_0}^{y_1} T(y)dy \; + \; T(y_1)y_1$$

with $q_0 = T(y_0)$ and $q_1 = T(y_1)$. Applying this formula to $T = D_i$, $q_1 = H(\Lambda)q_i$ and $y_1 = F(\Lambda)p_i/w$, we obtain:

$$\int_{q_{0i}}^{H(\Lambda)q_i} D_i^{-1}(q)dq \; = \; -\int_{y_{0i}}^{F(\Lambda)p_i/w} D_i(y)dy \; + \; D_i(F(\Lambda)p_i/w)F(\Lambda)p_i/w \; - \; y_{0i}q_{0i}$$

with $y_{0i} = D_i(q_{0i})$ for each i. Moreover, note that we have:

$$\sum_i (p_i/w) D_i(F(\Lambda)p_i/w) \; = \; H(\Lambda).$$

Applying these equalities to the expression for direct utility provided in the text, we obtain (indirect) utility as a function of normalized prices:

$$\begin{aligned}
U \;&=\; \sum_i u_i(H(\Lambda)q_i) \; - \; \int_{\Lambda_0}^{\Lambda} F(\Lambda)H'(\Lambda)d\Lambda \\
&=\; \sum_i \int_{q=0}^{H(\Lambda)q_i} D_i^{-1}(x)dx \; - \; \int_{\Lambda_0}^{\Lambda} F(\Lambda)H'(\Lambda)d\Lambda \\
&=\; -\sum_i \int_{y_{0i}}^{F(\Lambda)p_i/w} D_i(y)dy \; + \; \sum_i D_i(F(\Lambda)p_i/w)F(\Lambda)p_i/w \; - \; \int_{\Lambda_0}^{\Lambda} F(\Lambda)H'(\Lambda)d\Lambda \; - \; \sum_i y_{0i}q_{0i} \\
&=\; -\sum_i \int_{y_{0i}}^{F(\Lambda)p_i/w} D_i(y)dy \; + \; F(\Lambda)H(\Lambda) \; - \; \int_{\Lambda_0}^{\Lambda} F(\Lambda)H'(\Lambda)d\Lambda \; - \; \sum_i y_{0i}q_{0i} \\
&=\; -\sum_i \int_{y_{0i}}^{F(\Lambda)p_i/w} D_i(y)dy \; + \; \int_{\Lambda_0}^{\Lambda} F'(\Lambda)H(\Lambda)d\Lambda \; + \; F(\Lambda_0)H(\Lambda_0) \; - \; \sum_i y_{0i}q_{0i}
\end{aligned}$$

$$= \; -\sum_i \int_{y_{0i}}^{F(\Lambda)p_i/w} D_i(y)dy \; + \; \int_{\Lambda_0}^{\Lambda} F'(\Lambda)H(\Lambda)d\Lambda \; + \; g_0$$

where $g_0 = F(\Lambda_0)H(\Lambda_0) - \sum_i y_{0i}q_{0i}$ is a constant term.

A counter-example when condition [A3]-ii) fails.

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.

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:

$$\left[\sum_i (G_i(\Lambda)\, p_i/w)^{1-\sigma(\Lambda)}\right]^{\frac{1}{1-\sigma(\Lambda)}} = 1 \tag{70}$$

$$\left[\sum_i (G_i(U)/q_i)^{\frac{1-\sigma(U)}{\sigma(U)}}\right]^{\frac{\sigma(U)}{1-\sigma(U)}} = 1 \tag{71}$$

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 (70), 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 (70) 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 1 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 (76) 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 1 again, we obtain that the whole term strictly increases with U.

Hence, in both cases, Λ and U are well defined by equations (70) and (76) 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 (21) We mention in the text that condition ii) of Proposition 4 is equivalent to inequality (21) in the main text 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:

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

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 (21) in the main 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 p_i/w.

Lemma 1: *"Generalized Mean" inequality: For any given set of $x_i \geq 0$ and $\alpha_i \geq 0$ such that $\sum_i \alpha_i = 1$, the following expression is monotonically increasing in $\rho \in (-\infty, +\infty)$:*

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

Proof of Lemma 1: A complete proof is provided in Hardy, Littlewood and Polya (*Inequalities*, Cambridge University Press, 1934), see equation (2.9.1) on p.26, and is also referred to as the "Generalized Mean" inequality (e.g. on Wikipedia).

For convenience, here I report the proof for the case where both ρ and ρ' are positive. The same approach works for the case where they are negative. Consider two values $0 < \rho < \rho'$ 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:

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

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 $\rho = 0$.

Counter-examples when condition [A4] fails.

Here I provide counter-examples to show that Λ or U are not 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 (21) 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 this equation (equation 22 in the main text):

  $$\sum_i (q_i/G_i(U))^{\frac{\sigma(U)-1}{\sigma(U)}} = 1 \tag{72}$$

  but where implicit utility would depend negatively on some of the quantities. This amounts to showing that the following expression:

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

decreases with U and for at least some of the q_i's.

Suppose that U_0 is the solution of equation (72) 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 (72) . In particular, choose q_i^* such that U_0 is still the solution of (72) and such that:

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

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:

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

which corresponds to the left-hand side of equation (72). 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.

Proof of Proposition 5

About Λ . Before we prove parts i) and ii) of Proposition 5, note that conditions [A5] ensure that Λ can be implicitly defined by the budget constraint as either a function of quantities q or normalized prices p/w. Focusing on characterizing Λ as a function of quantities, we can follow the same approach as in Step 1 of the proof of Proposition 3.

First, using part ii) of restrictions [A5], we can see that the elasticity of $D_i(F(\Lambda,U)y_i)/H(\Lambda,U)$ w.r.t. Λ is given by $\varepsilon_F\varepsilon_{Di} - \varepsilon_H$ which is assumed to be negative, hence it strictly decreases with Λ. Symmetrically, we obtain that $D_i^{-1}(H(\Lambda,U)q_i,U)/F(\Lambda,U)$ also strictly decreases with Λ: its elasticity w.r.t. Λ is $\varepsilon_H/\varepsilon_{Di} - \varepsilon_F$, which is also negative given that ε_{Di} is negative and $\varepsilon_H - \varepsilon_{Di}\varepsilon_F$ is positive. Note that the budget constraint can be written as:

$$\sum_i q_i D_i^{-1}(q_i H(\Lambda,U),U)/F(\Lambda,U) = 1. \tag{73}$$

Adding up across goods, we obtain that the left-hand side of equation (73) decreases strictly with Λ. This implies that the solution in Λ to equation (73) is unique (if it exists).

Existence is then guaranteed using condition [A5]-iii), which we can symmetrically reformulate in terms of quantities. We assume that, for any good i, $y_i > 0$ and V, there exists $\Lambda \in \mathbb{R}$ such that: $y_i D_i\left(y_i F(\Lambda,V),V\right)/H(\Lambda,V) = 1/N$. Denote by D_i^{-1} the inverse with respect to the first argument of D_i). Note that this equality is equivalent to $1/N = 1/(Ny_i)D_i^{-1}\left(1/(Ny_i)H(\Lambda,V),V\right)/F(\Lambda,V)$.

Hence, denoting $q_i = 1/Ny_i$, we obtain that for any good i, $q_i > 0$ and U, there exists $\Lambda \in \mathbb{R}$ such that:

$$q_i D_i^{-1}\left(q_i H(\Lambda, U), U\right)/F(\Lambda, U) = 1/N.$$

For a given vector of quantities q, for each good we obtain a Λ such that the equality above holds. Taking the maximum of the Λ's obtained across the N goods (and using the monotonicity property in Λ described just above), we obtain a Λ^{max} such that

$$\sum_i q_i D_i^{-1}\left(q_i H(\Lambda^{max}, U), U\right)/F(\Lambda^{max}, U) \leq 1$$

For the same vector q, by taking the minimum of such Λ's across goods, we obtain a Λ^{min} such that

$$\sum_i q_i D_i^{-1}\big(q_i H(\Lambda^{min}, U), U\big)/F(\Lambda^{min}, U) \geq 1.$$

As the left hand side of this expression is continuous in Λ,the intermediate-value theorem ensures that a solution to equation (73) exists between Λ^{min} and Λ^{max}.

Hence the budget constraint, i.e. equation (73), can be used to uniquely define Λ as a function of q and U, or just as a function of q when we evaluate U at $U(q)$. Moreover, as in Proposition 3, since the left-hand side of equation (73) has a strictly non-zero (negative) derivative in Λ, we can use the Implicit Function Theorem to compute the derivatives of Λ. Here, we now have an equation that also depend on U, but we can still use the results as before (see Proposition 3) to compute the derivative $\frac{\partial \Lambda}{\partial q_i}\Big|_U$ w.r.t. q_i for each goods i along an indifference curve, i.e. holding U constant. This partial derivative will be useful for the proof of quasi-concavity of U, as discussed in part ii) further below.

i) Characterizing utility

The first part of Proposition 5 provides an equation that must be satisfied if demand can be rationalized and takes the form:

$$q_i(p_i/w, \Lambda, V) \; = \; \frac{1}{H(\Lambda, V)} D_i\left(\frac{p_i F(\Lambda, V)}{w}\, , \, V\right)$$

or in terms of inverse demand:

$$\frac{p_i}{w} \; = \; \frac{D_i^{-1}(H(\Lambda, U) q_i\, , \, U)}{F(\Lambda, U)}$$

satisfying equation (73). If demand can be rationalized with a differentiable utility function $U(q)$, there exists a function $\lambda(q)$ (real mapping from $\mathbb{R}_+^N$ to $\mathbb{R}_+$, such that

$$\frac{\partial U}{\partial q_i} = \lambda(q) \frac{1}{H(\Lambda, U)} D_i\left(\frac{p_i F(\Lambda, U)}{w}\, , \, U\right).$$

Define a function $M(q, \Lambda, U)$ as:

$$M(q, \Lambda, U) = \sum_i \int_{q=q_{i0}}^{q_i H(\Lambda, U)} D_i^{-1}(q, U) dq \; - \; \int_{\Lambda'=\Lambda_0}^{\Lambda} \frac{\partial H}{\partial \Lambda}(\Lambda', U) F(\Lambda', U) d\Lambda' \qquad (74)$$

The partial derivative of M w.r.t. q is:

$$\frac{\partial M}{\partial q_j} = H(\Lambda, U) D_j^{-1}(q_j H(\Lambda, U), U)$$

The partial derivative of M w.r.t. Λ is:

$$\begin{aligned}\frac{\partial M}{\partial \Lambda} &= \frac{\partial H}{\partial \Lambda}(\Lambda, U) \sum_i q_i D_i^{-1}(q_i H(\Lambda, U), U) dq \; - \; \frac{\partial H}{\partial \Lambda}(\Lambda, U) F(\Lambda, U) \\ &= \frac{\partial H}{\partial \Lambda}(\Lambda, U) \left[\sum_i q_i D_i^{-1}(q_i H(\Lambda, U), U) dq \; - \; F(\Lambda, U) \right].\end{aligned}$$

Note that this partial derivative null at $\Lambda = \Lambda(q)$ and $U = U(q)$ if the budget constraint is satisfied (condition 73).

Now, define $\widetilde{M}(q) = M(q, \Lambda(q), U(q))$, i.e. equal to M where U and Λ are evaluated at $U(q)$ and $\Lambda(q)$ respectively rather than treated as arguments. Note that, if demand is rational, marginal utility must be itself proportional to inverse demand and thus:

$$H(\Lambda, U) D_j^{-1}(q_j H(\Lambda, U), U) \; = \; H(\Lambda, U) D_j^{-1}(q_j H(\Lambda, U), U) \; = \; \frac{H(\Lambda, U) F(\Lambda, U)}{\lambda(q)} \frac{\partial U}{\partial q_j}$$

where λ is the Lagrange multiplier associated with the budget constraint, and where U and Λ are evaluated at $U(q)$ and $\Lambda(q)$. We obtain that the gradient of $\widetilde{M}$ is proportional to the gradient of utility:

$$\begin{aligned}\frac{\partial \widetilde{M}}{\partial q_j} &= \frac{\partial M}{\partial q_j} + \frac{\partial M}{\partial \Lambda} \frac{\partial \Lambda}{\partial q_j} + \frac{\partial M}{\partial U} \frac{\partial U}{\partial q_j} \\ &= H(\Lambda, U) D_j^{-1}(q_j H(\Lambda, U), U) \; + \; 0 \; + \; \frac{\partial M}{\partial U} \frac{\partial U}{\partial q_j} \\ &= \left[\frac{H(\Lambda, U) F(\Lambda, U)}{\lambda} \; + \; \frac{\partial M}{\partial U} \right] \frac{\partial U}{\partial q_j}.\end{aligned}$$

Given that indifference curves are connected, this implies that there exist a function $\widetilde{\widetilde{M}}(U)$ of utility such that: $\widetilde{M}(q) = \widetilde{\widetilde{M}}(U(q))$ for all q (see e.g. Lemma 1 of Goldman and Uzawa, 1954, for a proof of this statement). Hence, combining with equation (74), we obtain that $U(q)$ satisfies:

$$\widetilde{\widetilde{M}}(U(q)) = \sum_i \int_{q=q_{i0}}^{q_i H(\Lambda(q), U(q))} D_i^{-1}(q, U(q)) dq \; - \; \int_{\Lambda'=\Lambda_0}^{\Lambda(q)} \frac{\partial H}{\partial \Lambda}(\Lambda', U(q)) F(\Lambda', U(q)) d\Lambda'. \tag{75}$$

Defining $G(\Lambda, U)$ as:

$$G(\Lambda, U) \; = \; \widetilde{\widetilde{M}}(U) \; + \; \int_{\Lambda'=\Lambda_0}^{\Lambda(q)} \frac{\partial H}{\partial \Lambda}(\Lambda', U(q)) F(\Lambda', U(q)) d\Lambda'.$$

As described in Proposition 5, we obtain that $U(q)$ must satisfy:

$$\sum_i \int_{q=q_{i0}}^{q_i H(\Lambda(q),U(q))} D_i^{-1}(q,U(q))dq \;-\; G(\Lambda(q),U(q)) \;=\; 0. \tag{76}$$

ii) Properties of the utility function

We now examine the converse of part i): assuming that such equation admits a solution in U, does it yield a well-behaved utility function that is monotonic in each q_i, continuous and quasi-concave?

First, continuity is ensure by the fact that the left-hand side of equation (76) is continuous in q, Λ and U, and is assumed to strictly decrease with U (and Λ is itself a differentiable function of q). Hence we can solve for U as a continuous function of q.

Second, note that the left-hand side of equation (76) is strictly increasing in q_i, with a partial derivative (holding U constant) given by: $\frac{\partial M}{\partial q_j} = HD_j^{-1}(q_j H, U) > 0$ (with the partial derivative in Λ being null). As we assume that the left-hand side of equation (76) strictly decreases with U, the solution for U must be strictly increasing in q_i for each good i.

Quasi-concavity of U. Third and least obvious, we need to prove that the solution for utility U is quasi-concave in q. To do so, we can however build up on Step 2 of the proof of Proposition 3 provided earlier. In Proposition 3, we have already shown the quasi-concavity of the following function B, holding U constant:

$$B(q,U) = \sum_i \int_{q=q_{i0}}^{q_i H(\Lambda^*(q,U),U)} D_i^{-1}(q,U)dq - G(\Lambda^*(q,U),U)$$

(function B replaces the former utility function U in Proposition 3) with $\Lambda^*(q,U)$ defined such that the following condition holds:

$$\sum q_i D_i^{-1}(q_i H(\Lambda(q),U),U) \;=\; F(\Lambda,U)$$

again for a given U, where F is such that $\frac{\partial G}{\partial \Lambda}(\Lambda,U) = \frac{\partial H}{\partial \Lambda}(\Lambda,U)F(\Lambda,U)$.

We can then use the quasi-concavity of B (holding U constant) to prove the quasi-concavity of U, defined implicitly by $B(q,U(q) = 0$ (this implicit definition is equivalent to equation 76). The quasi-concavity of B implies that for any q and q' such that $B(q,U) = B(q',U) = 0$, we must have:

$$B(\alpha q + (1-\alpha)q', U) \geq B(q,U) = 0.$$

We can check also that the derivative of B in U is negative if $B(q,U) = 0$. Hence we obtain that: utility U' evaluated at $(\alpha q + (1-\alpha)$ is larger than $U(q) = U(q')$:

$$U' \equiv U(\alpha q + (1-\alpha)q') \geq U$$

for any $\alpha \in (0,1)$, since B is strictly decreasing in U and since U' must satisfy:

$$B(\alpha q + (1-\alpha)q', U') = 0.$$

The fact that $U(\alpha q + (1-\alpha)q') \geq U(q)$ whenever $U(q) = U(q')$ means that U is quasi-concave. We can also check that U is strictly quasi-concave if B is quasi-concave.

Indirect utility for the two-aggregator case

As for the single-aggregator case, we obtain:

$$\int_{q_{0i}}^{H(\Lambda,U)q_i} D_i^{-1}(q,U)dq = -\int_{D_i^{-1}(q_{0i},U)}^{F(\Lambda,U)p_i/w} D_i(y,U)dy + D_i(F(\Lambda,U)p_i/w,U)F(\Lambda,U)p_i/w - D_i^{-1}(q_{0i},U)q_{0i}$$

which holds for a given level of utility U. Moreover, note that we have:

$$\sum_i (p_i/w,V)D_i(F(\Lambda,V)p_i/w,V) \;=\; H(\Lambda,V).$$

Hence, summing across goods, we obtain:

$$\sum_i \int_{q_{0i}}^{H(\Lambda,U)q_i} D_i^{-1}(q,U)dq \;=\; -\sum_i \int_{D_i(q_{0i},U)}^{F(\Lambda,U)p_i/w} D_i(y,U)dy \;+\; H(\Lambda,U)F(\Lambda,U) \;-\; \sum_i D_i(q_{0i},U)q_{0i}.$$

Applying these equalities to the expression for direct utility provided in the text, we obtain a similar condition characterizing (indirect) utility as a function of normalized prices:

$$\begin{aligned}
\sum_i \int_{q_{0i}}^{H(\Lambda,U)q_i} D_i^{-1}(q,U)dq \;&=\; G(\Lambda,U)\\
\Leftrightarrow \sum_i \int_{D_i^{-1}(q_{0i},U)}^{F(\Lambda,U)p_i/w} D_i(y,U)dy \;&=\; -G(\Lambda,U) \;+\; H(\Lambda,U)F(\Lambda,U) \;-\; \sum_i D_i^{-1}(q_{0i},U)q_{0i}.
\end{aligned}$$

Next, using our definition of G (using function $\widetilde{\widetilde{M}}(U)$ defined above in the proof of Proposition 5) and integrating by parts, note that we have:

$$\begin{aligned}
G(\Lambda,U) \;&=\; \widetilde{\widetilde{M}}(U) \;+\; \int_{\Lambda'=\Lambda_0}^{\Lambda} \frac{\partial H}{\partial \Lambda}(\Lambda',U)F(\Lambda',U)d\Lambda'\\
&=\; \widetilde{\widetilde{M}}(U) \;+\; H(\Lambda,U)F(\Lambda,U) \;-\; H(\Lambda_0,U)F(\Lambda_0,U) \;-\; \int_{\Lambda'=\Lambda_0}^{\Lambda} \frac{\partial F}{\partial \Lambda}(\Lambda',U)H(\Lambda',U)d\Lambda'
\end{aligned}$$

and thus the equality above is equivalent to:

$$\sum_i \int_{y_{0i}}^{F(\Lambda,V)p_i/w} D_i(y,V)dy \;=\; K(\Lambda,V)$$

where function K is defined as:

$$\begin{aligned}
K(\Lambda,V) \;\equiv\; & \int_{\Lambda'=\Lambda_0}^{\Lambda} \frac{\partial F}{\partial \Lambda}(\Lambda',V)H(\Lambda',V)d\Lambda'\\
& -\sum_i D_i^{-1}(q_{0i},V)q_{0i} \;-\; \widetilde{\widetilde{M}}(V) \;+\; H(\Lambda_0,V)F(\Lambda_0,V) \;+\; \sum_i \int_{y_{0i}}^{D_i^{-1}(q_{0i},V)} D_i(y,V)dy.
\end{aligned}$$

Notice that the second line only depends on V, not Λ, hence: $\frac{\partial K}{\partial \Lambda}(\Lambda,V) \;=\; \frac{\partial F}{\partial \Lambda}(\Lambda,V)H(\Lambda,V)$.

Section 4) Practical cases and applications

Different forms of separability as special cases

Implicit separability If H does not depend on the aggregator Λ, we have: $\frac{\partial G}{\partial \Lambda}(\Lambda, U) = 0$, hence $G(\Lambda, U) = G(U)$. Without loss of generality, we can rescale function D_i by $1/G$ and impose $G(U) = 1$ after scaling. Utility U is then implicitly defined by

$$\sum_i \int_{q=q_{i0}}^{q_i} D_i^{-1}(q, U)\, dq \quad = \quad 1. \tag{77}$$

In this case, F must be a monotonic function of the aggregator Λ. It is also without loss of generality to assume $F(\Lambda, U) = \Lambda$.

Then, if $D_i^{-1}(q_i, U)$ is strictly decreasing in U, and takes values from the full interval $(+\infty, 0)$ as U decreases (conditional on q_i), the utility function defined implicitly by this equation is uniquely defined, for any q, and well-behaved.

Indirect implicit separability If F does not depend on the aggregator Λ, we can rescale function D_i such that it is without loss of generality to assume that $F = 1$. This also implies that function K obtained in equation (26) (in the main text) only depends on V, since $\frac{\partial K}{\partial \Lambda}(\Lambda, V) = H(\Lambda, V)\frac{\partial F}{\partial \Lambda}(\Lambda, V) = 0$. Hence, indirect utility can then be seen as the implicit solution of

$$\sum_i \int_{y_{0i}}^{p_i/w} D_i(y, V)/K(V) dy \quad = \quad 1.$$

Again, by rescaling D_i by K, it is without loss of generality to assume $K = 1$.

If $D_i(y, V)$ is strictly decreasing in V, and takes values from the full interval $(+\infty, 0)$ as V decreases (conditional on q_i), the indirect utility function defined implicitly by this equation is uniquely defined for all sets of prices p/w and well-behaved.

Parameterized non-homothetic CES Suppose that demand is in Proposition 4, with the following parameterization:

$$q_i = \alpha_i U^{\gamma_i} (p_i/w)^{-\sigma(U)} \qquad \text{with} \qquad \sigma(U) = \bar{\sigma} + \sigma_1 \log U$$

This can be derived from indirect and direct utility implicitly defined by:

$$\sum_i \alpha_i V^{\gamma_i} (p_i/w)^{1-\sigma(V)} = 1 \qquad \text{and} \qquad \sum_i (\alpha_i U^{\gamma_i})^{\frac{1}{\sigma(U)}}\, q_i^{\frac{\sigma(U)-1}{\sigma(U)}} = 1$$

In these summations, each term corresponds to the expenditure share on good i, in the primal or dual. To verify the equivalence, we start with $q_i = \alpha_i V^{\gamma_i}(p_i/w)^{-\sigma(V)}$ and see that:

$$\sum_i (\alpha_i U^{\gamma_i})^{\frac{1}{\sigma(U)}}\, q_i^{\frac{\sigma(U)-1}{\sigma(U)}} \quad = \quad \sum_i (\alpha_i U^{\gamma_i})^{\frac{1}{\sigma(U)}} \left(\alpha_i U^{\gamma_i} (p_i/w)^{-\sigma(U)}\right)^{\frac{\sigma(U)-1}{\sigma(U)}}$$

$$= \sum_i (\alpha_i U^{\gamma_i})^{\frac{1}{\sigma(U)}} (\alpha_i U^{\gamma_i})^{\frac{\sigma(U)-1}{\sigma(U)}} (p_i/w)^{1-\sigma(U)} \quad = \quad \sum_i \alpha_i U^{\gamma_i} (p_i/w)^{1-\sigma(U)} \quad = \quad 1$$

To show that such demand is rational, we need to show:

$$\sum_i \exp\left(\frac{(\sigma(U)-1)^2 G_i'(U)}{\sigma_1 G_i(U)/U}\right) < 1 \tag{78}$$

When σ_1 is negative and $\sigma'(U) = \sigma_1/U < 0$, inequality (78) provides a condition for rationalization that conveniently simplifies. With $G_i(U) = (\alpha_i U^{\gamma_i})^{\frac{1}{1-\sigma(U)}}$ and $\sigma(U) = \bar{\sigma} + \sigma_1 \log U$, we take logs and obtain:

$$G_i'(U)/G_i(U) = \frac{1}{U}\frac{1}{1-\sigma(U)}\left[\gamma_i \ + \ \sigma_1 \log G_i(U)\right]$$

Hence the terms in the summation are:

$$(1-\sigma(U))^2 U G_i'(U)/[\sigma_1 G_i(U)] = (1-\sigma(U))\gamma_i/\sigma_1 \ + \ \log(\alpha_i U^{\gamma_i})$$

The inequality is then equivalent to:

$$\begin{aligned}
& \sum_i \exp\left(\frac{(\sigma(U)-1)^2 G_i'(U)}{\sigma_1 G_i(U)/U}\right) < 1 \\
\iff \quad & \sum_i \exp\left[(1-\bar{\sigma}-\sigma_1 \log U)\gamma_i/\sigma_1 \ + \ \log(\alpha_i U^{\gamma_i})\right] < 1 \\
\iff \quad & \sum_i \alpha_i \exp\left[\gamma_i \frac{1-\bar{\sigma}}{\sigma_1}\right] \ < \ 1
\end{aligned}$$

If we assume $\bar{\sigma} > 1$, $\sigma_1 < 0$ and $\gamma_i < 0$, the term in brackets is negative. Hence, this inequality is satisfied if we also assume $\sum_i \alpha_i = 1$.

Direct semi-separability Preferences as directly semi-separable if utility is:

$$U(q) = \frac{1}{G(\Lambda)} \sum_i R_i(H(\Lambda) q_i) \tag{79}$$

where H, G and R_i are twice continuously-differentiable, with $G' > 0$, $H' > 0$, $R_i' > 0$ and $R_i'' < 0$ and where Λ is such that:

$$\frac{\sum_i q_i R_i'(H(\Lambda) q_i)}{\sum_i R_i(H(\Lambda) q_i)} = \frac{F(\Lambda)}{G(\Lambda)} \tag{80}$$

where $F(\Lambda) \equiv G'(\Lambda)/H'(\Lambda)$. Again, it may be useful though not necessary to assume $R_i(0) = 0$ (i.e. that there is no gain from a new variety when its consumption is zero).

This demand system is a special case of Proposition 5, as this corresponds to defining $D_i(y_i, V) = R_i'^{-1}(V y_i)$ and specifying F, G, and H as functions of Λ only. Here I provide again a derivation of demand for this special case.

First, note that the derivative of the right-hand-side of (79) is equal to:

$$\frac{1}{G(\Lambda)^2}\left[\sum_i q_i R_i'(H(\Lambda) q_i) H'(\Lambda) G(\Lambda) - \sum_i R_i(H(\Lambda) q_i) G'(\Lambda)\right]$$

which is null if condition (79) is satisfied. Hence marginal utility is given by the derivative of (79) holding Λ constant. This yields:

$$\lambda p_i/w = \frac{\partial \widetilde{U}}{\partial q_i} = \frac{H(\Lambda)}{G(\Lambda)} R_i'(H(\Lambda)q_i).$$

The budget constraint implies:

$$\lambda \ = \ \lambda \sum_i q_i p_i/w = \frac{H(\Lambda)}{G(\Lambda)} \sum_i q_i R_i'(H(\Lambda)q_i) = \frac{H(\Lambda)}{H'(\Lambda)G(\Lambda)^2} \sum_i R_i(H(\Lambda)q_i)G'(\Lambda) = \frac{H(\Lambda)G'(\Lambda)U}{H'(\Lambda)G(\Lambda)}$$

And thus we obtain the following expression for inverse demand:

$$p_i/w = \frac{H(\Lambda)R_i'(H(\Lambda)q_i)}{\lambda G(\Lambda)} = \frac{H'(\Lambda)R_i'(H(\Lambda)q_i)}{UG'(\Lambda)} = \frac{R_i'(H(\Lambda)q_i)}{UF(\Lambda)}$$

where $F(\Lambda) = G'(\Lambda)/H'(\Lambda)$. Re-inverting, we obtain Marshallian demand for good i:

$$q_i = R_i'^{-1}(VF(\Lambda)p_i/w)/H(\Lambda)$$

Conditions [A5]-ii) required by Proposition 5 is met if $R_i'^{-1}(F(\Lambda)y_i)/H(\Lambda)$ has a strictly negative derivative in Λ. Conditions iii) is met if this expression goes from $+\infty$ to 0 (in the limit) as Λ increases. Hence equation (30) in the main text has a unique solution in the aggregator Λ.

Written as in Proposition 5, the condition characterizing utility is:

$$\frac{\sum_i R_i(H(\Lambda)q_i)}{UG(\Lambda)} \ = \ 1.$$

In this case, it is obvious that it is strictly decreasing in U (holding Λ and q constant), and that a solution in U exists.

Indirect semi-separability Preferences as indirectly semi-separable if indirect utility can be written:

$$V = \frac{\sum_i S_i(F(\Lambda)p_i/w)}{L(\Lambda)} \tag{81}$$

where F, L and S_i are twice continuously-differentiable, with $F' > 0$, $L' < 0$, $S_i' < 0$, $S_i'' > 0$, and where Λ is such that:

$$\frac{\sum_i (p_i/w)D_i(F(\Lambda)p_i/w)}{\sum_i S_i(F(\Lambda)p_i/w)} = \frac{H(\Lambda)}{K(\Lambda)} \tag{82}$$

where we define $D_i(y_i) = -S_i'(y_i)$ and $H(\Lambda) = -L'(\Lambda)/F'(\Lambda)$. Again, it may be useful though not necessary to assume $\lim_{y \to +\infty} S_i(y) = 0$ (i.e. that there is no gain from a new variety when its price is prohibitive).

Such indirect utility function is again a special case of the dual-aggregator form that we studied in Proposition 5, with $D_i(y_i, V) = -S_i'(y_i)/V$ and specifying F and H as functions of Λ only. Condition [A5]-ii) required by Proposition 5 is met if $D_i(F(\Lambda)y_i)/H(\Lambda)$ has a strictly negative derivative in Λ. Condition [A5]-iii) is met if this term goes from $+\infty$ to 0 (in the limit) as Λ increases.

Using Roy's identity, we can check that demand for good i equals:

$$q_i = \frac{D_i(F(\Lambda)p_i/w)}{VH(\Lambda)}.$$

We can switch for a characterization of indirect utility to a characterization of direct utility by integrating by part.

From equation (81), we obtain:

$$\sum_i S_i(D_i^{-1}(UHq_i)) \;=\; LU.$$

From the budget constraint, we get:

$$\sum_i (UHq_i)D_i^{-1}(UHq_i)) \;=\; UHF.$$

Adding up the previous two equalities, we obtain:

$$\sum_i S_i(D_i^{-1}(UHq_i)) \;+\; \sum_i (UHq_i)D_i^{-1}(UHq_i)) \;=\; LU \;+\; UHF.$$

Denote by $S_{0i} = \lim_{p\to+\infty} S_i(p)$, which is well defined since S_i is positive and decreasing. For each good i, we have the following geometric equality (integration by part):

$$S_i(D_i^{-1}(q)) + qD^{-1}(q) = S_{0i} + \int_0^q D^{-1}(q')dq'.$$

Plugging this into the previous equality and dividing by U, we obtain:

$$\sum_i \left[\frac{S_{0i}}{U} + \frac{1}{U}\int_0^{UHq_i} D^{-1}(q')\,dq'\right] \;-\; (L+HF) \;=\; 0.$$

This equation in U corresponds to the characterization of utility in Proposition 5, with $G(\Lambda) = L(\Lambda) + H(\Lambda)F(\Lambda)$. Note that the left-hand side is strictly decreasing in U so that the solution in U is unique.

Homothetic semi-separability An interesting case of semi-separability is the homothetic case. This happens when $G(\Lambda)$ and $H(\Lambda)$ are iso-elastic (see earlier for the case of direct semi-separability). In particular, suppose that utility is defined (with Q as an aggregator) as:

$$U(q)^{\frac{\eta}{\eta+1}} \;=\; \tfrac{\eta}{\eta+1}\,Q^{\frac{\eta}{\eta+1}}\sum_i R_i(q_i/Q) \tag{83}$$

with $\eta > 0$, and denote $D_i(q) = R_i'$, with the same assumptions on R_i and D_i as above.

Suppose also that aggregator Q is such that the partial derivative of the RHS in Q is null, so that:

$$\sum_i (q_i/Q)R_i'(q_i/Q) \;=\; \tfrac{\eta}{\eta+1}\sum_i R_i(q_i/Q) \;=\; (U/Q)^{\frac{\eta}{\eta+1}} \tag{84}$$

Marginal utility is proportional to $R_i'(q_i/Q)$, so expenditure shares must be proportional to $(q_i/Q)R_i'(q_i/Q)$. Given equality (84) above and given the budget constraint, we must have:

$$p_iq_i/w \;=\; (q_i/Q)R_i'(q_i/Q)(U/Q)^{-\frac{\eta}{\eta+1}}$$

Denote:

$$D_i(q) = R_i'^{-1}(q)$$

Marshallian demand is then:

$$q_i \;=\; QD_i\Big((U/Q)^{\frac{\eta}{\eta+1}}Qp_i/w\Big)$$

These preferences are well defined and satisfy the rationality conditions above if

$$\eta + 1 \;<\; -\varepsilon_{Di}$$

is satisfied for all goods and all consumption baskets.

If instead we have: $\eta+1 > -\varepsilon_{Di}$ across all goods and baskets, we can consider a change in variable $Q' = 1/Q$ to safisfy the conditions for rationality above in aggregator Q' instead of Q.

Defining $\Lambda = (U/Q)^{\frac{\eta}{\eta+1}}Q/w$, so that $Q \;=\; (w\Lambda)^{\eta+1}U^{-\eta}$, we obtain:

$$q_i \;=\; (w\Lambda)^{\eta+1}U^{-\eta}D_i\Big(\Lambda p_i\Big) \;=\; w\Lambda^{\eta+1}P^{\eta}D_i\Big(\Lambda p_i\Big)$$

with $U = w/P$, where P denotes the price index. This corresponds to the equation in the text.

An alternative way is to define the price index as:

$$P^{-\eta} = \eta\Lambda^{\eta}\sum_i S_i(\Lambda p_i)$$

with Λ implicitly defined as a function of prices by:

$$\eta\sum_i S_i(\Lambda p_i) \;=\; \sum_i(\Lambda p_i)D_i(\Lambda p_i)$$

with $D_i = -S_i' > 0$. Since expenditure shares are proportional to $(\Lambda p_i)D_i(\Lambda p_i)$ and since:

$$P^{-\eta}\Lambda^{-\eta} \;=\; \eta\sum_i S_i(\Lambda p_i) \;=\; \sum_i(\Lambda p_i)D_i(\Lambda p_i)$$

expenditure shares must coincide with $(\Lambda p_i)D_i(\Lambda p_i)P^{\eta}\Lambda^{\eta}$, and thus again we obtain:

$$q_i \;=\; wP^{\eta}\Lambda^{\eta+1}D_i(\Lambda p_i)$$

Symmetric homothetic QMOR

Taking $D_i(y) = \alpha_i y^{r-1} + \beta_i y^{\kappa-1}$ and $F(\Lambda) = \Lambda$ and $H(\Lambda) = \Lambda^{r-1}$, we obtain that the ideal price index P is then implicitly defined by:

$$\sum_i \alpha_i\left(\frac{p_i\Lambda}{P}\right)^r + \frac{1}{\kappa}\sum_i \beta_i\left(\frac{p_i\Lambda}{P}\right)^{\kappa r} \;-\; \Lambda^r \;=\; c_0.$$

for some constant term c_0, and where aggregator Λ satisfies:

$$\sum_i \alpha_i\left(\frac{p_i\Lambda}{P}\right)^r \;+\; \sum_i \beta_i\left(\frac{p_i\Lambda}{P}\right)^{\kappa r} \;=\; \Lambda^r.$$

Taking the difference between the previous two equations leads to:

$$(P/\Lambda)^{\kappa r} \;=\; \frac{1}{c_0}\left(\frac{1}{\kappa}-1\right)\sum_i \beta_i p_i^{\kappa r}.$$

Normalizing $\frac{1}{c_0}\left(\frac{1}{\kappa}-1\right)=1$ so that $\Lambda^{-\kappa r} \;=\; \sum_i \beta_i \left(\frac{p_i}{P}\right)^{\kappa r}$, we obtain a price index of such form:

$$P^r \;=\; \sum_i \alpha_i p_i^r \;+\; \left(\sum_i \beta_i p_i^{\kappa r}\right)^{\frac{1}{\kappa}}.$$

Taking the log derivative w.r.t. log price p_i, we obtain the expenditure share in good i (Shepard's Lemma):

$$\frac{p_i q_i}{w} \;=\; \alpha_i \left(\frac{p_i}{P}\right)^r \;+\; \beta_i \left(\frac{p_i}{P}\right)^{\kappa r} \Lambda^{-r(1-\kappa)}$$

and thus:

$$\begin{aligned} q_i \;&=\; \frac{\alpha_i w}{P}\left(\frac{p_i}{P}\right)^{r-1}\left[1+\frac{\beta_i}{\alpha_i}\left(\frac{\Lambda p_i}{P}\right)^{-r(1-\kappa)}\right] \\ &=\; \frac{w}{P}\left(\frac{p_i}{P}\right)^{r-1}\left[\alpha_i + \beta_i p_i^{-r(1-\kappa)}\left(\sum_j \beta_j p_j^{\kappa r}\right)^{\frac{1-\kappa}{\kappa}}\right]. \end{aligned}$$

With $\kappa = 1/2$, $\alpha_i = \alpha$ and $\beta_i = \beta$, we get symmetric QMOR used in Freenstra (2010). When $\alpha_i > 0$ and $\beta_i < 0$, note that we get a finite reservation price (choke price).

A non-homothetic version of QMOR

Here we adopt the notation from Mrazova and Neary (2013). The notation used previously for homothetic case corresponds to $r = 1-\nu$ and $\kappa = (\sigma-1)/(\nu-1)$.

We have then:

$$\sum_i \alpha_i(V)\left(\frac{p_i}{w}\Lambda\right)^{1-\nu} + \frac{\nu-1}{\sigma-1}\sum_i \beta_i(V)\left(\frac{p_i}{w}\Lambda\right)^{1-\sigma} \;-\; \Lambda^{1-\nu} \;=\; c_0$$

where aggregator Λ satisfies:

$$\sum_i \alpha_i(V)\left(\frac{p_i}{w}\Lambda\right)^{1-\nu} + \sum_i \beta_i(V)\left(\frac{p_i}{w}\Lambda\right)^{1-\sigma} \;-\; \Lambda^{1-\nu} \;=\; 0.$$

Taking the difference between the last two equations, we obtain:

$$\left(\frac{\nu-\sigma}{\sigma-1}\right)\sum_i \beta_i(V)\left(\frac{p_i}{w}\Lambda\right)^{1-\sigma} \;=\; c_0.$$

Hence, setting $c_0 = \left(\frac{\nu-\sigma}{\sigma-1}\right)$, we get:

$$\Lambda^{\sigma-1} \;=\; \sum_i \beta_i(V)\left(\frac{p_i}{w}\right)^{1-\sigma}.$$

Plugging into the previous equation for Λ, we get:

$$\sum_i \alpha_i(V)\left(\frac{p_i}{w}\right)^{1-\nu} + \left(\sum_i \beta_i(V)\left(\frac{p_i}{w}\right)^{1-\sigma}\right)^{\frac{1-\nu}{1-\sigma}} \;=\; 1.$$

Demand for good i is then:

$$q_i \;=\; \alpha_i(V)\left(\frac{p_i}{w}\right)^{-\nu} + \beta_i(V)\Lambda^{\nu-\sigma}\left(\frac{p_i}{w}\right)^{-\sigma}.$$

We obtain the equation in the main text by plugging the expression for Λ.

Linear demand

Even with a simple linear demand in partial equilibrium, there are multiple ways to rationalize such demand functions with one or two aggregators.

Suppose that demand is linear for each good i (with the caveat that preferences are satiated above a certain level). In the most general case with two aggregators Λ and V, we obtain that demand must take the form:

$$q_i \;=\; \frac{\alpha_i(V) \;-\; F(\Lambda,V)p_i/w}{H(\Lambda,V)\gamma_i(V)}$$

(or zero if the latter is negative), where V is indirect utility and where Λ satisfies:

$$\sum_i (p_i/w)\max\left\{0\,,\,\frac{\alpha_i(V) \;-\; F(\Lambda,V)p_i/w}{H(\Lambda,V)\gamma_i(V)}\right\} \;=\; 1$$

and which can be obtained from a utility that satisfies:

$$\sum_i \left[\alpha_i(U)H(\Lambda,U)\,q_i \;-\; \frac{1}{2}\gamma_i(U)H(\Lambda,U)^2\,q_i^2\right] \;-\; G(\Lambda,U) = 0$$

where each q_i must not exceed $\frac{\alpha_i(U)}{H(\Lambda,U)\gamma_i(U)}$. Λ is uniquely defined if H and F are both increasing in Λ, and a solution in Λ always exists if H and F span from 0 to $+\infty$ at the limit. In turn, the solution in U is unique if we have the following monotonicity conditions (sufficient conditions), with strict monotonicity for at least one of them: $\alpha_i(U)$ decreases in U, $\gamma_i(U)$ increases in U, $H(\Lambda,U)$ decreases in U and $G(\Lambda,U)$ increases in U.

To illustrate the versatility of this approach and the many ways to specify the demand shifters, several special cases are worth noting:

- Directly-additive preferences can generate such linear demand and yield: $q_i \;=\; \frac{\alpha_i - \Lambda p_i/w}{\gamma_i}$
- Indirectly-additive preferences yield: $q_i \;=\; \frac{\alpha_i - p_i/w}{\Lambda\gamma_i}$
- Single aggregator preferences yields: $q_i \;=\; \frac{\Lambda\alpha_i - \Lambda^2 p_i/w}{\gamma_i}$
- Homothetic preferences yield: $q_i \;=\; \frac{w}{P}\,.\,\frac{\alpha_i - F(\Lambda)p_i/P}{H(\Lambda)\gamma_i}$
- Directly implicitly-separable preferences yield: $q_i \;=\; \frac{\alpha_i(V) - \Lambda p_i/w}{\gamma_i(V)}$

- Indirectly implicitly-separable preferences yield: $q_i = \frac{\alpha_i(V) - p_i/w}{\Lambda\gamma_i(V)}$
- Directly semi-separable preferences yield: $q_i = \frac{\alpha_i - F(\Lambda)Vp_i/w}{H(\Lambda)\gamma_i}$
- Indirectly semi-separable preferences yield: $q_i = \frac{\alpha_i - F(\Lambda)p_i/w}{VH(\Lambda)\gamma_i}$

Translog cost function

Translog costs functions have been studied in a variety of contexts, from consumer theory to productivity estimation. While a general formulation specifies the price index as:

$$\log P = \alpha_0 + \sum_i \alpha_i \log p_i + \frac{1}{2}\sum_{i,j} \gamma_{ij} \log p_i \log p_j$$

with $\alpha_i > 0$, $\sum_i \alpha_i = 1$ and $\gamma_{ij} = \gamma_{ji}$ required for rationalization, applications often typically impose a symmetric parameterization across the γ's, i.e. assume $\gamma_{ii} = \gamma/N - \gamma$ and $\gamma_{ij} = \gamma/N$ if $i \neq j$, with $\gamma > 0$.

As shown by Bergin and Feenstra (2009), the Symmetric Translog case leads to the following expenditure shares once we account for unavailable goods (or, equivalently, goods with prices above the choke price):

$$\frac{p_i q_i}{w} = \alpha_i + (1 - n\bar{\alpha}) + \gamma \left[\overline{\log p} - \log p_i\right]$$

where $\overline{\log p}$ denotes the average price across available varieties and $\overline{\alpha}$ is the average shifter α_i across available varieties, and n is the number of available varieties with $q_i > 0$. Defining the aggreagtor as $\log \Lambda = -\overline{\log p} - (1 - n\bar{\alpha})/\gamma$, we can reformulate the expenditure share as:

$$\frac{p_i q_i}{w} = \alpha_i - \gamma \log(\Lambda p_i/w).$$

This corresponds to demand in Proposition 3 with $D_i(y) = \alpha_i - \gamma \log y$, $F(\Lambda) = 1/H(\Lambda) = \Lambda$, and is well defined even if such demand has a choke price. One can then notice that aggregator Λ is uniquely determined by the budget constraint:

$$\sum_i \max\{0, \alpha_i - \gamma \log(\Lambda p_i/w)\} = 1$$

and that the price index can be obtained as:

$$\log P = \sum_i \alpha_i \log(\Lambda p_i/w) - \frac{\gamma}{2}\sum_i \Big(\log(\Lambda p_i/w)\Big)^2 - \log \Lambda.$$

Proof of Proposition 6

Suppose that demand take the form:

$$q_i = Q D_i(F p_i/w) \tag{85}$$

where Q and F are two aggregators that are functions of normalized prices p/w (i.e. functions homogeneous of degree zero of prices and income). Suppose also that utility is strictly quasi-concave and that the functions D_i are invertible.

As a first step, we can see that similar properties applies to inverse demand. Inverting expression (86), we obtain that we can write expenditure shares W_j as a function of own quantity q_j and the two aggregators Q and F by defining:

$$W_i(q_i, Q, F) \;=\; (Q/F)\, r_i(q_i/Q) \tag{86}$$

where we define r_i as:

$$r_i(q) \;=\; q D_i^{-1}(q).$$

Aggregators Q and F are initially defined as functions of vector of normalized prices, p/w. But since utility is assumed to be strictly quasi-concave, p/w can be expressed as a function of the vector of quantities q. Hence Q and F can also be viewed as aggregators that are functions of quantities q, so that expenditure shares can be written as $W_i(q_j, Q(q), F(q)) = (Q(q)/F(q))\, r_i(q_i/Q(q))$.

As stated, suppose that the set of gradients $\left\{\frac{\partial Q}{\partial \log p_i}, \frac{\partial F}{\partial \log p_i}\right\}$ is of rank two for all (p, w). Invertibility of demand (q as a function of p/w and vice-versa) also ensures that the rank of: $\left\{\frac{\partial Q}{\partial \log p_i}, \frac{\partial F}{\partial \log p_i}\right\}$ (as a function of normalized prices) is the same as the rank of $\left\{\frac{\partial Q}{\partial \log q_i}, \frac{\partial F}{\partial \log q_i}\right\}$ (as a function of quantities) evaluated at $q = q(p/w)$.

Differentiating the budget constraint $\sum_i r_i(q_i/Q) = F/Q$ implies:

$$\frac{\partial r_j}{\partial \log q_j} \;-\; \left(\sum_i \frac{\partial r_i}{\partial \log q_i}\right) \frac{\partial \log Q}{\partial \log q_j} \;=\; \frac{\partial (F/Q)}{\partial \log q_j}.$$

Hence, $\frac{\partial r_j}{\partial \log q_j}$ is colinear to the gradients $\frac{\partial Q}{\partial \log q_j}$ and $\frac{\partial F}{\partial \log q_j}$:

$$\frac{\partial r_j}{\partial \log q_j} \;=\; \frac{1}{Q}\frac{\partial F}{\partial \log q_j} \;+\; \left(\sum_i \frac{\partial r_i}{\partial \log q_i}\right) \frac{1}{Q}\frac{\partial Q}{\partial \log q_j} \;-\; \frac{F}{Q^2}\frac{\partial Q}{\partial \log q_j}. \tag{87}$$

If demand is rational and can be derived from utility maximization, we must have:

$$\frac{\partial U}{\partial \log q_i} \;=\; (\lambda Q/F)\, r_i(q_i/Q) \;\equiv\; \Lambda\, r_i(q_i/Q)$$

where we define the new aggregator $\Lambda = \lambda Q/F$ as a function of marginal utility λ and the two aggregators Q and F. Differentiating, we get:

$$\frac{\partial U}{\partial \log q_i \partial \log q_j} \;=\; \frac{\partial \Lambda}{\partial \log q_j} r_i \;-\; \Lambda \frac{\partial r_i}{\partial \log q_i}\frac{\partial \log Q}{\partial \log q_j}.$$

The cross derivative must be symmetric, hence, dividing by Λ we obtain:

$$\frac{1}{\Lambda}\frac{\partial \Lambda}{\partial \log q_j} r_i \;-\; \frac{\partial r_i}{\partial \log q_i}\frac{\partial \log Q}{\partial \log q_j} \;=\; \frac{1}{\Lambda}\frac{\partial \Lambda}{\partial \log q_i} r_j \;-\; \frac{\partial r_j}{\partial \log q_j}\frac{\partial \log Q}{\partial \log q_i}.$$

Rearranging, and using again $\Lambda r_i = \frac{\partial U}{\partial \log q_i}$, we obtain:

$$\frac{1}{\Lambda^2}\frac{\partial \Lambda}{\partial \log q_j}\frac{\partial U}{\partial \log q_i} \;+\; \frac{\partial r_j}{\partial \log q_j}\frac{\partial \log Q}{\partial \log q_i} \;=\; \frac{1}{\Lambda^2}\frac{\partial \Lambda}{\partial \log q_i}\frac{\partial U}{\partial \log q_j} \;+\; \frac{\partial r_i}{\partial \log q_i}\frac{\partial \log Q}{\partial \log q_j}. \tag{88}$$

Incorporating (87) into (88) and simplifying, we obtain:

$$\frac{1}{\Lambda^2}\frac{\partial \Lambda}{\partial \log q_j}\frac{\partial U}{\partial \log q_i} + \frac{1}{Q^2}\frac{\partial F}{\partial \log q_j}\frac{\partial Q}{\partial \log q_i} = \frac{1}{\Lambda^2}\frac{\partial \Lambda}{\partial \log q_i}\frac{\partial U}{\partial \log q_j} + \frac{1}{Q^2}\frac{\partial F}{\partial \log q_i}\frac{\partial Q}{\partial \log q_j}. \tag{89}$$

The remainder of the proof exploits this symmetry condition (89) to show that (Λ, U) can provide an alternative set of aggregators to (Q, F).

Take a vector x such that $\sum_i x_i \frac{\partial \Lambda}{\partial \log q_i} = 0$. Multiplying equation (89) by x_i and summing across goods i, we obtain:

$$\frac{1}{\Lambda^2}\frac{\partial \Lambda}{\partial \log q_j}\left(\sum_i x_i \frac{\partial U}{\partial \log q_i}\right) + \frac{1}{Q^2}\frac{\partial F}{\partial \log q_j}\left(\sum_i x_i \frac{\partial Q}{\partial \log q_i}\right) = \frac{1}{Q^2}\left(\sum_i x_i \frac{\partial F}{\partial \log q_i}\right)\frac{\partial Q}{\partial \log q_j} \tag{90}$$

If for all x, we also get $\sum_i x_i \frac{\partial U}{\partial \log q_i} = 0$, then we can see that the gradients of Q and F are colinear, which contradicts the assumption that they are not. Hence there exists x such that $\sum_i x_i \frac{\partial U}{\partial \log q_i} \neq 0$ while we still have $\sum_i x_i \frac{\partial \Lambda}{\partial \log q_i} = 0$. We can see from equation (90) that is implies that the gradient of U is colinear with the gradients of F and Q.

Similarly, since the gradients of U and Λ are not colinear, we can find a vector z such that $\sum_i z_i \frac{\partial U}{\partial \log q_i} = 0$ and $\sum_i z_i \frac{\partial \Lambda}{\partial \log q_i} \neq 0$. Multiplying equation (89) by z_i and summing across goods i, we obtain:

$$\frac{1}{Q^2}\frac{\partial F}{\partial \log q_j}\left(\sum_i z_i \frac{\partial Q}{\partial \log q_i}\right) = \frac{1}{\Lambda^2}\frac{\partial \Lambda}{\partial \log q_j}\left(\sum_i z_i \frac{\partial U}{\partial \log q_i}\right) + \frac{1}{Q^2}\left(\sum_i z_i \frac{\partial F}{\partial \log q_i}\right)\frac{\partial Q}{\partial \log q_j} \tag{91}$$

This implies that the gradient of Λ is also colinear with the gradients of F and Q. Since the gradients of Λ and U are not colinear with each other, we obtain that the gradients of Λ and U offers an alternative basis on which we can project the gradients of F and Q.

Aggregates F and Q can thus be written as functions of $U(q)$ and such an aggregate $\Lambda(q)$. Conversely, coming back to Marshallian demand instead of inverse demand, this also proves that we can express F and Q as a function of indirect utility $V(p/w)$ and an aggregate $\Lambda(p/W)$ that is function of normalized prices p/w. Hence, such demand system is a special case of Proposition 2 and 5.

Section 5) Application to monopolistic competition

Frechet differentiability with a continuum of goods

The continuum of goods is $[0, \bar{N}]$, and a consumption profile is defined as $q \in L^2[0, \bar{N}]$. From here onward, we denote the Lebesgue space $L^n[0, \bar{N}]$ by L^n to simplify notation.

We would like to define utility implicit as a mapping from L^2 to $\mathbb{R}$ that satisfies:

$$\frac{\int_{i=0}^{\bar{N}} \int_{q=0}^{q_i H(\Lambda,U)} D^{-1}(q, U)\, dq\, di}{G(\Lambda, U)} = 1 \tag{92}$$

where aggregator Λ is itself a solution to:

$$\frac{\int_0^{\bar{N}} q_i D^{-1}(q_i H(\Lambda, U), U) di}{F(\Lambda, U)} = 1. \tag{93}$$

For utility to be well-defined and Frechet differentiable in $q \in L^2$, the following conditions are needed:

- First, note that the two integral sums in equations (92) and (93) are well-defined and finite for any $q \in L^2$. For the first one, we have:

$$\int_{i=0}^{\bar{N}} \int_{q=0}^{Hq_i} D^{-1}(q,U)dqdi < \int_{i=0}^{\bar{N}} \int_{q=0}^{HA} D^{-1}(q,U)dqdi + D^{-1}(HA,U) \int_{i=0}^{\bar{N}} (q_i - A)\mathbb{1}_{\{q_i > A\}}\, di < +\infty$$

  for any constant term $A > 0$, since $D^{-1}(q,U)$ is decreasing in q. This integral is finite as we already assume that $\int_{q=0}^{q_i} D^{-1}(q,U)\,dq$ is finite, and $q \in L^2$ (which implies that $q \in L^1$ since we are working over a bounded segment $[0,\bar{N}]$). For the second one, note that we already assume $\lim_{q_i \to 0} q_i D^{-1}(q_i,U) = 0$ (i.e. the marginal utility form a good increases by less than $1/q$ when q decreases).

$$\int_0^{\bar{N}} q_i D^{-1}(q_i H,U)di < \int_0^{\bar{N}} q_i D^{-1}(q_i H,U)\mathbb{1}_{\{q_i > A\}}di + D^{-1}(AH,U) \int_0^{\bar{N}} q_i \mathbb{1}_{\{q_i > A\}}di < +\infty$$

- Next, as we define U implicitly as the solution of the system of equations (92) and (93), we need the Jacobian of the LHS to be well defined. The derivatives w.r.t. U depend on:

$$\int_{i=0}^{\bar{N}} \int_{q=0}^{q_i} \frac{\partial D^{-1}}{\partial U}(q,U)\,dq\,di \qquad \text{and} \qquad \int_{i=0}^{\bar{N}} q_i \frac{\partial D^{-1}}{\partial U}(q_i,U)\,di.$$

  We need to assume that those are well-defined and finite for any $q \in L^2$, a property that is not necessarily implied by the other assumptions made above.

  The derivatives w.r.t. Λ are $\int_{i=0}^{\bar{N}} q_i D^{-1}(q_i,U)\,di$ and $\int_{i=0}^{\bar{N}} q_i^2 \frac{\partial D^{-1}}{\partial q}(q_i,U)\,di$. The former one is is finite, as shown above. The latter is finite if $\left|\frac{\partial D^{-1}}{\partial q}\right|$ is bounded among large enough values of q and if it does not exceed A/q_i^2 for some constant term A in the limit $q_i \to 0$.

  Note also that the Jacobian is triangular and invertible thanks to the assumptions that the derivative of LHS of equation (92) is strictly negative in U, zero in Λ (this is implied by the budget constraint), and derivative of the LHS of equation (92) is strictly negative in Λ.

- Finally, for utility and Λ to be Frechet differentiable, we need to assume that $\int_{i=0}^{\bar{N}} \int_{q=0}^{q_i} D^{-1}(q,U)dqdi$ and $\int_{i=0}^{\bar{N}} q_i D^{-1}(q_i,U)di$ are Frechet differentiable in q. The derivatives are $\int_{i=0}^{\bar{N}} D^{-1}(q_i,U)h_i di$ and $\int_{i=0}^{\bar{N}} \left(D^{-1}(q_i,U) + q_i \frac{\partial D^{-1}}{\partial q}\right) h_i di$ respectively, for any $h \in L^2$. Hence Frechet differentiability requires that:

$$\int_{i=0}^{\bar{N}} \int_{q_i}^{q_i+h_i} D^{-1}(q,U)dqdi - \int_{i=0}^{\bar{N}} D^{-1}(q_i,U)h_i di \; = \; o(||h||_2)$$

  and

$$\int_{i=0}^{\bar{N}} (q_i + h_i)\left(D^{-1}(q_i + h_i,U) - D^{-1}(q_i,U)\right) di - \int_{i=0}^{\bar{N}} q_i \frac{\partial D^{-1}}{\partial q} h_i di \; = \; o(||h||_2)$$

  as h converges to zero, where $||\cdot||_2$ denotes the L^2 norm.

Proof of Proposition 7

Uniqueness and existence Suppose that demand is Gorman-Pollak with:

$$q \; = \; Q/L \; = \; D(F(\Lambda)p/w)/H(\Lambda)$$

Firms choose $Q = Lw$ to maximize profits:

$$\pi = Q(p-c) - f = Q[wD^{-1}(H(\Lambda)Q/L)/F(\Lambda) - c] - f$$

Assuming a unique quantity level to maximize profits (second order condition), we obtain that optimal profits can be expressed as a function of Λ (after optimizing over p or q). Moreover, we can show that profits strictly decrease with Λ. This can be seen by applying the envelop theorem and noticing that $D^{-1}(H(\Lambda)Q/L)/F(\Lambda)$ must have a negative partial derivative in Λ as one of our assumption for rationalizing Gorman-Pollak demand, conditions [A3]:

$$\frac{\partial \pi}{\partial \Lambda} \; = \; Qw \frac{\partial}{\partial \Lambda} \left\{ D^{-1}(H(\Lambda)Q/L)/F(\Lambda) \right\} \; < \; 0$$

Hence, the zero-profit condition uniquely pins down $\Lambda = \Lambda^*$. From the profit maximization, we also obtain $Q(\Lambda)$ as a function of Λ, hence we obtain a unique firm size $Q^* = Q(\Lambda^*)$. Then, prices as obtained as a function of Q and Λ, so we also have a unique equilibrium price p^*. Finally, the budget constraint (or equivalently the resource constraint) uniquely determines the equilibrium number of firms N^*, given Q^* and Λ^*.

To prove existence, we can see that $QD^{-1}(HQ/L)$ is null in the limit cases where Q is zero and infinite, while $D^{-1}(HQ/L)$ goes to infinity at $Q \approx 0$. Thus, along with the second-order condition, profit maximization leads to a unique and finite $Q^*(\Lambda) > 0$ that maximizes profits, for any given Λ. As we assume that $D^{-1}(H(\Lambda)Q/L)/F(\Lambda)$ spans 0 to $+\infty$ when Λ goes from $+\infty$ to zero for any given Q, we also obtain that variable profits with $Q = Q^*(\Lambda)$ can take all values from 0 to $+\infty$ (for any given c) and thus can equal any value of fixed costs f.

Comparative statics It is again useful to denote:

$$r(q) = qD^{-1}(q)$$

as well as $\rho(q) = \varepsilon_r$ the elasticity of $r(q)$ w.r.t q. The price elasticity of demand (in absolute value) is then $\sigma = \frac{1}{1-\rho}$.

The second-order condition for profit maximization can then be simply stated as r having a negative second derivative, which is equivalent to assuming a negative elasticity of r':

$$Profit\ SOC \iff \varepsilon_{r'} \; < \; 0$$

As shown above, equilibrium conditions on maximized profits being zero determine Λ and Q. First, we can see that:

$$\rho(H(\Lambda)Q/L) \; = \; \frac{cQ}{cQ+f}$$

Hence, differentiating w.r.t Q, Λ and L, we obtain:

$$(1-\rho)d\log Q = \varepsilon_\rho d\log(HQ/L)$$

Note that $\varepsilon_\rho = \varepsilon_{r'} + 1 - \rho$, so this is equivalent to:

$$-\varepsilon_{r'} d\log Q = \varepsilon_\rho \varepsilon_H d\log\Lambda - \varepsilon_\rho d\log L \tag{94}$$

Note that $\varepsilon_{r'} < 0$ (SOC in profits) hence this uniquely characterizes how Q changes depending on Λ and population L.

Next, consider the zero-profit condition:

$$wQD^{-1}(H(\Lambda)Q/L)/F(\Lambda) \ - \ (cQ + f) \ = \ 0$$

Given that Q maximizes profits, the derivative in Q is null. Hence we obtain how a change in population L and income w affects aggregator Λ: Differentiating the zero profit condition yields:

$$[\varepsilon_F + (1-\rho)\varepsilon_H] d\log\Lambda = \ (1-\rho) d\log L + d\log w$$

where again ρ is the elasticity of $r(q) = qD^{-1}(q)$. Rationalization of the demand system and the definition of the aggregator Λ requires that $\varepsilon_F + (1-\rho)\varepsilon_H$ is positive, so we obtain that Λ increases with population L and income w. Plugging into equation (94) describing the changes in firm size Q, and multiplying by $\varepsilon_F + (1-\rho)\varepsilon_H$, we obtain:

$$\begin{aligned}
-\varepsilon_{r'}[\varepsilon_F + (1-\rho)\varepsilon_H] d\log Q \ &= \ \varepsilon_\rho \varepsilon_H [\varepsilon_F + (1-\rho)\varepsilon_H] d\log\Lambda - \varepsilon_\rho[\varepsilon_F + (1-\rho)\varepsilon_H] d\log L \\
&= \ [-\varepsilon_\rho[\varepsilon_F + (1-\rho)\varepsilon_H] + \varepsilon_\rho\varepsilon_H(1-\rho)] d\log L \ + \ \varepsilon_\rho \varepsilon_H d\log w \\
&= \ -\varepsilon_\rho \varepsilon_F d\log L \ + \ \varepsilon_\rho \varepsilon_H d\log w
\end{aligned}$$

So we find that firm size Q increases with L if and only if $\varepsilon_\rho \varepsilon_F$ is negative and increases with income w if and only if $\varepsilon_\rho \varepsilon_H$ is positive. Since the sign of ε_ρ is the same as the sign of ε_σ, we obtain Proposition 7.

Lemma 2

For Propositions 8, 9 and 10, we rely on the following Lemmas linking sub/superconvexity and the sign of $\rho - r/R$, governing whether markups exceed the gains from variety (see Proposition 10).

Denote $r(q, U) = qD^{-1}(q, U)$ and its elasticity $\rho = \varepsilon_r$ w.r.t q as well as:

$$R(q, U) = \int_0^q D^{-1}(q', U) dq'$$

which is well defined according to our assumptions from Section 5.1. Note that we also have $\rho = \varepsilon_r \in (0, 1)$, $R(0, U) = 0$, $r(0, U) = 0$, $r(q, U) > 0$ for any U and $q > 0$. The price elasticity of demand corresponds to $\frac{1}{1-\rho}$ (in absolute terms), hence demand is superconvex if $\varepsilon_\rho > 0$ and subconvex if $\varepsilon_\rho < 0$ (where ε_ρ refers to the elasticity in q).

Lemma 2a: *Suppose $\varepsilon_\rho < 0$ (subconvex demand), then $\rho < r/R$.*

Given the definition of ρ, we have $\rho = xr'/r$ and thus:

$$r/x = r'/\rho \tag{95}$$

hence, since ρ decreases in x ($\varepsilon_\rho < 0$):

$$R(q,U) = \int_0^q [r(x,U)/x]dx = \int_0^q \frac{\partial r}{\partial x}\frac{1}{\rho(x,U)}dx < [\int_0^q \frac{\partial r}{\partial x}dx]/\rho(q,U) = r(q,U)/\rho(q,U)$$

Lemma 2b: *Suppose $\varepsilon_\rho > 0$ (superconvex demand), then $\rho > r/R$.*

We start again from equation (95). Since ρ increases with q, we obtain:

$$R(q,U) = \int_0^q [r(x,U)/x]dx = \int_0^q \frac{\partial r}{\partial x}\frac{1}{\rho(x,U)}dx > [\int_0^q \frac{\partial r}{\partial x}dx]/\rho(q,U) = r(q,U)/\rho(q,U)$$

Now consider the dual, looking here just at demand D with a single argument for the sake of simplicity, and define:

$$S(y) = \int_{y'=y}^{+\infty} D(y')dy'$$

(assumptions on the integrability of D^{-1} around zero implies that the integral above is also well defined). Assuming also that the price elasticity is larger than one, $yD(y)$ has a derivative $yD' + D$ that is negative, and thus $yD(y)$ is decreasing in y. We also assume that it converges to 0 as y goes to infinity (this is the dual equivalent to assuming that $qD^{-1}(q)$ converges to zero at $q \approx 0$). Define μ as:

$$\mu(y) \ = \ \frac{yD'(y)}{yD'(y) + D(y)}$$

As $\mu = \sigma/(\sigma - 1)$, μ increases in y if and only if the price elasticity $\sigma = -\frac{yD'(y)}{D(y)}$ decreases in y. Demand is subconvex if $\varepsilon_\mu > 0$ and superconvex if $\varepsilon_\mu < 0$.

Lemma 2c: *Suppose $\varepsilon_\mu > 0$ (subconvex demand), then $\varepsilon_S > 1 + \varepsilon_D$.*

Note that:

$$\mu(y) - 1 \ = \ -\frac{D(y)}{yD'(y) + D(y)}$$

hence we have $D = (-yD' - D)(\mu - 1)$ and thus, for any $y < y_0$:

$$S(y) - S(y_0) = \int_y^{y_0} D(x)dx = \int_{x=y}^{y_0} (-xD'(x) - D(x))(\mu(x) - 1)\, dx$$

Since $\mu(y)$ increases in y, we obtain:

$$S(y) - S(y_0) \ > \ (\mu(y) - 1)\int_y^{y_0} [-xD' - D]dx \ = \ (\mu(y) - 1)[yD(y) \ - \ y_0 D(y_0)]$$

As y_0 goes to infinity, $S(y_0) = 0$ and $y_0 D(y_0) = 0$, so we obtain:

$$S(y) > yD(y)(\mu(y) - 1)$$

and thus, noticing that $1/(\mu - 1) = \sigma - 1$, we obtain:

$$\varepsilon_S \ = \ -yD(y)/S(y) \ > \ -1/(\mu - 1) \ = \ 1 - \sigma \ = \ 1 + \varepsilon_D$$

Lemma 2d: *Suppose $\varepsilon_\mu < 0$ (superconvex demand), then $\varepsilon_S < 1 + \varepsilon_D$.*

As above, we have

$$S(y) - S(y_0) = \int_y^{y_0} D(x)dx = \int_{x=y}^{y_0} (-xD'(x) - D(x))(\mu(x) - 1)\, dx$$

Hence, since $\mu(y)$ now decreases in y, we have:

$$S(y) - S(y_0) \ < \ (\mu(y) - 1) \int_y^{y_0} [-xD' - D]dx \ = \ (\mu(y) - 1)[yD(y) \ - \ y_0 D(y_0)]$$

As y_0 goes to infinity, $S(y_0) = 0$ and $y_0 D(y_0) = 0$, so we obtain:

$$S(y) \ < \ yD(y)(\mu(y) - 1)$$

and thus, noticing that $1/(\mu - 1) = \sigma - 1$, we obtain:

$$\varepsilon_S \ = \ -yD(y)/S(y) \ < \ -1/(\mu - 1) \ = \ 1 - \sigma \ = \ 1 + \varepsilon_D$$

Combining Lemma 2c and 2d, we can also see that $\varepsilon_\mu(1 + \varepsilon_D - \varepsilon_S) < 0$ with both subconvex and superconvex demand.

With implicitly-additive preferences

With directly-implicitly-additive preferences based on equation (27) in the main text, and with symmetric demand over a continuum of goods, utility satisfies:

$$N \int_0^q D^{-1}(q', U)dq' = 1$$

where D^{-1} is strictly decreasing in q and U. This implicitly determines how utility U depends on N and q.

This leads to prices:

$$p \ = \ w\, D^{-1}(q, U)/\Lambda$$

with Λ determined by the budget constraint:

$$r(q, U)/\Lambda \ = \ 1/N$$

Free-entry condition leads to:

$$r(q, U)/\Lambda \ = \ \frac{cLq + f}{Lw}$$

We get:

$$\int_0^q D^{-1}(q', U)dq' \ = \ \frac{cLq + f}{Lw}$$

From this, we obtain utility U as a function of q, a function that is increasing in q if and only if $\varepsilon_R > \rho$. This is the case for subconvex demand (Lemma 2a).

The optimal pricing condition leads to:

$$\rho(q, U) \;=\; \frac{cLq}{cLq + f}$$

Assuming the profit maximization second order condition, this can be used q as a implicit function of U, decreasing with U if $\rho(q, U)$ decreases with U. Combining with the previous equation, we therefore obtain a unique equilibrium in q and U in the case of subconvex demand if the price elasticity of demand $\sigma(q, U) = 1/(1 - \rho)$ decreases with utility U.[41]

Proof of Proposition 8

Directly semi-separable preferences Suppose that utility is given by: $U = \frac{NR(H(\Lambda)q)}{G(\Lambda)}$ and Λ by:

$$\frac{r(H(\Lambda)q)}{R(H(\Lambda)q)} = \frac{H(\Lambda)G'(\Lambda)}{H'(\Lambda)G(\Lambda)} \equiv \frac{F(\Lambda)H(\Lambda)}{G(\Lambda)}$$

where we denote $F(\Lambda) = \frac{G'(\Lambda)}{H'(\Lambda)}$ which we assume to be positive, and with:

$$R(q) \;=\; \int_0^q D^{-1}(q')dq' \qquad\qquad r(q) \;=\; q\, D^{-1}(q)$$

Also denote by $\rho = \varepsilon_r$ the elasticity of r (which must be between 0 and 1). The first-order condition of profit maximization yields:

$$\rho(H(\Lambda)Q/L) = \frac{cQ}{cQ + f}$$

Assume $r' < 0$ (SOC in profit maximization) and note $\varepsilon_\rho = \varepsilon_{r'} + 1 - \rho$.

Equilibrium in Q and Λ can be summarized by these two conditions:

$$\begin{aligned}
\rho(H(\Lambda)Q/L)\left(1 + \frac{f}{cQ}\right) \;&=\; 1 \\
\frac{r(H(\Lambda)Q/L)}{R(H(\Lambda)Q/L)}\frac{G(\Lambda)}{F(\Lambda)H(\Lambda)} \;&=\; 1.
\end{aligned}$$

The left-hand side of the first equation is decreasing in Q given the profit maximization second order condition: $\varepsilon_\rho + \rho - 1 = \varepsilon_{r'} < 0$, and thus the first equation has a unique solution in Q as a function of Λ and population L:

$$Q \;=\; Q^*(\Lambda, L)$$

with:

$$-\varepsilon_{r'} d\log Q^* \;=\; \varepsilon_\rho \varepsilon_H d\log\Lambda - \varepsilon_\rho d\log L$$

So $Q^*(\Lambda, L)$ increases with Λ if $\varepsilon_\rho \varepsilon_H > 0$; decreases with Λ if $\varepsilon_\rho \varepsilon_H < 0$. Moreover, $Q^*(\Lambda, L)$ increases

[41] Conversely, in the case of superconvex demand, a sufficient condition for uniqueness is that the price elasticity of demand increases with utility.

with L (conditional on Λ) if and only if ε_ρ is negative (subconvex demand).

For the demand system to be well-defined, recall that we assume that the left-hand side of the second equation is decreasing in Λ, i.e. $\varepsilon_F + (1-\rho)\varepsilon_H > 0$. This leads to Λ as a unique solution as a function Q/L.

$$\Lambda = \Lambda^*(Q/L)$$

with:

$$(\varepsilon_F + (1-\rho)\varepsilon_H) d\log\Lambda \ = \ (\varepsilon_r - \varepsilon_R) d\log(Q/L)$$

so $\Lambda^*(Q/L)$ increases in Q/L if and only if $\varepsilon_r - \varepsilon_R = \rho - r/R > 0$.

Suppose that ε_H is negative and that demand is subconvex ($\varepsilon_\rho < 0$). Lemma 2a implies $\rho - r/R < 0$. In this case, we obtain that $Q^*(\Lambda, L)$ strictly increases with Λ while $\Lambda^*(Q/L)$ strictly decreases with Q/L. A solution in (Q, Λ) is thus unique. From Q and the resource constraint, we obtain the number of firms,and we obtain utility from the definition of utility above. The equilibrium is thus unique. Moreover, in this case, $Q^*(\Lambda, L)$ increases with L (while $\Lambda^*(Q/L)$ decreases in Q/L), so an increase in L leads to an upward shift (towards larger firm size) for both curves, and a larger firm size Q in equilibrium.

Suppose that ε_H is negative and that demand is superconvex ($\varepsilon_\rho > 0$). Lemma 2b implies $\rho - r/R > 0$. In this case, we obtain that $Q^*(\Lambda, L)$ strictly decreases with Λ while $\Lambda^*(Q/L)$ strictly increases with Q/L. A solution in (Q, Λ) is thus unique. Again, using the resource constraint and the definition of utility, N and U are also unique.

Thus, the equilibrium is unique when ε_H is negative and demand is either subconvex or superconvex, as the latter implies $\varepsilon_\rho(\varepsilon_r - \varepsilon_R) > 0$. When ε_H is positive, a sufficient condition for uniqueness of equilibrium is obtained by comparing the slopes of Λ^* and Q^* defined just above. In the case of superconvex demand ($\varepsilon_\rho > 0$), both curves are upward-slopping. Plotting both with Q on the vertical axis and Λ on the horizontal axis, curve Q^* is steeper than Λ^* if and only if:

$$\varepsilon_H \varepsilon_\rho (\varepsilon_r - \varepsilon_R) \ > \ (-\varepsilon_{r'})(\varepsilon_F + (1-\rho)\varepsilon_H)$$

With subconvex demand ($\varepsilon_\rho < 0$), both curves are downward slopping and the same condition indicates that the Q^* curve is steeper than Λ^*. A sufficient condition for uniqueness is that the sign of $\varepsilon_H \varepsilon_\rho (\varepsilon_r - \varepsilon_R) - (-\varepsilon_{r'})(\varepsilon_F + (1-\rho)\varepsilon_H)$ does not change in equilibrium.

Finally, assuming that inverse demand goes from zero to infinity as Λ goes from infinity to zero (conditional on its own quantity), the curve $\Lambda = \Lambda^*(Q/L)$ defining Λ necessarily intersect the other curve dictated by profit maximization, $Q^*(\Lambda, L)$. For any given Q, we can then obtain $N = Lw/(cQ + f)$ and then obtain utility U conditional on any N, Q and Λ. Hence an equilibrium exists.

Indirectly semi-separable preferences Suppose that indirect utility can be expressed as:

$$V = \frac{N}{L(\Lambda)} S(F(\Lambda)p/w)$$

where we denote:

$$S(y) = \int_{y'=y}^{\infty} D(y')dy'$$

and where F, L and D are twice continuously differentiable, with $F' > 0$, $L' < 0$, $D > 0$ and $D' < 0$ (assumptions on the integrability of D^{-1} around zero implies that the integral above is also well defined). Suppose that Λ is such that:

$$\frac{(p/w)D(F(\Lambda)p/w)}{S(F(\Lambda)p/w)} = \frac{H(\Lambda)}{L(\Lambda)} \tag{96}$$

where we denote $H(\Lambda) = -L'(\Lambda)/F'(\Lambda)$ (assumed to be positive). We assume that the left-hand side of equation (96) is decreasing in Λ (this assumption is needed to ensure that such indirect utility function is well-behaved, as in Proposition 5), with $\varepsilon_H - \varepsilon_D\varepsilon_F > 0$. We obtain $\Lambda(p/w)$ implicitly defined by this equation, with:

$$(\varepsilon_H - \varepsilon_D\varepsilon_F)d\log\Lambda \;=\; (1+\varepsilon_D-\varepsilon_S)d\log(p/w) \tag{97}$$

Profits are given by:

$$\pi = Q(p-c) - f \;=\; L(p-c)\frac{D(Fp/w)}{VH(\Lambda)} \;-\; f$$

So maximizing profits w.r.t. p leads to:

$$(p-c)/p \;=\; -D(Fp/w)/[(Fp/w)D'(Fp/w)] \;\equiv\; 1/\sigma(Fp/w)$$

or equivalently, using relative markups p/c:

$$\frac{p}{c} \;=\; \frac{(Fp/w)D'(Fp/w)}{(Fp/w)D'(Fp/w)+D(Fp/w)} \;\equiv\; \mu(Fp/w)$$

Note that SOC in profit maximization implies that $\varepsilon_\mu < 1$. Moreover, demand is subconvex if $\varepsilon_\mu < 0$ and superconvex if $\varepsilon_\mu > 0$.

Differentiating w.r.t. p, Λ and w, we obtain:

$$d\log p \;=\; \varepsilon_\mu d\log(p/w) \;+\; \varepsilon_\mu\varepsilon_F d\log\Lambda$$

Hence we get:

$$(1-\varepsilon_\mu)\; d\log p \;=\; -\varepsilon_\mu d\log w \;+\; \varepsilon_\mu\varepsilon_F d\log\Lambda$$

Comparing how p depends on Λ to maximize profits with how Λ depends on p/w (equation 97), we obtain that the equilibrium is unique if one curve is always steeper, i.e. if the sign of:

$$(\varepsilon_H - \varepsilon_D\varepsilon_F)(1-\varepsilon_\mu) \;-\; (1+\varepsilon_D-\varepsilon_S)\varepsilon_\mu\varepsilon_F$$

never flips in equilibrium. On the one hand, the second-order condition in profits and the condition for integrability imply that $(\varepsilon_H - \varepsilon_D\varepsilon_F)(1-\varepsilon_\mu)$ is positive. On the other hand, Lemma 2c and 2d imply that the sign of $(1+\varepsilon_D-\varepsilon_S)\varepsilon_\mu$ is always negative with both the subconvex and superconvex demand. Hence, a sufficient condition for uniqueness is that F increases with Λ.

If demand is subconvex, ε_μ is positive and profit maximization implies that prices increase with income w, conditional on Λ. In turn, since $1+\varepsilon_D-\varepsilon_S$ is negative when demand is subconvex, equation (97) implies that Λ increases with w, conditional on prices (and decreases in prices, conditional on w). Hence equilibrium prices increase with income.

As for direct semi-separability, assuming that demand goes from zero to infinity as Λ goes from infinity to zero (conditional on its own normalized price), we can always find a pair (Λ, p) that satisfy the definition of Λ and profit maximization. Then, for any given price, we obtain firm size Q and thereby obtain $N = Lw/(cQ+f)$. Finally, we obtain utility U conditional on any value of N, Q and

Λ. Hence an equilibrium exists.

Proof of Proposition 9

We adopt the same notation as in Proposition 7. Suppose that preferences are homothetic, as described in Section 4.2, and that U satisfies:

$$NR(qH(\Lambda)/U) \;=\; G(\Lambda)$$

where R is such that:

$$R(x) \;=\; \int_{q=0}^{x} D^{-1}(q)dq$$

This leads to demand proportional to $D^{-1}(qH(\Lambda)/U)$. Demand is then:

$$p/w \;=\; D^{-1}(H(\Lambda)q/U) \;/\; [UF(\Lambda)]$$

with $G' = H'F$. Denote $r(q) = qD^{-1}(q)$. The budget constraint can be written:

$$Nr(qH(\Lambda)/U) \;/\; [H(\Lambda)F(\Lambda)] \;= 1$$

The free-entry condition (zero profits) yields:

$$\frac{Lw}{N} \;=\; cLq + f$$

which we can combine with the budget constraint to obtain:

$$\frac{r(qH(\Lambda)/U)}{F(\Lambda)H(\Lambda)} \;=\; \frac{cLq+f}{Lw} \tag{98}$$

Since the left-hand side is strictly decreasing in Λ, this can be used to implicitly define $\Lambda = \Lambda^*(q,U,L,w)$. Notice that, in equilibrium, q maximizes profits which implies that the derivative of $\Lambda^*(q,U,L,w)$ w.r.t q is null.

Taking differentials, equation (98) yields:

$$[\varepsilon_F + (1-\rho)\varepsilon_H]d\log\Lambda^* \;=\; -\rho d\log U + (1-\rho)d\log L + d\log w$$

Maximizing profits leads to the first-order condition:

$$\rho(H(\Lambda)q/U) \;=\; \frac{cLq}{cLq+f} \tag{99}$$

where $\rho = \varepsilon_r$ is the elasticity of r w.r.t q. The second order condition implies that the derivative in q of the left-hand side is strictly smaller than the RHS. Hence this equality can be used to implicitly define q as a function of U, Λ and L around an equilibrium. Differentiating, we obtain:

$$-\varepsilon_{r'}d\log q^* \;=\; -\varepsilon_\rho d\log U + \varepsilon_\rho\varepsilon_H d\log\Lambda \;-\; (1-\rho)d\log L$$

where $\varepsilon_{r'}$ is strictly negative (profit maximization second order condition).

Finally, we can also combine the free-entry condition with the equation defining utility to obtain:

$$\frac{R(qH(\Lambda)/U)}{G(\Lambda)} \;=\; \frac{cLq+f}{Lw} \tag{100}$$

Note that the derivative of the left-hand-side w.r.t Λ is null if Λ also satisfies equation (98). As the left-hand side is strictly decreasing in U, this can be used to implicitly characterize U as a function U^* of q, Λ and L. Differentiating equation (100) yields:

$$(r/R)d\log U^* \;=\; -(\rho - r/R)d\log q \;+\; (1-\rho)d\log L + d\log w$$

Conditions for uniqueness. Assume for now that L and w remain fixed. Combining with $q = q^*$ and $\Lambda = \Lambda^*$ based on equations (98) and (99), we obtain:

$$d\log U^* \;=\; -(R/r)(\rho - r/R)/(-\varepsilon_{r'})\left[-\varepsilon_\rho d\log U + \varepsilon_\rho \varepsilon_H d\log\Lambda\right]$$

$$=\; -(R/r)(\rho - r/R)/(-\varepsilon_{r'})\left[-\varepsilon_\rho d\log U + \varepsilon_\rho\varepsilon_H\left(-\rho d\log U\right)/(\varepsilon_F + (1-\rho)\varepsilon_H)\right]$$

The equilibrium is unique if the derivative in U in the RHS is always smaller or larger than unity. This happens when the sign of:

$$\Theta_0 \;\equiv\; 1 \;+\; (R/r)(\rho - r/R)/(-\varepsilon_{r'})\left[-\varepsilon_\rho + \varepsilon_\rho\varepsilon_H\left(-\rho\right)/(\varepsilon_F + (1-\rho)\varepsilon_H)\right]$$

is either always positive or negative. Rearranging, we find:

$$sign(\Theta_0) \;=\; sign\left[(-\varepsilon_{r'})(r/R)(\varepsilon_F + (1-\rho)\varepsilon_H) \;-\; \varepsilon_\rho(\rho - r/R)(\varepsilon_F + \varepsilon_H)\right] \tag{101}$$

Note that $\varepsilon_\rho(\rho - r/R)$ is positive when demand is subconvex or superconvex (Lemma 2a and 2b). Moreover, both $(-\varepsilon_{r'})$ and $\varepsilon_F + (1-\rho)\varepsilon_H$ are positive (profit SOC and definition of Λ). Hence the first term is always positive. The second term is also positive if $\varepsilon_F + \varepsilon_H$ is negative, hence $\varepsilon_F + \varepsilon_H < 0$ is a sufficient condition to ensure that Θ_0 is positive and that the equilibrium must be unique.

Now, suppose instead that demand is subconvex ($\varepsilon_\rho < 0$) and that ε_H is negative. We obtain:

$$sign(\Theta_0) \;=\; sign\left[(1-\rho)(r/R)(\varepsilon_F + (1-\rho)\varepsilon_H) \;-\; \varepsilon_\rho\rho(\varepsilon_F + (1-\rho)\varepsilon_H) \;-\; \varepsilon_\rho(\rho - r/R)\rho\varepsilon_H\right]$$

The first term is positive given that $\rho < 1$ and $\varepsilon_F + (1-\rho)\varepsilon_H > 0$ (to ensure a well-defined aggregator Λ). The second term is positive if demand is subconvex. The third term is also positive if we assume $\varepsilon_H < 0$, given that $\varepsilon_\rho(\rho - r/R)$ is positive in the subconvex case (Lemma 2a). Hence, $\varepsilon_H < 0$ is a sufficient condition for uniqueness of equilibrium in the subconvex case.

Comparative statics. If we now examine how changes in population L and w influence equilibrium outcomes, we can again use equations (98), (99) and (100). Incorporating $q = q^*$ into the equation defining utility, we obtain:

$$(r/R)d\log U^* \;=\; -(\rho - r/R)/(-\varepsilon_{r'})\left[-\varepsilon_\rho d\log U + \varepsilon_\rho\varepsilon_H d\log\Lambda \;-\; (1-\rho)d\log L\right] + (1-\rho)d\log L + d\log w.$$

Incorporating $\Lambda = \Lambda^*$ and rearranging, we obtain:

$$d\log(LU^*) \;=\; \Theta_U d\log(Lw)$$

with

$$\Theta_U \;=\; \left[1 - \frac{(\rho - r/R)\varepsilon_\rho \varepsilon_H}{(-\varepsilon_{r'})[\varepsilon_F + (1-\rho)\varepsilon_H]}\right] / \Theta_0$$

Using equation (98), differentiated:

$$[\varepsilon_F + (1-\rho)\varepsilon_H] d\log\Lambda^* \;=\; -\rho d\log(LU) + d\log(Lw)$$

we obtain the changes in aggregator Λ:

$$d\log\Lambda^* \;=\; \Theta_\Lambda d\log(Lw)$$

with

$$\Theta_\Lambda \;=\; \frac{1 - \rho\Theta_U}{\varepsilon_F + (1-\rho)\varepsilon_H}$$

Finally, thanks to equation (99):

$$-\varepsilon_{r'} d\log(Lq^*) \;=\; -\varepsilon_\rho d\log(LU) + \varepsilon_\rho \varepsilon_H d\log\Lambda$$

we obtain the changes in firm size $Q = Lq$:

$$d\log(Lq^*) \;=\; \Theta_Q d\log(Lw)$$

with

$$\Theta_Q \;=\; \frac{-\varepsilon_\rho \Theta_U + \varepsilon_\rho \varepsilon_H \Theta_\Lambda}{-\varepsilon_{r'}}$$

After much simplifying, we obtain:

$$\Big[-\varepsilon_\rho(\rho - r/R)[\varepsilon_F + \varepsilon_H] - \varepsilon_{r'}(r/R)[\varepsilon_F + (1-\rho)\varepsilon_H]\Big] d\log Q$$

$$= \; -\varepsilon_\rho \Big[[\varepsilon_F + (1-\rho)\varepsilon_H] \; + \; (\rho - r/R)\varepsilon_H \Big] d\log(Lw)$$

As shown in equation (101) and below, the coefficient in front of $d\log Q$ is positive when demand is subconvex and ε_H is negative. In that case, we can also see that the coefficient in front of Lw is also positive, since $-\varepsilon_\rho > 0$, $\varepsilon_F + (1-\rho)\varepsilon_H > 0$ and $\rho - r/R < 0$ (Lemma 2a).

In the superconvex case $\varepsilon_\rho > 0$ with $\varepsilon_F + \varepsilon_H < 0$, the effect of market size on firm size is positive if and only if:

$$\varepsilon_F + \varepsilon_H < r/R\varepsilon_H$$

This does not necessary hold, even under the sufficient assumption $\varepsilon_F + \varepsilon_H < 0$ for uniqueness of equilibrium.

Homothetic semi-separability. Item iii) of Proposition 9 is a direct corrolary of proposition 8. When preferences are homothetic, firm size depends on just Lw. If preferences are also directly semi-separable, we know that firm size does not vary with w, hence it does not also vary with w.

Also, as we noted earlier in Section 4, homothetic direct semi-separability is equivalent to homothetic indirect semi-separability.

Proof of Proposition 10

Gains from variety Suppose that U satisfies:

$$NR(qH(\Lambda, U), U) \ = \ G(\Lambda, U)$$

with $R(x, U) = \int_0^x D^{-1}(x', U)dx'$. Denote $r(x) = xD^{-1}(x)$. The budget constraint is then:

$$Nr(qH(\Lambda, U), U)/H(\Lambda, U) \ = \ G'(\Lambda, U)/H'(\Lambda, U) \ \equiv \ F(\Lambda, U)$$

with $G' = H'F$ and where the ́denotes the derivative w.r.t Λ.

Assuming identical demand across goods, what are the benefits of increased consumption q per good relative to increased number of product varieties? This is obtained by how much we need to change q and N jointly to keep utility constant (slope of indifference curves in terms of q and N). Differentiating the first equation, given that the derivative in Λ is null, we obtain:

$$d\log N + \varepsilon_R d\log q \ = \ \beta dU$$

with $\varepsilon_R = r/R$, for some $\beta > 0$, and thus we obtain:

$$\frac{\frac{\partial U}{\partial \log N}}{\frac{\partial U}{\partial \log q}} \ = \ 1/\varepsilon_R \ \equiv \ v$$

which is the inverse of the elasticity of $R(x, U) = \int_0^x D^{-1}(x', U)dx'$, evaluated at $x = H(\Lambda, U)q$ (with $q = Q/L$ is individual consumption).

First best entry and production Suppose that total resources are fixed (total costs are given by total GDP Lw), so that:

$$Nf \ + \ NcLq \ = \ Lw$$

Differentiating (holding Lw constant), we obtain:

$$(f + cLq)\, dN \ + \ NcL\, dq \ = \ 0$$

Thus, at optimum, maximizing utility U as function of q and N under that constraint, the FOC of the first-best allocation implies:

$$\frac{\frac{\partial U}{\partial N}}{\frac{\partial U}{\partial q}} \ = \ \frac{cLq + f}{NcL}$$

which is equivalent to:

$$\frac{\frac{\partial U}{\partial \log N}}{\frac{\partial U}{\partial \log q}} \ = \ \frac{cLq + f}{cLq}$$

Hence, at the first best allocation we must have:

$$\frac{cLq + f}{cLq} \ = \ v$$

where $v = R/r = 1/\varepsilon_R$ refers to the gains from variety.

Market equilibrium

In the market equilibrium, firms follow the pricing rule:

$$\rho(H(\Lambda, U)q\,,\,U) \;=\; \frac{cLq}{cLq+f}$$

where $\rho = \varepsilon_r = (\sigma-1)/\sigma$ is the elasticity of r w.r.t q, and where σ refers to the price elasticity of demand. In addition, the zero profit condition is equivalent the resource constraint $N(f+cLq) = Lw$ as above.

Hence the market equilibrium is not the first-best allocation as long as ρ differs from ε_R, or equivalently $\sigma/(\sigma-1) = 1/\rho$ differs from $v = 1/\varepsilon_R$.

Moreover, if demand is subconvex, Lemma 2a implies that $\rho < \varepsilon_R$ and entry is inefficiently large:

$$\frac{cLq+f}{cLq} < \frac{\frac{\partial U}{\partial \log N}}{\frac{\partial U}{\partial \log q}}$$

(marginal utility from increased q relative to marginal utility from increased entry N is too high).

If demand is superconvex, Lemma 2b implies that $\rho > \varepsilon_R$ and entry is inefficiently low:

$$\frac{cLq+f}{cLq} > \frac{\frac{\partial U}{\partial \log N}}{\frac{\partial U}{\partial \log q}}$$