CONSUMER DEMAND WITH PRICE AGGREGATORS AND LOW-RANK CROSS-PRICE EFFECTS

THIBAULT FALLY AND ETHAN LIGON

Abstract. Estimating consumer demands is a bread-and-butter undertaking in applied economics. In general, demand for each good depends on the prices of all goods and services, but for most applications it is impractical to estimate models of such high dimension. In this paper, we consider consumer demand with a low rank of the matrix of cross-price effects, a property implicitly assumed in most empirical settings. First, we show that imposing a low rank is equivalent to introducing functions that we call "aggregators", where each aggregator maps information from an arbitrarily large vector of prices (and perhaps income) into a scalar. We then provide a complete characterization of the preferences that rationalize demand systems with such aggregators. These results can be used to derive new and flexible forms of demand that can be tailored to applications in various fields of economics. Most commonly-used demand systems (including directly-additive, indirectly-additive, non-homothetic CES and Kimball preferences) can be described with one or two aggregators where the price index may coincide with one of the aggregators. Nested and mixed logit require as many aggregators as nests or consumer types. Aggregators can also be naturally expressed as a function of observed product attributes. Using barcode data on purchases of ready-to-eat cereals, we illustrate how to estimate a simple yet flexible specification of such a demand system with K aggregators, with or without using information on product attributes.

Thibault Fally	Ethan Ligon
fally@berkeley.edu	ligon@berkeley.edu
University of California	University of California
Berkeley, CA 94720–3310	Berkeley, CA 94720–3310

Date: March 27, 2026.
Keywords. Preferences; Cross-price effects; Generalized Separability; Latent variables; Low rank.
This remains very much a work in progress—please do not distribute without consulting us first.
Many thanks to Scott Kjorlien and Cassandra Turk for excellent research assistance, Nano Barahona, Kirill Borusyak, Pablo Fajgelbaum, Scott French, Joonkyo Hong, Eunhee Lee, Monika Mrazova, Jakob Munch, Mathieu Parenti, Filip Premik, Frederic Robert-Nicoud, Philip Ushchev and seminar participants at UC Berkeley, HKU, U Geneva, Columbia U, ULB, LSU, UCLA, NUS, Monash U (METC Conf), UNSW, PSU (Midwest Trade Conf) and SNU (KETC Conference).

1. Introduction

Estimating consumer demands is a bread-and-butter undertaking in applied economics. Any such undertaking must somehow contend with the fact that, in general, demand for any good will depend on the prices of all goods and services. However, it is highly impractical to estimate models with such high-dimensional price interactions, and so researchers invariably reduce dimension, often implicitly via a choice of functional forms. Examples include nested and mixed logit which extend the basic logit specification in order to generate more flexible cross-price effects by grouping goods or combining heterogeneous types of consumers.[1]

These functional form assumptions lead to strong restrictions in terms of income effects (typically assuming quasi-linearity or homotheticity) and cross-price effects (e.g. by relying on predetermined groups of goods or assumptions on the patterns of consumer heterogeneity). Of course, sometimes these limitations are features: we may need to restrict price and income effects to be able to construct a representative consumer, or we lack the data to estimate important patterns of substitution, but we may have lost an appreciation for the critical limitations of these functional forms and we lack a unifying perspective.

Our goal is to provide a disciplined approach to introduce more or less flexibility in functional forms of demand, depending on the data available and the needs of the researcher. In this paper we provide such an approach through the introduction of functions we call "aggregators". An aggregator maps information from an arbitrarily large set of prices (and perhaps income) into a scalar, thus summarizing information on many prices. A price index is a familiar example of an aggregator, but aggregators need not satisfy the usual requirements of a price index (for example, aggregators need not be homogeneous in prices). In a first step, we show that the number of aggregators coincides with the rank of the matrix of cross-price effects, adjusting for own-price effects. This number provides a key metric that captures the complexity of a demand system. For instance, when the rank of cross-price effects is one, all cross-price effects can be captured by a single aggregator and demand for each good can then be expressed as a function of its own price and this aggregator.

This notion of rank of cross-price effects is also tightly linked to the complexity of the estimation problem and the number of parameters to identify, akin to low-rank approximations in machine learning. With a general demand system without rank restrictions, the complexity of cross-price effects grows as a quadratic function of the number of goods, and the estimation of cross-price effects becomes impossible with datasets covering a large number of goods or product varieties. By imposing a parameterization with low-rank cross-price effects, estimation with large datasets remains manageable as the number of parameters to be estimated grows at most linearly with the number of goods.

In a second step, we further assume that demand is rational, i.e., derived from the maximization of a utility function. We establish that rational demands with K aggregators must then take particular functional forms, and we show what these forms must be. In parallel, we show how integrating these demands gives rise to

[1] See, among others, Boyd and Mellman (1980), Berry (1994), Berry, Levinsohn, and Pakes (1995), D. McFadden and Train (2000).

a rationalizing utility function, and thus provide a complete characterization of the preferences which rationalize a demand system having K aggregators. What we call K-aggregator preferences and demands nest many of the preferences and demand structures that exist in the literature, but also include novel forms. As a practical example, we provide a specification that allows for flexible non-parametric own-price effects combined with parametric cross-price effects captured by a symmetric matrix of any given rank K.

We begin with the homothetic case, allowing us to focus on cross-price effects, and subsequently consider Hicksian demand as a function of both prices and utility, which remains homogeneous of degree one in prices. With utility as an additional aggregator, we can extend our results for homothetic demand to construct demand with more flexible income effects for each good.

We further show how our demand system can be represented as a "perturbed utility model" (PUM), treating consumer decisions as random realizations where probabilities are chosen to maximize a perturbed utility function that is linear in log prices. This representation connects our framework to the literature on stochastic choice and random utility models (RUM), while being more general. We also discuss how our demand system relates to those commonly used in the literature. In particular, mixed logit can be interpreted as demand with price aggregators, where we have at most one aggregator per consumer type; nested logit similarly requires as many aggregators as nests. The more recent inverse product differentiation logit (IPDL) of Fosgerau, Monardo, and De Palma (2024) can also be obtained as a special case of our framework. As with IPDL, a key advantage of our approach is that cross-price effects enter linearly once we invert the own-price demand curve, which avoids the computationally intensive nonlinear inversion required by BLP (Berry, Levinsohn, and Pakes 1995) and leads to more transparent identification. Moreover, unlike additive random utility models, our specification allows for complementarity between goods, providing greater flexibility in substitution patterns.

We estimate a special case of our demand system where cross-price effects are parameterized by a symmetric matrix (of rank K to be estimated), which we further project onto product attributes. As with BLP, we further assume that substitution patterns are driven by observable attributes, which reduces the problem to estimating an $M \times M$ matrix, where M denotes the number of product attributes. We consider the instruments proposed by Gandhi and Houde (2019) and also introduce a new set of instruments based on interaction terms between attributes and some weighted average prices of all goods in a market. We apply our approach to estimating demand for ready-to-eat cereals using NielsenIQ data for the 2010–2019 period (monthly), and find that the data support a substitution matrix of rank $K = 2$. Adding demographic taste shifters on product characteristics proves essential for identification.

We estimate both our model and a BLP random-coefficients logit on the same data with matched instruments and demographic controls. A Rivers-Vuong test decisively favours the K-aggregator specification, though neither model fully encompasses the other. Adding demographic taste shifters proves essential for both models, stabilizing price-elasticity estimates and sharpening the eigenvalue separation in the substitution matrix. Strikingly, despite incorporating random coefficients designed to enrich substitution patterns beyond logit, the BLP estimates yield a cross-price substitution matrix that is effectively rank one across all the instrument sets that we considered. In contrast, our K-aggregator estimation identifies richer substitution patterns, with two dimensions of substitution with eigenvalues of comparable magnitude.

Literature. This paper primarily aims to contribute to the literature on modeling cross-price effects, which has a long tradition, not only in industrial organization (e.g. to understand the effects of competition) but also in other fields such as macroeconomics (e.g. cost-of-living estimation), international trade (e.g. in gravity equations), and development (understanding household consumption choices).

Earlier models of demand which allow for flexible cross-price effects include the Rotterdam model, Translog, PIGL and PIGLOG (including AIDS, developed by Deaton and Muellbauer 1980). In most of these specifications, prices enter linearly or log-linearly, but are only valid demand systems at best in a local sense, not over the full range of prices and income. In this vein, EASI (Lewbel and Pendakur 2009) may be the latest and most flexible. For any specific good, it allows for flexible Engel curves along with a general matrix of cross-price effects. But, as in all these demand specifications with flexible cross-price effects, the form of own-price effects is very restrictive (log linear) and does not allow for, e.g., flexible pass-through and variable elasticities or semi-elasticities.

A simple but powerful way to generate non-trivial cross-price effects is to construct nests, allowing for different elasticities of substitution between vs. within different nests. Nested logit has been widely used in industrial economics (see e.g. Berry 1994) while nested CES is a standard specification in macroeconomics and international trade (see e.g. Broda and Weinstein 2006). These types of preferences generate low-rank cross-price effects where the rank corresponds to the number of nests. However, for estimation, these nests have to be specified ex ante, and impose a lot of structure on substitution patterns. A recent class of demand proposed by Fosgerau, Monardo, and De Palma (2024), "inverse product-differentiation logit", proposes a more flexible way to construct nests without relying on hierarchies. Our approach here provides a generalization of nests while keeping the low rank of price substitution patterns.

The now standard specification in industrial organization is that of "BLP" (Berry, Levinsohn, and Pakes 1995), itself a form of mixed logit (e.g. McFadden and Train 2000). While each consumer is typically assumed to have logit demand, thus with simple rank-one cross price effects, the BLP approach obtains more complex substitution patterns for aggregate demand, where the rank of cross-price effects can be as large as the number of types of consumers that are lumped together. In practice, the BLP approach involves adopting statistical assumptions regarding the distribution of consumer values. It allows for complicated cross-price elasticities, but these are typically generated by ad hoc assumptions on heterogeneity of price coefficients and attributes in logistic models. It is also difficult to link the patterns of cross-price effects to the distributional parameters to be identified in these models. In particular, BLP estimators involve a non-linear inversion of expenditure shares to express those as a linear function of prices and average valuation of product attributes. This non-linear inversion must account for parameters governing the heterogeneity in tastes, and thus cross-price effect terms. In comparison, our specification also involves an inversion but only with respect to own-price effects. In our baseline specification, cross-price effects enter linearly.

Non-trivial income effects can also be incorporated in our demand systems with low-rank cross-price effects. A first approach is to consider Hicksian demand and the expenditure function, which remains homogeneous in prices, conditional on utility. Thus, using utility as an additional aggregator, we are able to incorporate very flexible income effects, and their interaction with prices. This approach can be used for instance to generalize demand as in EASI (Lewbel and Pendakur 2009), or AIDS (Deaton and Muellbauer 1980), and obtain specifications that are globally regular. Another approach involves expressing aggregators as a function of normalized prices (prices relative to income) instead of prices. We obtain results that are similar (up to multiplicative terms) to the homothetic case. Note that our concept of rank sharply differs from the notion of income rank introduced by Lewbel (1991), with higher rank demand systems admitting more complicated Engel curves. Our notion of rank focuses on cross-price effects, and demand system with K aggregators (hence rank K) can have any rank in terms of income effects.[2]

In applied theory, it is often useful to reduce the dimensionality of cross-price effects in order to improve tractability, while maintaining the assumption of consumer rationality. With price aggregators, the characterization of equilibrium is reduced to examining a few variables instead of potentially many prices. Focusing on a few aggregators also speeds up numerical solutions for general equilibrium models. Demand with aggregators also simplifies the analysis of industry equilibrium and interactions between firms under imperfect competition: firms under monopolistic competition take such aggregators as given, and firms under Bertrand competition account for how their choices influence such aggregators.[3]

Beyond their applications to economics, these results can be useful for machine learning. Assuming a low rank is useful for various applications, such as principal component analysis, signal processing and matrix completion (e.g. "Netflix problem"), image compression, word embeddings and Latent Semantic Analysis (LSA), data imputation, etc. Proposition 1 characterizes any function with low-rank interactions between variables, while Proposition 2 can be used to construct homogeneous functions with a specific rank of the Hessian after removing diagonal terms. Our empirical approach has similarities with the estimation of latent factor models (e.g., Bai and Ng 2002), but imposes additional structure on how factors (i.e. aggregators) and factor loadings are related.

2. Preamble: aggregators and the rank of cross-price effects

Consider a setting where we have goods indexed by $i = 1, \dots, J$, and demand $F_i(x)$ for good i that depends on its own price x_i, and potentially also on all other prices

[2]Our definition of rank and aggregators also differs from the notion of "latent separability" introduced by Blundell and Robin (2000). We focus on the rank of demand while Blundell and Robin (2000)'s concept relates more closely to the rank of income effects and homothetic aggregators used as part of utility and expenditure functions.

[3]Under Cournot competition it may then be easier to work with inverse demand and express aggregators as functions of quantities, a formulation we adopt later in the paper.

or inputs $x_1, \ldots, x_J$.[4] The Jacobian of F with respect to all prices is then a $J \times J$ matrix, which can be decomposed into a diagonal matrix of own-price effects σ, and a matrix of cross-price effects Σ:

$$D_x F(x) = \sigma(x) + \Sigma(x). \tag{1}$$

Under general assumptions, the matrix of cross-price effects Σ can have any rank up to J. Empirically, this implies that the number of cross-price elasticities to identify grows quadratically with the number of goods. As typical consumer data often includes more goods than markets or time periods (e.g. with barcode-level data), identification of fully-flexible cross-price effects becomes impossible.

A common approach in applied statistics and machine learning is to approximate these interactions by imposing a lower rank (any matrix can be approximated by a lower-rank matrix) and thereby reducing the dimensions of the estimation problem. Here, the low-rank approximation would only apply to the matrix of cross-price effects Σ, as the diagonal term σ typically needs to retain full rank in standard settings, where a usual assumption is that demand for each good depends negatively on the price of that good.

We obtain a low-rank Σ when demand for a good i depends on its own price x_i and a few scalar functions $\Lambda_k(x)$, $k = 1, ..., K$ with $K \leq J$, that summarize the effects of other prices. We call these functions "aggregators". If demand $F_i(x)$ for good i coincides with a function $S_i\left(x_i\,, \Lambda_1(x)\,, ..., \Lambda_K(x)\right)$, we can see that the matrix of cross-price effects can be expressed as a product of two smaller matrices:

$$\frac{\partial F_i}{\partial x_j} = \sum_k \frac{\partial S_i}{\partial \Lambda_k} \frac{\partial \Lambda_k}{\partial x_j} \quad \text{for } i \neq j$$

Since the matrices with coefficients $\frac{\partial S_i}{\partial \Lambda_k}$ and $\frac{\partial \Lambda_k}{\partial x_j}$ have K columns or rows, where K is the number of such aggregators, the resulting matrix of cross-price effects has at most a rank K.

In Proposition 1, we show that the connection between the rank and the number of aggregators is even stronger. In particular, we show that the converse holds under some additional regularity conditions: if cross-price effects have a low rank K, the corresponding demand functions must depend on some K aggregator functions in addition to their own price.[5] This result applies to smooth functions with rank and other regularity conditions (see Appendix A for details), but we do not impose rationality of demand at this stage. This is not yet our main theoretical result but a key motivation for our analysis of demand with price aggregators:

[4]More generally, we could consider any function where the i^{th} component $F_i(x)$ depends monotonically on the i^{th} component of variable x as well as other components potentially. Here we refer to goods and prices for convenience.

[5]When own price effects are null, this result is a simple application of the "Rank Theorem" in the theory of smooth manifolds: If we consider a function with a Jacobian of constant rank K, its image corresponds to a K-dimensional embedded manifold. Here the Rank Theorem does not apply and the difficulty arises from the presence of own-price effects, so that the Jacobian has full rank. Hence the image does not provide a simple way to define the K aggregators.

Proposition 1. *Under assumptions [C1], [C1'], [C2], [C3] and [C4], there exist K real functions* $\Lambda_1(x)$, *...* , $\Lambda_K(x)$ *and J functions* $S_i(\Lambda, x_i)$ *such that:*

$$F_i(x) \; = \; S_i\left(x_i\,,\, \Lambda_1(x)\,, ...,\, \Lambda_K(x)\right)$$

e.g., demand for each good can be reduced to a function of its own price and K "aggregators" $\Lambda_k(x)$.

The proof relies on tools from the theory of smooth manifolds. We defer the proof to the Appendix, and here we provide a less formal description of the conditions and their role in the proof of Proposition 1:

- C1 **(Rank)** At any price x, we assume that matrix $\frac{\partial F_i}{\partial x_j}(x)$ is the sum of a diagonal matrix $\sigma(x)$ with diagonal elements $\sigma_i(x)$ and a matrix $\Sigma(x)$ of rank K. The diagonal matrix σ captures own-price effects. Matrix Σ, which we refer to as the substitution matrix, then captures substitution patterns across goods, conditional on their own prices. Thus, for convenience, we rule out situations where the rank varies over the price space.
- C1' **(Identifying goods)** Furthermore, we assume that there are K goods, i.e. K columns or rows for which the substitution matrix Σ (which varies with x) are always independent. This facilitates the construction of aggregators. If we remove this condition, aggregators remain K dimensional but must be defined in a more abstract space (see Corollary 1 in Appendix). Thanks to this assumption (along with C4 below), we can write demand for each good $j > K$ as a function of its own price x_j, the price of the first K goods, and the demand for the first K goods. Hence, demand is reduced to a function of $2K + 1$ arguments. With such assumption, we rule out situations where some of these K reference goods are no longer purchased at high prices.
- C2 **(Stability of the rank)** We assume that the rank of the "substitution matrix" is preserved when we drop some goods and when we condition on some prices. The exact equivalence between the rank of the substitution matrix and the number of aggregators might be broken without such an assumption. Specifically, we need this condition to ensure that some aggregators are not just capturing the price of a single good. We exploit this assumption to obtain that the own-price elasticity σ_k of the K reference goods can be expressed as a function of the prices and demands for the K reference goods.
- C3 **(No escaping)** We assume that if the demand for some of the K reference goods diverges, then demand for some other goods also diverges. Under such conditions, we can use the solutions of some autonomous differential equations to construct some transformations and reduce the $2K + 1$ arguments into just $K + 1$ arguments (including the own price).
- C4 **(Connectedness)** We assume that level sets for the K "identifying goods" are connected (in a topological sense), even if we condition on their K prices. This ensures some form of monotonicity and allows us to construct aggregators implicitly, using Lemma 1 from Goldman and Uzawa (1964, see Appendix A).

Note that these assumptions on the stability of the rank imply that the matrix of cross-price effects has a "low" rank in a statistical sense; i.e., its rank K is smaller than $J/2$.[6] Note also that, locally, assumption [C1'] is implied by [C1], hence it is not a strong assumption for the existence of aggregators. Globally, if [C1'] is not satisfied, we can define aggregators in a K-dimensional manifold instead of the K-dimensional Euclidian space. We provide results in this more general case in Corollary 1 in Appendix.

Thus, Proposition 1 provides mathematical foundations to justify the use of "price aggregators" when we want to impose low-rank cross-price effects. Next, we combine these tools with assumptions on rational behavior that are standard in consumer theory.

3. Rational demand with price aggregators: homothetic case

We now examine the implications of rationality; i.e., that consumer demand is derived from utility maximization (under a standard budget constraint). As a well-known result in microeconomics, it implies that (compensated) cross-price effects are symmetric ("Slutsky symmetry"), an additional restriction on the Jacobian. We maintain the assumption that demand depends on its own price, as well as a given number K of aggregators (themselves functions of all other prices). For the ease of exposition, in this section we start with the case of homothetic demand, and move to more general cases in the next section.

In the first subsection below, we precisely lay out hypotheses in addition to other topology assumptions, and in Proposition 2 we provide the functional forms of all demand functions that satisfy these properties. In addition, we provide a form of utility that yields such demand systems, assuming rational behavior under a standard budget constraint. This first set of results can be understood as a set of necessary conditions that demand must satisfy in order to depend on K aggregators and be derived from a utility function.

These necessary conditions however are not sufficient to ensure that such a utility function is quasi-concave (or, equivalently, obtained from a concave price index). We address this issue by providing (mild) additional sufficient conditions. We examine properties of the price substitution matrix in these cases.

These results provide practical ways to construct utility and demand with low-rank cross-price effects. As an illustration, we then provide an example of a very tractable form of demand that allows for flexible own price effects as well as cross-price effects parameterized by arbitrary positive semi-definite symmetric matrices, with a chosen rank K corresponding to the number of aggregators.

3.1. Set up and functional form in the homothetic case.

Under homothetic preferences, demand q_i for each good $i \in \{1, ..., J\}$ is proportional to income w, hence expenditure shares $S_i \equiv \frac{p_i q_i}{w}$ only depend on prices p. More specifically, our goal here is to describe expenditure shares that depend on their own prices, respectively, and a vector of K price aggregators. This implies a rank K of cross-price effects, as discussed previously. We denote these aggregators by $\Lambda = (\Lambda_1, ..., \Lambda_K) \in \mathbb{R}^K$ (using

[6] It also implies that the decomposition into a diagonal matrix and a low-rank matrix is unique. Uniqueness can be shown by applying the results of Chandrasekaran et al. (2009).

subscript k to refer to an aggregator), Hence the expenditure share on good i is expressed as:

$$\frac{p_i q_i}{w} = S_i(p_i, \Lambda_1(p), ..., \Lambda_K(p)) \tag{2}$$

where p refers to the full vector of prices, and p_i is the price of good i. We assume that the number of goods is larger than the number of aggregators (specifically, $J > K+3$), as our analysis will be particularly useful in settings where the number of goods is very large and where we want to reduce the dimensionality of interactions between goods. We consider only one period, with a balanced budget:

$$\sum_i S_i(p_i, \Lambda(p)) = 1 \tag{3}$$

Rationality. We focus on demand from a rational consumer who is maximizing a quasi-concave utility U. This is equivalent to assuming the existence of a price index $P(p)$ that is homogeneous of degree one and concave in prices p. It is also the same as imposing that the expenditure function is proportional to $P(p)$ (i.e. multiplicatively separable in utility U and prices p). Shephard's Lemma implies that the expenditure share on good i must equal the derivative of $\log P$ w.r.t $\log p_i$, hence:

$$\frac{\partial \log P(p)}{\partial \log p_i} = S_i(p_i, \Lambda(p)) \tag{4}$$

For most of the analysis, we assume that demand (and utility) is smooth.

For the first proposition, we also make the following assumptions on price effects and topology:

Additional assumptions:

A1. Own price elasticity: *for each good, the own price effect is negative, i.e:* $\frac{\partial S_i}{\partial p_i}(p_i, \Lambda) < 0$*, holding* Λ *constant, evaluated at any* p *and* Λ*.*

A2. *For any* Λ *and* $y > 0$*, there exists a real* $t > 0$ *such that:* $\sum_i S_i(t, \Lambda) = y$*.*

A3. Rank of ∂S: *the matrix with coefficients* $\left\{\frac{\partial S_i}{\partial \Lambda_k}\right\}$ *has full rank* K*, where* K *denotes the number of aggregators.*

A4. Rank of $\partial\Lambda$: *the matrix with coefficients* $\left\{\frac{\partial \Lambda_k}{\partial \log p_i}\right\}$ *has maximal rank* K*, even if we drop one good* i *from the set of goods.*

A5. Connectedness. *The level sets of* Λ*,* $\left\{p \in \mathbb{R}^J_{++} \mid \Lambda(p) = \Lambda_0\right\}$*, are connected, for any* $\Lambda_0 \in \mathbb{R}^K$*.*

A6. No escaping: $\exists p \in \mathbb{R}^J$ *such that* $\max_i |\log S_i(p_i, \Lambda^{(t)})|$ *goes to infinity for any sequence of* $\Lambda^{(t)} \in \mathbb{R}^K$ *that escapes any compact set (i.e. is unbounded).*

Assumption 1 might be reversed, i.e. it may be possible to have all own price effects be positive, but we cannot have a mix of positive and negative own-price effects. The case of positive own price effects is considered for instance in Matsuyama and Ushchev (2017) for the special case of homothetic preferences with a single aggregator.

With the two rank assumptions (A3 and A4) we assume that aggregators capture different types of information relevant to consumers. With those, we assume that K is the minimum number of aggregators needed to explain demand. If the gradient of an aggregator was colinear with the gradient of other aggregators, we could then express this aggregator as a function of the other one, thereby reducing the number of aggregators.

Then, two topology assumptions put restrictions on the space of aggregators. The connectedness assumption A5 can be interpreted as a monotonicity assumption. It implies that if there are two sets of conditions in p that are associated with the same values of aggregators Λ, there is a continuous path indexed by $t \in [0,1]$ of intermediate conditions $p(t)$ from one to the other with the same aggregator values $\Lambda(p(t)) = \Lambda$. On the contrary, topology assumption A6 implies that, when some aggregators diverge, there are some relative expenditures that also diverge (for some reference price level). Also, given the rank and topology assumptions, up to a change in variables, we can assume that $\Lambda(p)$ spans all $\mathbb{R}^K$.[7]

We will use Lemma 1 of Goldman and Uzawa (1964) repeatedly: if the gradient of a real function f (defined on a Euclidean space) is colinear with the gradients of other real functions g_1, ..., g_n, and if the level sets of g are connected, we can express f as a function of $(g_1, ..., g_n)$. Assumption A5 on connectedness is useful for this.

We can now move onto the first main result on the functional form of homothetic demand:

Proposition 2. *Homothetic demand that depends on aggregators Λ and satisfies all assumptions A1-A6 above must take the form:*

$$S_j(p_j, \Lambda) \;=\; D_j(p_j/\lambda, \Lambda') \tag{5}$$

where $\Lambda = \Phi(\lambda, \Lambda')$ and Φ is a one-to-one re-mapping from aggregators $\Lambda \in \mathbb{R}^K$ to aggregators $\lambda \in \mathbb{R}$ and $\Lambda' \in \left\{\Lambda \;\middle|\; \sum_i S_i(1,\Lambda) = 1\right\}$ (a submanifold of dimension $K-1$). Moreover, it is derived from price index $P(p)$ that satisfies:

$$\log P(p) \;=\; \log \lambda \;-\; G(\Lambda') \;+\; \sum_j \int_{t=1}^{p_j/\lambda} D_j(t, \Lambda')\, d\log t \tag{6}$$

for some real function $G(\Lambda')$, and where the aggregators (λ, Λ') are such that the partial derivatives of the RHS in (λ, Λ') are null. Aggregator λ is then homogeneous of degree one in p while Λ' is homogeneous of degree zero.

Not every function S_i of prices and aggregators Λ coincides with a rational demand system: a rational demand system with K aggregators Λ must take the form described above, with a utility characterized by (6). In particular, there must be one aggregator (denoted λ) that plays a special role. We can think of this aggregator as adjusting in order to obtain a balanced budget: the first order condition in λ in the RHS of (6) is equivalent to imposing $\sum_i D_i = 1$. In the special case of directly-additive preferences (see Section 5 and Fally 2022 for examples), aggregator λ coincides with the budget multiplier. Regarding other aggregators Λ', note that: i) the good-specific demand function D_i can be a flexible function of Λ'; ii) aggregators Λ' are determined by the first-order condition (zero derivative of RHS) – we discuss in the next subsection the implications for price and income effects.

[7]A more natural yet abstract approach would be to define aggregators Λ as part of a smooth K-dimensional manifold. The proofs can be reformulated in this setting.

We defer the formal proof of Proposition 2 to the Appendix but we provide here some intuition behind it. By examining the gradients and using Goldman and Uzawa's lemma, the first step is to show that V takes the following form: $\log P = -M(\Lambda) + \sum_i \int_1^{p_i} S_i(t,\Lambda) d\log t$. In addition, the rank of the gradients of Λ and $\log P$ imply that the derivatives of the right-hand side must be null for each aggregator Λ_k (which we will refer to as the first-order condition in Λ_k). The envelope theorem then implies that the derivative of $\log P$ in $\log p_i$ equals $S_i(p_i,\Lambda)$, as desired. At this stage, however, nothing guarantees that such expenditure shares S_i sum up to one across goods.

The budget constraint (or, equivalently, homogeneity of P) further imposes functional form restrictions on the demand function. For such expenditure shares to add up to unity, it must be that the first-order condition in one of the aggregators (or a combination of those) implies the budget constraint. This is intuitively why one specific aggregator such as λ in equations (5) and (6) plays a specific role. It must enter symmetrically (across goods) as a price shifter.

Differentiating the budget constraint and using the rank assumptions, we obtain that the own price effect $\frac{\partial S_i}{\partial \log p_i}$ must be colinear with the derivatives of S_i in Λ. Moreover, we obtain that coefficients of colinearity can be expressed as functions of Λ. We can then construct a "flow" Φ (transformation within the Λ space) that must keep each $S_i(tp_i, \Phi(t,\Lambda))$ invariant w.r.t t (assumptions 1, 2 and 6 are useful to define Φ globally). We use this flow to project the aggregators onto λ and Λ'.[8]

The homogeneity of $\lambda(p)$ (degree one) and Λ' (degree zero) is then simply obtained by checking that λ and Λ' are solutions of the first-order conditions when prices are multiplied by λ. These results follow without initially imposing any homogeneity assumptions on the aggregators, aside from the assumption of homotheticity of preferences.

For the remainder of the paper, we denote the aggregators λ and Λ, with Λ rather than Λ'.

3.2. **Rationalization and concavity.** Conversely, we can obtain sufficient conditions under which the P function defined above is concave in prices, which would then imply that there is a well-defined quasi-concave utility from which we can derive such demand systems. As for Proposition 2, we consider smooth functions $D_i(p_i,\Lambda)$ and $G(\Lambda)$ where $\frac{\partial D_i}{\partial p_i} < 0$ and define:

$$\log \mathcal{P}(p,\lambda,\Lambda) \;=\; \log\lambda \;-\; G(\Lambda) \;+\; \sum_j \int_{t=1}^{p_j/\lambda} D_j(t,\Lambda)\, d\log t \tag{7}$$

This coincides with the price index when $\lambda(p)$ and $\Lambda(p)$ are such that the derivative of the right-hand side is null in (λ,Λ). The elasticity of P in each price p_i would then provide the expenditure shares described previously. What is left to provide are sufficient conditions for the concavity of P. This is obtained naturally by imposing concavity or convexity in Λ, conditions that are simpler to check.

[8]For these steps it is helpful to consider the "Lyapunov" function $\sum_i S_i(1,\Phi(t,\Lambda))$, that is strictly decreasing in t.

Lemma 1 *Suppose that the function* $\log \mathcal{P}(p, \lambda, \Lambda)$ *is defined as above, and is either convex in* Λ *or concave in* $(\Lambda, \log p)$*:*

i) If $\log \mathcal{P}(p, \lambda, \Lambda)$ is convex in Λ, define $P(p) = \max_\lambda\{\min_\Lambda \mathcal{P}(p, \lambda, \Lambda)\}$.
ii) If $\log \mathcal{P}(p, \lambda, \Lambda)$ is concave in $(\Lambda, \log p)$, define $P(p) = \max_{\lambda,\Lambda} \mathcal{P}(p, \lambda, \Lambda)$.

In both cases, we obtain a well-behaved price index function $P(p)$ *that is concave and homogeneous of degree one in* p*.*

Conditional on aggregators λ and Λ, one can verify that $\mathcal{P}(p, \lambda, \Lambda)$ is concave in log prices. The max and/or min operations in Lemma 1 preserve concavity. First, in terms of Λ, this includes taking the maximum in Λ, noticing that the domain of Λ is assumed to be $\mathbb{R}^K$ and is convex, so the max of $\log \mathcal{P}(p, \lambda, \Lambda)$ remains concave. Conversely, taking the minimum always preserves concavity, regardless of the domain of Λ.

By introducing λ, we take the "perspective" of such function,[9] which allows us to go from log concavity to concavity in p, and also provides homogeneity in prices when we take the maximum over λ. Perspective functions have been used recently in game theory, mean-field games, machine learning, transportation theory, among others (Combettes 2018, Combettes and Müller 2018).

A more general alternative is to assume that $\Lambda = (\Lambda^+, \Lambda^-)$ can be separated into two sets of aggregators, one set for which we have convexity and the other one where we have concavity. Our results carry by jointly taking the maximum over Λ^+ and the minimum over Λ^- if Slater conditions in (Λ^+, Λ^-) are satisfied.

When $P(p)$ is concave, demand can be alternatively derived from maximizing a concave utility function that is homogeneous of degree one in quantities. Here, this utility function can be expressed with a similar functional form as P.

Proposition 3. *Under the assumptions of Lemma 1, the demand system can be obtained from the maximization (under the budget constraint) of the following utility function:*

$$\log U \;=\; -\log \lambda \;-\; G(\Lambda) \;+\; \sum_i u_i(q_i \lambda, \Lambda)$$

where each $u_i(q_i, \Lambda)$ *is the Fenchel concave conjugate of* $\int_{t=0}^{\log p_i} D_i(t, \Lambda)\, dt$ *in* p_i *(conditional on* Λ*), and where we take the minimum/maximum over* Λ *depending on convexity/concavity.*

To prove this result, it is useful to note that the log of utility is the Fenchel concave conjugate of the log of the price index (see Appendix). Fenchel concave or convex conjugates are used implicitly in various contexts in economics, for instance to retrieve

[9]Take a real function $f(x)$ with $x \in \mathbb{R}^J$. For $t > 0$, the function $tf(x/t)$ is the "perspective" of f. It is concave (resp. convex) in (x, t) if and only if f is concave (resp. convex) in x. See Combettes (2018) for properties of perspectives.

cost and profit functions.[10] Here, in particular, each u_i is defined as:

$$u_i(q_i, \Lambda) \;=\; \min_{p_i} \Big\{ p_i q_i - \int_{t=1}^{p_i} D_i(t, \Lambda)\, d\log t \Big\}$$

and is concave in q_i.

Note that, for the optimal consumption basket as a function of prices, λ and Λ in this formulation coincide with λ and Λ in the dual.

3.3. **Implications for price effects.** When preferences are rational and allow for K aggregators, Proposition 2 states that one of such aggregators must play a special role and must enter symmetrically as a price shifter while demand can be a flexible function of other aggregators.

Furthermore, one should note that the first-order conditions in aggregators Λ imply some symmetry relating how Λ depends on prices p to how demand D depends on Λ itself. Differentiating the first-order conditions in λ and Λ, we obtain that $\frac{\partial \Lambda_k}{\partial \log p_i}$ is tightly linked to $\frac{\partial D_i}{\partial \Lambda_k}$:

$$\frac{\partial D_j}{\partial \Lambda_k} \;=\; -\sum_{k'} \mathcal{H}_{kk'} \frac{\partial \Lambda_{k'}}{\partial \log p_j} \;+\; \Big(\sum_j \frac{\partial D_j}{\partial \Lambda_k}\Big) \frac{\partial \log \lambda}{\partial \log p_j} \tag{8}$$

where matrix $\mathcal{H}$ is the Hessian of $\mathcal{P}$ in Λ in the right-hand-side of equation (7). If the right-hand-side of equation (7) is convex in Λ, as in case i) of Lemma 1, matrix $\mathcal{H}$ is definite positive. In case ii) of Lemma 1, it is definite negative.

Using this relationship, we can illustrate more directly the rank of cross-price effects and the influence of aggregators. Holding aggregator λ constant, we can exploit the J by K matrix $\frac{\partial S}{\partial \Lambda}$ and obtain the following expression for cross-price effects:

$$\frac{\partial S_i}{\partial \log p_j}\bigg|_{\lambda} \;=\; -\sum_{k,k'} \mathcal{H}^{-1}_{kk'} \frac{\partial D_j}{\partial \Lambda_{k'}} \frac{\partial D_i}{\partial \Lambda_k} \tag{9}$$

where $\mathcal{H}^{-1}_{kk'}$ are the coefficients of the inverse of the $\mathcal{H}_{kk'}$ matrix above (Hessian of $\mathcal{P}$ in Λ), and has a rank $K-1$. This matrix of cross-price effects inherits properties of $\mathcal{H}$ (e.g., it is positive semi-definite if $\mathcal{H}$ is positive definite, in the convex case) but its cells do not necessarily have same sign. In either case, it allows for complementarity between some of the goods, unlike additive random utility models (ARUM) such as mixed logit. This is illustrated more concretely in the specification in the next section.

4. Example of semi-parametric specifications for estimation

In this section, we illustrate the usefulness of these results by exploring a specification based on a linear relationship between own prices and aggregators. Special cases of it include nested logit/CES and the more recent Inverse Product Differentiation Logit model (IPDL, Fosgerau et al 2024).

[10] The Fenchel conjugate of a convex function $f(x)$ is $f^*(y) = \max\{y.x - f(x)\}$, and instead we use the minimum for a concave function. The conjugate of the conjugate is equal to the original function (Fenchel-Moreau theorem).

Suppose that the price index is:

$$\log P \ = \ \max_{\lambda,\Lambda} \Big\{ -\log\lambda \ - \ \sum_k g_k(\Lambda_k) \ - \ \sum_i W_i\Big(\log p_i + \log\lambda + \sum_k b_{ik}\Lambda_k\Big)\Big\}$$

with $W_i(x_i) = \int_{x_i}^{+\infty} D_i(t)dt$ and $D_i' < 0$, where $b_{ik} \in \mathbb{R}$ are parameters and where g_k are convex functions with $g_k' > 0$ and $g_k'' > 0$. Also, suppose that aggregators are such that the derivative of $\log P$ in λ and Λ is null. The right-hand side is concave in $\log p$, $\log\lambda$ and Λ, so aggregators are uniquely identified and the resulting demand system is well defined, as described in Lemma 1.

Each good can have a flexible own-price demand schedule D_i, as long as it is downward sloping. Note also that the substitution parameters b_{ik} can have any sign.

The first-order condition in λ yields the adding-up condition: $\sum_i D_i = 1$, so we obtain that the expenditure share on good i is given by each of these terms, with $S_i \ = \ D_i\Big(\log p_i + \log\lambda + \sum_k b_{ik}\Lambda_k\Big)$. In turn, the first-order condition for each aggregator Λ_k yields a simple expression as a sum of the b_{ik}'s (for each k) weighted by expenditure shares:

$$g'(\Lambda_k) \ = \ \sum_i b_{ik} D_i\Big(\log p_i + \log\lambda + \sum_{k'} b_{ik'}\Lambda_{k'}\Big) \ = \ \sum_i b_{ik} S_i \tag{10}$$

4.1. **Quadratic case for estimation.** Several specifications of functions g_k lead to very tractable solutions. For instance, as we describe later, nested logit/CES and IPDL (Fosgerau et al 2024) can be obtained by choosing exponential for each g_k and iso-elastic functions D_i. Here we explore an even simpler case by imposing a quadratic separable specification for the g's. The expression for Λ_k is the simplest with $g_k = \frac{1}{2}\Lambda_k^2$, so that:

$$\Lambda_k \ = \ \sum_i b_{ik} S_i$$

In that case, we can express expenditure shares more directly as:

$$S_i \ = \ D_i\Big(\log p_i + \log\lambda + \sum_j \beta_{ij} S_j\Big) \tag{11}$$

where β is a positive semi-definite symmetric matrix with coefficients $\beta_{ij} = \sum_k b_{ik} b_{jk}$.[11] This matrix is constant (does not vary with prices) and depends only on the primitive parameters b_{ik} to estimate.

Using the results in the previous section (expression 9 above), we obtain that the price effects are determined by each slope D_i' of the demand curve for good i as well as parameters b_{ik} that parameterize the cross-price effects:

$$\left.\frac{\partial S_i}{\partial \log p_j}\right|_{\lambda} \ = \ D_i'\, D_j' \sum_{k,k'} \mathcal{H}^{-1}_{kk'}\, b_{ik}\, b_{jk'} \tag{12}$$

where $\mathcal{H}^{-1}$ is the inverse of the matrix with coefficients $\mathbb{1}_{(k=k')} - \sum_i b_{ik} b_{ik'} D_i'$. This demand system is very flexible already as it can fit a wide range of substitution

[11]We can also simply check homotheticity by shifting all prices (in log) by a common constant term. This will increase $\log\lambda$ and $\log P$ by that same constant term, without affecting expenditure shares and other aggregators, hence $P(p)$ is homogeneous of degree 1 in prices.

patterns. To be more precise, conditional on own-price effects $\{D_i'\}$, we can choose parameters b to fit any rank-K positive semi-definite matrix $\mathcal{H}^{-1}$ with a spectral radius less than one (i.e. that does not have eigenvalues larger than one). This includes the mixed logit and BLP specifications (see Appendix for a proof).

In this specification, recovering welfare is not difficult. Once own demand and cross-price effects are estimated, we can recover $W_i(x_i) = \int_{x_i}^{+\infty} D_i(t)dt$ by integration. In turn, we can recover $\frac{1}{2}\sum_k \Lambda_k^2 = \frac{1}{2}\sum_{i,j}\beta_{ij}S_iS_j$ as the quadratic form associated with β and evaluated using expenditure shares.

While such demand can be derived from the price index specification above, we can alternatively derive it from a direct utility function (homogeneous of degree one). Using Proposition 3, it can be expressed as:

$$\log U(q) \;=\; -\log\lambda \;+\; \frac{1}{2}\sum_k \Lambda_k^2 \;+\; \sum_i u_i\Big(\log q_i + \log\lambda - \sum_k b_{ik}\Lambda_k\Big)$$

where aggregators are now defined as a function of q, and such that the right-hand-side has a zero derivative in $\lambda(q)$ and $\Lambda(q)$.

4.2. **Overlapping nests and relationship to IPDL (Fosgerau et al 2024).** Constructing nests is a practical way to increase the rank of a demand system. Consider different partitions of the set of goods. For instance, among yogurt, one partition could be composed of the set of vanilla-flavored yogurt, the set of plain yogurt, and the set with other flavors, so that each yogurt product is in either one of these sets. On top of this, we can consider another partition depending on the fat content, or quality labels, etc. A standard way to model a partition is nested logit, but this only allows for one partition, with more or less fine sets within that partition. A recent paper by Fosgerau, Monardo, and De Palma (2024) provides a useful way to consider several of such partitions at once, with a different substitution parameter for each partition, which they call IPDL (inverse product differentiation logit model). This can be viewed as a special case of our preferences, because demand can be expressed in terms of its own price, as well as an aggregator for each set of each partition. Here below we provide an intermediate generalization where preferences remain homothetic (using log prices instead of price levels in IPDL) and where we can pick any functional form of demand for own-price effects instead of the logit/CES formulation.

Denote each partition by $\mathcal{P}$. and by $\mathcal{S} \in \mathcal{P}$ the sets in that partition. For each such set, we define an aggregator $\Lambda_{\mathcal{S},\mathcal{P}}$. We show here that this is a special case of the specification highlighted in Section 4.1 where the G function is the exponential and where we specify $b_{i,\mathcal{P},\mathcal{S}} = \mu_{\mathcal{P},\mathcal{S}}\mathbb{1}_{(i\in\mathcal{S})}$, i.e. the inclusion function for set $\mathcal{S}$ and a parameter $\mu_{\mathcal{P},\mathcal{S}}$ that will capture substitution within vs. across goods of sets $\mathcal{S} \in \mathcal{P}$.

Suppose that the price index is:

$$\log P = \max_{\lambda,\Lambda}\Big\{\log\lambda - \sum_{\mathcal{P}}\sum_{\mathcal{S}\in\mathcal{P}} g(\Lambda_{\mathcal{S},\mathcal{P}}) - \sum_i W_i\Big(\log(p_i/\lambda) + \sum_{\mathcal{P}}\sum_{\mathcal{S}\in\mathcal{P}}\mathbb{1}_{(i\in\mathcal{S})}\mu_{\mathcal{S},\mathcal{P}}\Lambda_{\mathcal{S},\mathcal{P}}\Big)\Big\}$$

If we take a convex function g, the RHS is concave in the $\Lambda_{\mathcal{S},\mathcal{P}}$'s, so we can readily apply Lemma 1 and Proposition 3. Consider two practical cases: quadratic vs. exponential.

When g is quadratic as above, we obtain a special case of our specification (equation 11) where $\Lambda_{\mathcal{S},\mathcal{P}} = \mu_{\mathcal{S},\mathcal{P}}S_{\mathcal{S},\mathcal{P}} = \mu_{\mathcal{S},\mathcal{P}}\sum_{i\in\mathcal{S}} S_i$ is the aggregate expenditure share on goods i in the set $\mathcal{S}$ (for that partition $\mathcal{P}$). The expenditure share on good j is then:

$$S_j \;=\; D_j\Big(\log(p_j/\lambda)\;-\;\sum_{\mathcal{P}}\sum_{\mathcal{S}\in\mathcal{P}}\mathbb{1}_{(j\in S)}\mu^2_{\mathcal{S},\mathcal{P}}S_{\mathcal{S}}\Big)$$

When g is the exponential, we obtain a generalization of the IPDL demand system. Again, the first-order conditions in Λ's yield simple expressions: $\Lambda_{\mathcal{S},\mathcal{P}} = \log S_{\mathcal{S},\mathcal{P}} + \log\mu_{\mathcal{S},\mathcal{P}}$. The resulting expenditure share on good j is now:

$$S_j \;=\; D_j\Big(\log(p_j/\lambda)\;-\;\sum_{\mathcal{P}}\sum_{\mathcal{S}\in\mathcal{P}}\mathbb{1}_{(j\in S)}\mu_{\mathcal{S},\mathcal{P}}\log S_{\mathcal{S}} + \chi_j\Big)$$

for some constant term $\chi_j = \sum_{\mathcal{P}}\sum_{\mathcal{S}\in\mathcal{P}}\mathbb{1}_{(j\in S)}\mu_{\mathcal{S},\mathcal{P}}\log\mu_{\mathcal{S},\mathcal{P}}$.

In both cases, the terms in $S_{\mathcal{S}}$ account for different substitution patterns with goods that are in the same set, allowing for i) different layers or partitions $\mathcal{P}$; ii) different intensity of substitution parameters $\mu_{\mathcal{S},\mathcal{P}}$ across sets and partitions. The approach taken in our semi-parametric specification is however more general, as it can account for substitution patterns that may not just depend on partitions or sets but also continuous attributes (e.g. calories or sugar contents). With our approach, note also that we do not require sets to be organized in partitions, i.e. goods need not be exactly in one set for each partition.

4.3. **Relationship to mixed logit/CES.** Another common way to model non-trivial cross-price effects is to assume that a market is the aggregation of heterogeneous consumers (e.g. Berry 1994, Berry et al. 1995). Each consumer has Logit or CES preferences, but heterogeneous price elasticities and heterogeneous demand shifters (often modeled as heterogeneous evaluations of various product attributes). Two products have greater cross-price effects (i.e. are more substitutes) if they tend to be purchased by the same types of consumers. At the aggregate level, we show here that we can interpret such mixed logit demand as demand with price aggregators, where we have at most one aggregator by consumer type. Hence, a more complex demand system with a larger number of consumer types leads to a greater number of aggregators and a higher rank of cross-price effects.

Formally, suppose that demand is the aggregation of several types of consumers, indexed by k, each of which has an expenditure share given by a multinomial logit structure as standard in discrete-choice models.[12] Expenditure shares for type k of consumers are then given by:

$$\widetilde{S}_{ik} \;=\; \frac{e^{-\alpha_k\log p_i + b_{ik}}}{\sum_j e^{-\alpha_k\log p_j + b_{jk}}} \tag{13}$$

Denote by ω_k the aggregate income share of consumers of type k. The aggregate expenditure share is then:

$$S_i = \sum_k \omega_k\widetilde{S}_{ik} \;=\; \sum_k \omega_k\Lambda_k e^{-\alpha_k\log p_i + b_{ik}} \tag{14}$$

[12] As standard in the literature, we can assume that goods i differ in terms of their characteristics h, with ζ_{ih} describing the content of good i in characteristics h. Suppose that each type k of consumers has a valuation B_{kh} of characteristics h, we could have then: $b_{ik} = \sum_h B_{kh}\zeta_{ih}$.

with $\Lambda_k = \left(\sum_j e^{-\alpha_k \log p_j + b_{ik}}\right)^{-1}$. This specification coincides with a rank-K demand system, where the price index is defined as:

$$\log P = \max_{\lambda,\Lambda} \left\{ -\log \lambda - \sum_i \sum_k \Lambda_k e^{-\alpha_k(\log p_i + \log \lambda) + b_{ik}} + \gamma_0 \Pi_k \left(\Lambda_k\right)^{\gamma \omega_k / \alpha_k} \right\} \tag{15}$$

Note that the patterns of substitution are also constrained under mixed logit, and in particular goods must all be substitutes. We can show that the patterns of substitution are also less general than the ones obtained with our baseline specification given by equation 11. We describe in Appendix H how we can fit our baseline specification with rank-K cross-price effects to match locally (for any vector of prices) price effects from any given CES/logit mixed across K types of consumers.

As the number of consumer types goes to infinity, as in BLP with a continuum of consumers having heterogeneous valuations of product attributes, it becomes harder to characterize the rank of such demand system on aggregate. However, taking the second-order approximation in the scale of heterogeneity as in Salanié and Wolak (2022) (see also Borusyak, Bravo, and Hull 2025 for an intuitive description and application), we find that the rank of such demand system where we have M product attributes (with uncorrelated random coefficients) is bounded by $2M$.[13] In practice, most papers estimating BLP focus on just a small number of product attributes and therefore implicitly impose a low rank of cross-price effects, up to a 2nd-order approximation. In fact, our empirical results below indicates that the BLP substitution patterns are well approximated by low-rank matrices.

5. Heterogeneity and non-homotheticity

In this section, we extend our approach to account for non-homotheticity, i.e. non-trivial income effects. We then illustrate how the demand system can be viewed as "perturbed utility model" (PUM) treating consumer decisions as random realizations, where probabilities are chosen to maximize a perturbed utility function that is linear in prices or log prices.

5.1. **Non-homothetic demand with price aggregators.**

In the previous results, expenditure shares are assumed to be derived from a price index that is homogeneous in prices. Under non-homotheticity, we can apply the same results to the expenditure function, conditional on utility. We would then be considering demand where expenditure shares are functions:

$$\frac{p_i q_i}{w} = S_i(p_i, \Lambda_1(p, U), ..., \Lambda_K(p, U), U) \tag{16}$$

[13]To prove this claim, one can see in the formulation provided by Borusyak, Bravo, and Hull (2025) that demand can be expressed as a function of $2M$ aggregators (means and covariance terms) in addition to the own price. This implies a rank at most $2M$, where M is the number of attributes.

The rationality condition is then expressed using the expenditure function $e(p, U)$ which must have the same properties as the price index in p (homogeneity and concavity) but may also depend on utility. Specifically, Shephard's Lemma requires:

$$\frac{\partial \log e(p, U)}{\partial \log p_i} \;=\; S_i(p_i, \Lambda(p, U), U) \tag{17}$$

Rank and topological conditions [A1]-[A6] remain identical in terms of prices and aggregators. Under these conditions, using Proposition 3, we obtain the following functional forms for demand and the expenditure function:

Corollary 4. *Demand that depends on aggregators, including utility U, and satisfies all assumptions A1-A6 above must take the form:*

$$S_j \;=\; D_j(p_j/\lambda, \Lambda, U) \tag{18}$$

Moreover, it is derived from an expenditure function $e(p, U)$ that satisfies:

$$\log e(p, U) \;=\; \log \lambda \;-\; G(\Lambda, U) \;+\; \sum_j \int_{t=1}^{p_j/\lambda} D_j(t, \Lambda, U)\, d\log t \tag{19}$$

for some real function $G(\Lambda, U)$, and where the aggregators (λ, Λ) are such that the partial derivatives of the RHS in (λ, Λ) are null.

Conversely, if the right-hand side satisfies the concavity or convexity conditions highlighted in Lemma 1, the corresponding expenditure function is concave and homogeneous of degree one in prices, hence can be associated with rational preferences.

In this setting, one can think of utility as an additional aggregator, and replace U by indirect utility if we want to express demand as a function of prices and income. Aggregators would then depend on income (through indirect utility) and would be homogeneous of degree zero jointly in income and prices.

Example 1. These results can be readily applied to the specification discussed in the section above. Perhaps the most simple way to obtain flexible good-specific Engel curves is to incorporate an additive shifter $\alpha_i(U)$ that is itself a function of utility. Expenditure shares, expressed as Hicksian demand, would then be:

$$S_i \;=\; D_i\Big(\alpha_i(U) + \log p_i - \log \lambda + \sum_k b_{ik} \Lambda_k\Big)$$

while keeping the same expressions for aggregators Λ_k as in equation (10).

Example 2. A well-known and simple demand system that allows for flexible Engel curves combined with high-rank cross-price effects is the EASI demand system developed by Lewbel and Pendakur (2009), which can be seen as a generalization of the AIDS by Deaton and Muellbauer (1980).

Here we can obtain EASI preferences as a special case of our demand systems by considering the following expenditure function:

$$\log e(p, U) \;=\; \log \lambda \;-\; \frac{1}{2\gamma_k} \sum_k \Lambda_k^2 \;+\; \sum_i \int_0^{\log(p_i/\lambda)} D_i\big(t, \Lambda, U\big)\, dt \tag{20}$$

combined with linear demand functions D_i:

$$D_i\big(t, \Lambda, U\big) \ = \ \alpha_i(U) \ - \ \theta_i \, t \ - \ \sum_k b_{ik} \Lambda_k \tag{21}$$

where we impose $\sum_i \alpha_i(U) = 1$ for all levels of utility U, as well as $\sum_i b_{ik} = 0$, and where Λ and λ are such that the RHS of (20) has zero derivatives in Λ and λ (taking the maximum or minimum depending on the sign of γ_k).

Solving for Λ and λ in this maximization (see details in appendix), we obtain the following expenditure shares on good i with log-linear price effects and flexible Engel curves for each good (dictated by α_i):[14]

$$S_i \ = \ \alpha_i(U) \ + \ \sum_j \beta_{ij} \log p_j$$

where the coefficients $\beta_{ij} = \left[-\theta_i \mathbb{1}(i=j) + \frac{\theta_i \theta_j}{\sum_{j'} \theta_{j'}}\right] - \sum_k \gamma_k b_{ik} b_{jk}$ satisfy the standard conditions imposed with EASI, i.e. $\sum_j \beta_{ij} = 0$ and $\beta_{ij} = \beta_{ji}$.

Conversely, note that any symmetric matrix β with column sums equal to zero can be decomposed in this manner. Adjusting for own-price effects, the rank of β determines the number of aggregators K that are needed. Eckart-Young-Mirsky theorem then states that the largest eigenvalues of β (net of own price effects) would determine the quality of a lower-rank approximation.

Example 3. If we want to shut down all cross-price effects, yet retain flexible demand curves (own-price effects) as well as flexible Engel curves (income effects), we need two aggregators: λ to ensure that the budget constraint holds and utility U for income effects. Applying Corollary 4, demand then takes the form:

$$S_i \ = \ D_i \left(p_i/\lambda \, , \, U \right)$$

and must be obtained from the following expenditure function:

$$\log e(p, U) \ = \ \max_{\lambda} \Big\{ \log \lambda + \sum_j \int_{t=0}^{\log(p_j/\lambda)} D_j(t, U) \, dt \Big\}$$

As long as good-specific demand functions D_j are decreasing in its first argument, this expenditure function $e(p, U)$ is concave and homogeneous of degree one in prices. In addition, one must also ensure that it is increasing with utility U. This is a special case of Fally (2022) where the special aggregator is homogeneous in prices.

While cross-price effects are then very simple under this specification (rank 1 or 2), this demand specification is already flexible enough to construct demand systems with flexible own-price effects as well as flexible Engel curves by choosing how each function D_i depends on p_i/λ and utility U.

[14] Here for the sake of exposition we omit the interaction terms between price effects and utility. Such interactions can be obtained by adding a term $\sum_l \gamma_{il} \Psi_l$ with coefficients satisfying $\sum_i \gamma_{ik} = 0$, combined with additional aggregators Ψ_l.

5.2. **Heterogeneity and "perturbed utility" representation.** A recent formulation of consumer preferences with "Perturbed Utility Models" (PUM) has been proposed by D. L. McFadden and Fosgerau (2012), Fosgerau, Paulsen, and Rasmussen (2022) and Allen and Rehbeck (2019). This class of utility functions is more general than random-utility models (RUM). Yet it can still be interpreted naturally as the aggregation of consumers with heterogeneous tastes, and the aggregation of perturbed utility (where heterogeneity is in the "perturbation" function) is again a perturbed utility (Allen and Rehbeck 2019). This class of utility is actually very general, and the preferences described here can be expressed as PUM with simple formulations. Here we provide a PUM representation as a function of log prices where the dual variables are expenditure shares. In the Appendix, we provide an alternative PUM representation as a function of prices where dual variables are quantities.

Under the assumptions of Lemma 1, preferences described above can be reformulated as:

$$\log V(p) \;=\; \max_{S\in\Delta}\Big\{-\sum_i S_i \log p_i \;-\; F^*(S)\Big\}$$

where Δ is the simplex (i.e. imposing $\sum_i S_i = 1$). The share S_i can be interpreted as the probability of picking good i. In the limit case where the perturbation function is linear, only a single good is being chosen. Convexity leads to non-trivial probabilities for the choice of each good. Here, the perturbation function takes the form:

$$F^*(S) \;=\; \min_{\Lambda}\Big\{G(\Lambda) \;+\sum_j W_j^*(S_j,\Lambda)\Big\} \tag{22}$$

where W_j is again a Fenchel conjugate, as above, but now in terms of logs and shares:

$$W_j^*(S_j,\Lambda) = \min_{p_j}\Big\{S_j \log p_j - \int_{t=1}^{p_j} D_j(t,\Lambda)\, d\log t\Big\}$$

Similar to the random-utility model (RUM), we can interpret the expenditure share S_i as the probability of picking good i. The terms W_j^* in the perturbation function ensure that the probabilities are not degenerate and the $G(\Lambda)$ term captures interactions across these probabilities.

Example. The semi-parametric demand for estimation (see section 4) can be derived from a simple PUM representation. Consumers can be described to choose their expenditure shares S_i to maximize their utility based on the following expression:

$$\log V \;=\; \max_{S\in\Delta}\Big\{-\sum_i S_i \log p_i \;-\; F^*(S)\Big\}$$

where the perturbation function is:

$$F^*(S) \;=\; \sum_j W_j^*\big(S_j\big) + \frac{1}{2}\sum_i\sum_j \beta_{ij} S_j S_i$$

The terms W_j^* are additively separable and depend on the demand curve for each good j. A more convex adjustment W_j leads to a less price sensitive expenditure share S_j. In turn, the second term is quadratic in expenditure shares and captures all cross-price effects, where β is the positive semi-definite matrix defined above. With: $\frac{\partial^2 \log V}{\partial S_i\, \partial S_j} \;=\; -\beta_{ij} \;=\; -\sum_k b_{ik} b_{jk}$, we can see that the b terms parameterize how the expenditure share S_j influences the marginal utility from consuming good i. Since there is no restriction on the sign of coefficients b, we can potentially have complements, so that a higher expenditure share on a good j may potentially increase the marginal utility from consuming another good i, i.e. $-\beta_{ij} > 0$.

6. Estimation approach

Here we are interested in situations with many goods (large J) where it is not practically possible to estimate fully flexible cross-price effects of the order of J^2. Demand with aggregators provides a flexible framework to modeling and estimating demand with non-trivial cross-price effects when the number of goods is large and we need to reduce the rank of cross-price substitution matrix. Our theoretical results impose discipline on the structure of these effects, exploiting the mathematical structure of demands for rational utility-maximizing consumers.

Our main specification follows Section 4.1. To keep the focus on cross-price effects, we shut down income effects by developing a demand system which assumes homothetic preferences. In a set of extensions we then explore different ways to account for more flexible income effects.

6.1. **Data.** The NielsenIQ home scanner data[15] are collected through hand-held scanner devices that households use at home after their shopping in order to scan each individual transaction they have made. The data cover a quarter billion dollars of grocery expenditures, from about 60,000 individual households spread evenly across 53 "Scantrack" markets in the US (which approximately coincide with large metropolitan areas). Here as an illustration we focus on ready-to-eat cereals as in Nevo (2001) and Backus, Conlon, and Sinkinson (2021), and further narrow the scope of our analysis to the 2010-2019 time period. We augment our dataset with information on product attributes for the top-50 brands, including the content in sugar, fiber, fat, protein and sodium (see Table 1).

For this application, we aggregate expenditure by market at the monthly level. Hence, our dataset is similar to those typically used in the Industrial Organization literature, following Berry et al (1995), estimating demand across granular products and markets (or stores).

6.2. **Specification with additional functional form assumptions.** As indicated above, we adopt a homothetic specification so as to focus on cross-price effects. Our starting point is the specification highlighted above (equation 11 in section 4.1) where aggregators enter linearly in combination with the own price of the good. We also assume that variation across markets and products is influenced by unobserved taste shocks ε_{it}.

[15]Disclaimer: Researchers' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researchers and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Table 1. Summary Statistics across attributes

Attribute	Mean	Stdev	Min	Max
Sugar	8.047	3.315	1.465	15.79
Fiber	1.561	1.038	0.000	3.885
Fat	1.114	0.859	0.000	3.400
Protein	2.003	0.776	0.751	5.128
Sodium	143.4	50.29	2.286	221.4

Inverting the own demand curve D_i for each good i, we obtain the following specification for each market/time t:

$$D_i^{-1}(S_{it}) = \log p_{it} - \log \lambda_t + \sum_j \beta_{ij} S_{jt} + \varepsilon_{it} \tag{23}$$

Intuitively, own price effects are influenced by the shape of D_i while cross-price effects are determined by the rank-K matrix β with coefficients $\beta_{ij} = \sum_k b_{ik} b_{jk}$. Both have to be estimated. Aggregator λ_t can be regarded as a fixed effect that is uniform across all goods.

In all that follows, we interpret the error term as idiosyncratic demand shock at the market t by good j level, which we treat as an additive price shifter in logs. Notice that, after inverting the left-hand side, there is no restriction on the domain of ε_{it}.

6.2.1. *Parameterization of D.* Various parameterizations would be convenient for D_i; all that is required is that it be strictly decreasing. A first specification that we favor is iso-elastic, parameterized by a product-specific shifter α_i and potentially product-specific elasticity θ_i:

$$D_i(t) = \exp\left[-\alpha_i - \theta_i t\right] \tag{24}$$

so that its inverse $D_i^{-1}(S_{it}) = -\frac{1}{\theta_i}(\log S_i + \alpha_i)$ leads to the simple log-linear specification

$$\log S_{it} = -\alpha_i - \theta_i \log p_{it} + \theta_i \log \lambda_t - \theta_i \sum_j \beta_{ij} S_{jt} - \theta_i \varepsilon_{it}.$$

We can also consider a parameterization of D_i that allows for a choke price, i.e. such that demand is null above a certain reservation price. In the following specification:

$$D_i(t) = \nu \left[e^{-\alpha_i - \theta_i t} - 1\right] \tag{25}$$

demand for i is positive if and only if the right hand side of (23) is smaller than $-\alpha_i$.

6.2.2. *Projecting on attributes: our baseline estimation equation.* Following BLP and related approaches, a natural assumption is that goods with similar characteristics are closer substitutes than others. Here, a simple way to capture this idea is to project the good/aggregator-specific demand shifter b_{ik} onto the space of product characteristics. Specifically, we use data ζ_{il} on attributes l across goods i, informing on whether good i has attribute l (in which case ζ_{il} is a dummy variable) or the intensity of that attribute (ζ_{il} is then a scalar). Using such data, we now impose:

$$b_{ik} = \sum_l B_{kl} \zeta_{il}$$

With such specification, cross-price effects between goods i and j are then fully determined by how i and j differ in terms of their attributes. Aggregators are now determined by expenditure shares across attributes:

$$\Lambda_{kt} \;=\; \sum_{l} B_{kl} Z_{lt},$$

where Z_{lt} is defined as the sum of expenditure shares S_{it} across goods i weighted by their observed attribute ζ_{il}:

$$Z_{lt}(\zeta, S) \;=\; \sum_{i} \zeta_{il} S_{it}.$$

When ζ_{il} is a dummy for a label (e.g., whether yogurt i has a vanilla flavor), Z_{lt} is simply the observed share of expenditures across products that have such a label. In turn, cross-price effects are determined by an interaction between observed characteristics ζ_{il}, weighted expenditures $Z_{l't}$ on characteristics l', and matrix $\Gamma_{ll'} = \sum_k B_{kl} B_{kl'}$ capturing substitution patterns between attributes l and l'. Matrix Γ now becomes the key positive semi-definite matrix to be estimated.

With this projection on attributes, the estimating equation can be written as:

$$D_i^{-1}(S_{it}) = \log p_{it} - \log \lambda_t + \sum_{l,l'} \Gamma_{ll'}\, \zeta_{il} Z_{l't} + \varepsilon_{it}. \tag{26}$$

6.2.3. *Relation to Berry et al (1995) demand inversion.* A major advantage of specification (26) is that the price substitution effects enter linearly and do not enter the inversion of own expenditure shares on the left-hand side. In contrast, the inversion in BLP (Berry, Levinsohn and Pakes, 1995) relies on the mixing parameters that both characterize the heterogeneity of tastes across consumers and the cross-price effects, and is highly non-linear in these parameters. With BLP, one must therefore invert again to re-evaluate the orthogonality condition for each new set of heterogeneity parameters. Here, the inversion can rely on analytical solutions, and the linear specification in Γ leads to more transparent identification and helps to avoid the problem of weak instruments highlighted in Gandhi and Houde (2019), while retaining flexible substitution patterns and allowing for complementarity unlike BLP.

6.3. **IV and GMM formulation.** We describe two alternative estimation strategies: i) a linear IV estimator imposing Γ to be symmetric but not necessarily positive semi-definite; ii) a GMM specification of a non-linear estimator imposing both symmetry and positive semi-definiteness on Γ.

6.3.1. *Linear Problem.* Requiring the estimated matrix $\hat{\Gamma}$ to be positive semi-definite involves a set of non-linear constraints (or alternatively, makes the objective function nonlinear in parameters). But if we set aside the requirement that Γ must be positive semi-definite, then we can express the estimation problem as a linear problem.

Starting from (26), combined with the log-linear specification of D_i and a common θ, we obtain the linear expression:

$$-\log S_{it} = \alpha_i + \theta \log \lambda_t + \theta \log p_{it} + \theta \sum_{l,l'} \Gamma_{ll'}\, \zeta_{il} Z_{l't} + \varepsilon_{it}, \tag{27}$$

α_i and $\theta \log \lambda_t$ serve as product and market–time fixed effects. With symmetry imposed linearly ($\Gamma_{ll'} = \Gamma_{l'l}$), (27) can be estimated by conventional linear IV methods given suitable instruments. However, such estimates $\hat{\Gamma}_{\text{LIV}}$ need not be positive semi-definite. In what follows we address this by moving to a nonlinear GMM formulation that imposes the required curvature restrictions.

Observed variables are $\log S_{it}$, $\log p_{it}$ and $\zeta_{il} Z_{l't}$, with $Z_{l't} = \sum_i \zeta_{il'} S_{it}$. Linear coefficients to be estimated are $\theta \Gamma_{ll'}$ and θ, as well as α_i and λ_t which can be interpreted respectively as product and time/market fixed effects or nuisance parameters.

Further, as requiring $\Gamma_{ll'} = \Gamma_{l'l}$ for $l \neq l'$ is a linear set of constraints, we can impose symmetry on Γ without compromising the linearity of the estimator. This leaves us with the need to estimate $1 + M(M+1)/2$ parameters, and so a need for at least this many restrictions.

Suppose that we have appropriate instruments, both valid and sufficiently numerous to achieve identification (see discussion of the differentiation instruments below). Still, note that estimating this equation using conventional linear estimators will involve estimating $1 + M^2$ parameters (not counting the fixed effects α and $\log \lambda$). This is more than the available linear independent differentiation instruments $(1 + M(M+1)/2)$. As a formal matter symmetry implies $M(M-1)/2$ linear restrictions, so that the estimator is identified. However, textbook implementations of, say, two-stage least squares typically assume identification even absent additional linear restrictions (Greene and Seaks 1991 note a similar issue for the case of restricted OLS).

Do note, however, the resulting linear estimates Γ_{LIV} may be defective, as there is no guarantee that the linear estimates of Γ will yield a positive semi-definite matrix. We discuss an approach to correcting this using a non-linear GMM estimator below.

6.3.2. *Differentiation Instrumental Variables.* Gandhi and Houde (2019) consider the problem of estimating a Berry, Levinsohn, and Pakes (1995) style model of differentiated products. This is *not* our model, but is a non-linear GMM problem using aggregate data that poses issues similar to ours as the number of characteristics grows large. They are interested in particular with the problem of weak instruments, and argue (formally in the context of the BLP model, but heuristic arguments may also apply to our setting) that the appropriate instrument set should include what they call "differentiation IVs," which capture differences in characteristics across goods.

In our case, characteristics for J different products are encoded in a $J \times M$ matrix ζ. Thus, differences across products for characteristic m can be written as

$$d_m = [\zeta_{jm} - \zeta_{j',m}]$$

yielding a $J \times J$ matrix of differences for each attribute m. Similarly, $d_p^s = [\log p_{js} - \log p_{j',s}]$ gives differences in log prices across goods for a market-period s (note the important difference that d_m is invariant across s).

Importantly, not all goods are available in all market-periods. Let A be an $S \times J$ "availability" matrix of ones and zeros. We can then construct a symmetric measure of difference for characteristic m using $A d_m^2$, yielding an $S \times J$ instrument matrix. Extending this idea, what Gandhi and Houde (2019) call the "instrument function" can be written as

$$GH(\mathbf{p},\zeta) = \begin{cases} A\,d_p^2 & \text{Price differentiation,} \\ A\,d_m^2 & \text{Differentiation in attribute } m = 1,\ldots,M, \\ A\,(d_m \odot d_\ell) & \text{Interaction between attributes } m \text{ and } \ell, \end{cases} \tag{28}$$

where $\odot$ denotes the Hadamard product. This yields $1 + (M+1)M/2$ instruments for both the linear and nonlinear specifications.

6.3.3. *A new set of instruments.* A practical issue with the instruments proposed by Gandhi and Houde (2019) arises when all or most products are available in all or most market-period combinations. When such an issue arises, the only subset of Gandhi-Houde instruments with some variation would consist of the price differentiation instruments as well as interactions between the price and differentiation for a specific attribute m. The number of available instruments would then only grow linearly in the number of product attributes, and would not allow for the estimation of a full matrix of covariance of taste shocks across all pairs of attributes.

Building on both our approach and Gandhi and Houde (2019), instead we consider:

$$FL(\mathbf{p},\zeta)_{ms,jj'} \;=\; \zeta_{j,m} \sum_{j' \in J_s} \zeta_{j',l} \Delta \log p_{j's}$$

where $\Delta \log p_{j's} = \log p_{j's} - \overline{\log p}_s$ is the log price of product j' in market-time s relative to the average log price for that market (taking the sum over products $j' \in J_s$ available in market s). It closely fits the right-hand-side of equation (26), where we consider a weighted average of characteristics, using relative prices as weights instead of market shares. This set of instruments is defined across all pairs of product attributes (m,l) and varies across markets even if the set of available products remains invariant, as long as prices vary both across markets/periods and goods with different attributes.

6.3.4. *Instrument sets.* Both the Gandhi–Houde IV and our new construction admit several variants depending on whether one uses level or log prices and whether one includes polynomial or interaction terms. In the empirical application, we consider seven instrument specifications, summarized in Table 2. Our preferred specification uses our new instruments based on log-price instruments (henceforth FL), whose log-price construction matches the demand parameterization of both our model and BLP, permitting a structurally clean comparison. The FL instruments generate sufficient cross-market variation from the linear interaction of characteristics and prices alone; the Gandhi–Houde instruments (GH), which rely on pairwise product differences, require polynomial expansion to achieve comparable identifying power.

6.3.5. *Generalized method of moments.* To guarantee that the estimated matrix Γ is symmetric and positive semi-definite of rank $K \leq M$, we normalize $\Gamma = \theta^{-1} BB^\top$ with $B \in \mathbb{R}^{M \times K}$. Since BR and B yield the same Γ for any orthogonal matrix $R \in O(K)$, the relevant parameter space for B can be described as a Riemannian quotient manifold of equivalence classes $[B] = \{BR : R \in O(K)\}$.[16] We denote this

[16]In brief, we optimize over the space of matrices B, treating two matrices B that yield the same BB^t as the same point in that space. That space, a Riemannian manifold, embodies a natural notion of distance and convergence.

Table 2. Instrument specifications

Abbreviation	Description	Excluded IVs
FL	Characteristic × price-weighted characteristic aggregates (log prices)	36
FL-level	Same construction as FL but using prices in level	36
FL-poly	Polynomial ($n = 2$) variant of FL	36
GH	Gandhi-Houde differentiation IVs, log price cross-product attributes	6
GH-poly	Polynomial expansion of Gandhi–Houde IVs	6
Combined	Union of FL, GH, FL-poly, GH-poly	78–92
Nevo	Mean of product j's price across other markets in the same geographic region, following Nevo (2001)	14

manifold of rank-K symmetric positive semi-definite matrices by

$$\mathcal{S}_+^{(M,K)} = \{\, BB^\top : B \in \mathbb{R}^{M\times K}, \operatorname{rank}(B) = K\,\}/O(K).$$

The full parameter space for the nonlinear estimator is thus the product manifold

$$\mathcal{M} = \mathbb{R} \times \mathcal{S}_+^{(M,K)},$$

corresponding to the scalar parameter θ and the structured component Γ.

With the instruments $IV = FL(\mathbf{p}, \zeta)$ or $GH(\mathbf{p}, \zeta)$ in hand, we form unconditional moment restrictions using $\epsilon_{sj}(\theta, B) = \log S_{sj} + \theta \log p_{sj} + \alpha_j + H_s + \zeta_j^\top BB^\top Z_s$. Defining $g_{sj}(\theta, B) = \epsilon_{sj}(\theta, B) \odot IV(\mathbf{p}, \zeta)$, the sample moments and continuously-updated GMM criterion are

$$m_N(\theta, B) = \frac{1}{S}\sum_s \frac{1}{J}\sum_j g_{sj}(\theta, B), \tag{29}$$

$$J_N(\theta, B) = N\, m_N(\theta, B)^\top \hat{\Omega}_N(\theta, B)^{-1} m_N(\theta, B), \tag{30}$$

where $\hat{\Omega}_N$ is the sample covariance of the moments.

Our primary estimator is the two-step GMM estimator: in the first step we minimize the criterion with $\hat{\Omega}_N$ set to the identity (or a diagonal matrix of moment variances), and in the second step we re-estimate using the efficient weighting matrix evaluated at the first-step estimates. The two-step estimator solves:[17]

$$(\hat{\theta}, \hat{B}) = \arg\min_{(\theta, B)\in\mathcal{M}} J_N(\theta, B). \tag{31}$$

[17] We also considered the continuously-updated estimator (CUE), which re-evaluates $\hat{\Omega}_N(\theta, B)$ at each candidate parameter value. In our application the CUE criterion surface exhibits a pathological attractor: regardless of the instrument set, the CUE criterion converges to approximately 21 with degenerate eigenvalues of Γ. This behavior is consistent across instrument sets and values of K, and we therefore report two-step results throughout.

6.3.6. *Computation.* The optimization problem (31) is defined on the smooth manifold $\mathcal{M}$. We compute the solution using `ManifoldGMM` (Ligon 2025), which provides routines for GMM estimation on product manifolds by extending `pymanopt` (Townsend, Koep, and Weichwald 2016). Optimization proceeds via the Riemannian trust-region algorithm of Baker, Absil, and Gallivan (2008), developed for smooth manifolds, with guarantees regarding computational complexity and global convergence established by Sheng and Yuan (2024). By searching directly over points on $\mathcal{S}_+^{(M,K)}$, the algorithm ensures that updates to B remain on the manifold and that the estimated matrix $\hat{\Gamma} = \hat{B}\hat{B}^\top/\hat{\theta}$ is symmetric and positive semi-definite by construction.

Given a solver for fixed rank K, we determine the appropriate number of aggregators by sequential testing. Starting from $K = 1$, we compute the test statistic J^K (asymptotically χ^2 with $\ell - MK - 1$ degrees of freedom under the null that the moment conditions are satisfied). We then test whether increases in K significantly improve fit, continuing until further increases yield no significant improvement.

6.3.7. *Demographic taste shifters.* In the baseline specification, the cross-price substitution matrix Γ and the price-elasticity parameter θ are homogeneous across markets. We extend the model by adding demographic $\times$ characteristic interaction regressors, analogous to the π matrix in BLP. The structural equation (26) becomes

$$D_i^{-1}(S_{it}) \;=\; \log p_{it} - \log \lambda_t \;+\; \sum_{l,l'} \Gamma_{ll'}\, \zeta_{il}\, Z_{l't} \;+\; \sum_{s,l} \Pi_{sl}\, \zeta_{il}\, \bar{d}_{st} \;+\; \varepsilon_{it}, \tag{32}$$

where $\bar{d}_s$ is a vector of market-mean demographics (in our application, household income and male head age) and Π is a $D \times M$ matrix of demographic taste shifters on characteristics to be estimated. This allows the baseline valuation of characteristics to vary with market demographic composition while keeping Γ (the substitution structure) homogeneous.

We consider three estimation settings. Our preferred specification incorporates demographic instruments and demographic controls Π, but we also report estimates that progressively remove demographic information: using only the base instruments, without demographic controls Π; and a specification that augments the instrument set with demographic $\times$ characteristic interactions but does not estimate Π. This three-way structure isolates the separate contributions of demographic instruments and demographic taste-shifting parameters.

6.3.8. *Bootstrap inference.* We conduct inference via a moment wild bootstrap with Rademacher weights. In each bootstrap replication, the observation-level moment contributions g_{sj} are multiplied by independent Rademacher random variables (± 1 with equal probability), and the two-step estimator is re-computed on the perturbed moments. Percentile confidence intervals for θ, the eigenvalues of Γ, and the elements of Π are constructed from the bootstrap distribution.

6.4. **Comparison to BLP.** For the sake of comparing our approach to Berry, Levinsohn, and Pakes (1995), we re-estimate demand using the data described above, i.e. with the same expenditure shares, prices, and instruments. We borrow the approach described in Conlon and Gortmaker (2020) and use the PyBLP Python code.

To make it comparable to our approach described above, we allow for market-time fixed effects as well as product fixed effects. We estimate correlated coefficients in attributes by either assuming a diagonal covariance matrix or allowing for non-zero off-diagonal terms. Following Conlon and Gortmaker (2025), we also use demographics data drawn from individual household characteristics: log income, age of household head, and whether the household includes kids. As with the K-aggregator, we estimate BLP under three demographic settings—no demographics (Monte Carlo Halton draws), empirical demographic draws without π, and full demographics with π taste shifters—to isolate the contribution of demographic heterogeneity to identification.

7. Application: Demand for Dry Cereal

7.1. **Results.** We present estimation results for consumer demand in the ready-to-eat cereal market. We first report the K-aggregator estimates, then the BLP random-coefficients logit estimated on the same data and instruments, and finally a head-to-head comparison of the two models. Throughout, our preferred instrument set is denoted FL with a log-price construction that matches the demand parameterization of both models. Results for alternative instrument sets and demographic settings are collected in Appendix H.

7.1.1. *K-Aggregator Estimation.* How large is the rank K of cross-price effects? Table 3 reports two-step GMM estimates with demographic taste shifters for the FL instrument set at $K = 1, 2, 3$. Moving from $K = 1$ to $K = 2$ produces a meaningful second eigenvalue ($\lambda_2 = 0.706$) and a notable drop in $\hat{\theta}$. Moving from $K = 2$ to $K = 3$ leaves the first two eigenvalues essentially unchanged ($\lambda_1 = 5.82$, $\lambda_2 = 0.817$) while the third eigenvalue is negligible ($\lambda_3 = 0.055$), indicating that the optimizer cannot find a direction in which to place a third factor. In our specification projected on attributes, the data support at most rank 2.

Table 3. K-aggregator two-step GMM: FL instruments with demographic taste shifters (Π).

K	J	dof	$\hat{\theta}$	λ_1	λ_2	λ_3
1	220.47	34	1.16	1.13	≈ 0	≈ 0
2	213.47	29	1.18	5.84	0.706	≈ 0
3	212.98	24	1.18	5.82	0.817	0.055

Notes: Eigenvalues λ_k are those of $\hat{\Gamma} = \hat{B}\hat{B}^{\top}/\hat{\theta}$. Values below 10^{-2} are reported as ≈ 0.

The preferred specification ($K = 2$) yields $\hat{\theta} = 1.18$ for own-price effects and a two-factor substitution structure with eigenvalues $\lambda_1 = 5.84$ and $\lambda_2 = 0.706$. Moment wild bootstrap inference gives 95% confidence intervals of (1.05, 1.30) for θ, (1.08, 14.26) for λ_1, and (0.003, 2.56) for λ_2.[18] The first eigenvalue is clearly separated from zero; the second is borderline significant, with its confidence interval barely excluding zero.

[18]Bootstrap details: 796 Rademacher replicates; percentile confidence intervals.

The estimated demographic taste shifters $\hat{\Pi}$, reported in Table 4, reveal that higher income and older household heads are associated with stronger demand for fiber ($\hat{\Pi} \approx +0.09$ to $+0.13$) and, to a lesser extent, protein. Sodium coefficients are uniformly near zero.

TABLE 4. Estimated demographic taste shifters (FL instruments, $K = 2$)

	Sugar	Fiber	Protein	Fat	Sodium
Household income	0.002	0.094	0.030	0.020	−0.001
Male head age	0.003	0.126	0.042	0.004	−0.001

Notes: Each entry gives the effect of a one-unit increase in market-mean demographics on the valuation of the corresponding characteristic.

As we discuss below for BLP, adding demographic taste shifters is essential for identification. Table 5 compares the FL specification at $K = 2$ across the three demographic settings. Without Π (demographic instruments only), the second eigenvalue is negligible (0.062), which means that the rank K is effectively equal to one. With Π, both eigenvalues are large, giving the clearest evidence for a rank-2 substitution structure. Absorbing demographic heterogeneity into Π allows Γ to capture the pure substitution structure, lowering J and stabilizing $\hat{\theta}$.

TABLE 5. Effect of adding Π (FL instruments, $K = 2$)

Setting	λ_1	λ_2	J	dof	$\hat{\theta}$
No demographics	3.65	0.184	136.77	25	0.68
Demog instruments	3.94	0.062	236.92	39	1.11
Demog IV + Π	5.84	0.706	213.47	29	1.18

7.1.2. *BLP Estimation.* We estimate a BLP random-coefficients logit on the same data, using the same FL instruments augmented with demographic interactions.[19]

Without demographics, the random coefficients σ_k collapse to zero for prices and other attributes across most instrument sets.[20] Once the π demographic taste shifters are included, all specifications consistently estimate $\sigma_1 \approx 0.10$ (heterogeneous valuations of the outside option), while $\sigma_{\text{price}} = 0$ universally. The price coefficient moves from -5.31 to -3.41 when demographics are included.

Table 6 reports the preferred BLP specification. Controlling for demographic shifters (Π) leads to more stable parameter estimates, but the BLP objective function *rises* with demographics (from 913 to 931), a pattern that persists across instrument sets. This contrasts with the K-aggregator, whose J *falls* when π is added. The asymmetry is economically significant: the K-aggregator's flexible Γ can absorb the additional moment conditions that demographic instruments provide, while BLP's diagonal Σ restriction cannot.

[19]Estimation uses PyBLP with empirical demographic draws. The π taste shifters are initialized at small random values; full details are in Appendix H.

[20]All seven instrument sets and three demographic settings are reported in Appendix H.

Table 6. BLP estimates with FL instruments and demographics.

	No demographics	With demographics
β_{price}	-5.307	-3.412
σ_1 (constant)	0.000	0.101
σ_{sugar}	0.000	0.007
σ_{price}	0.000	0.000
Objective	912.970	930.690

Notes: No-demographics specification uses 200 Halton draws. Specification with demographics uses empirical demographic draws with taste shifters on income and age.

7.1.3. *Model Comparison.* With both models estimated on the same data, instruments, and demographic controls, we can compare them along several dimensions.

The own-price-sensitivity parameters are not directly comparable in scale across the two models, but both are robustly identified with demographics: the K-aggregator's θ clusters at 1.1–1.3 across instrument sets, while BLP's β_{price} clusters at -3.3 to -3.7. Without demographics, both parameters exhibit much greater sensitivity to the choice of instruments.

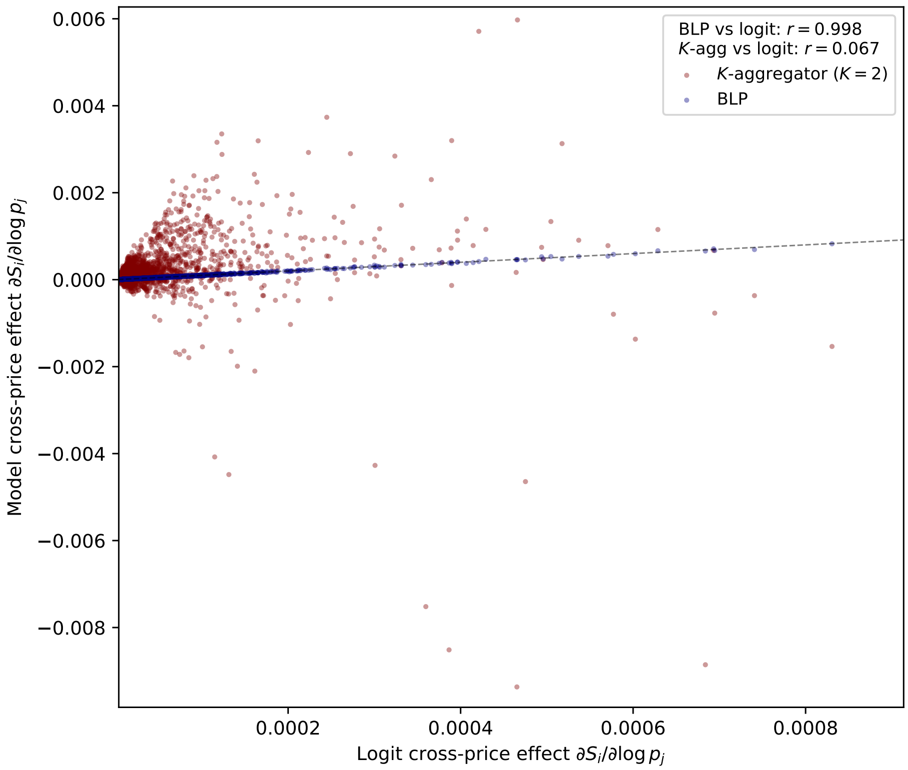
The K-aggregator's Γ varies substantially across instruments (eigenvalues range from 0.05 to 51.5), adapting its full positive semi-definite substitution matrix to the identifying variation. BLP's diagonal Σ is nearly invariant ($\sigma_1 \approx 0.10$ everywhere), reflecting its structural restriction to diagonal random coefficients. To compare the two models' substitution structures on a common footing, we decompose their implied $n \times n$ cross-price matrices directly.

Both models estimate a demographics $\times$ characteristics interaction matrix. Qualitative patterns agree on the two most important characteristics: higher income is associated with stronger demand for fiber (K-agg: $+0.094$; BLP: $+0.007$) and weaker demand for fat (K-agg: $+0.020$; BLP: -0.013). Magnitudes differ by roughly an order of magnitude, reflecting the different utility normalizations.

The models' elasticity matrices, evaluated at sample-mean prices and shares, display substantial overall disagreement. The cross-price elasticities are weakly negatively correlated (Pearson $r = -0.072$, $p = 0.001$), indicating that the two models disagree on which products are close substitutes. The relative Frobenius distance between the two elasticity matrices, $\|\Delta\|_F / \|E^{\text{BLP}}\|_F = 0.773$, confirms the scale of the discrepancy. Importantly, the eigenvalues of the difference matrix are small and uniform, indicating that the two models differ diffusely across the full product space rather than along a few dominant dimensions—precisely what one would expect given the rank difference documented above.

The distribution of cross-price effects implied by our estimates is considerably more dispersed with K-aggregators than with BLP, while BLP cross-price effects are all positive by construction (histograms are shown in Figure 2, Appendix H). To illustrate the relative dispersion of cross-price effects, in Figure 1 we plot our estimates of $\partial S_i / \partial \log p_i$ for each model against those implied by a Logit demand system. In spite of incorporating random coefficients with significant variance, BLP cross-price effects deviate only little from Logit. By contrast, cross-price effects with

FIGURE 1. Cross-price effects for BLP and K-Agg against Logit



K-aggregators are only mildly correlated with those in Logit, and span a wider range of negative and positive values, conditional on Logit price effects. CES may provide a more natural benchmark for K-aggregators, but we observe very similar patterns (see Figure 3 in Appendix H).

To illustrate the source of variation and complexity of substitution patterns, we provide an eigenvalue decomposition of the matrix of cross-price effects akin to a principal component analysis (PCA). Table 7 reports the eigenvalue decomposition of the structural cross-price matrices evaluated at sample-mean shares and prices.[21] The contrast is stark. The K-aggregator's first two eigenvalues are of comparable magnitude ($\lambda_1/\lambda_2 \approx 1.4$), and the third drops to zero. The sequential testing in Table 3 selects $K = 2$, and both identified dimensions carry substantial weight. BLP's cross-price matrix, by contrast, is dominated by a single eigenvalue that captures 93% of the total squared eigenvalue magnitude; the second is two orders of magnitude smaller ($\lambda_1/\lambda_2 \approx 98$), and the remaining eigenvalues are negligible. The K-aggregator identifies two dimensions of cross-price substitution; BLP identifies only one.

This rank-one pattern is robust to the choice of instruments: across all seven specifications considered in Appendix H.2 (see Table 14), BLP's leading eigenvalue captures about 93% of the squared eigenvalue magnitude, and the ratio λ_1/λ_2 ranges from 67 to 125.

[21]This corresponds to $\zeta\Gamma\zeta^\top$ for the K-aggregator demand, and for BLP it is matrix Σ defined in Proposition 5 in Appendix, i.e. a weighted sum of the product of market shares of the two goods.

TABLE 7. Eigenvalue decomposition of the structural cross-price matrices

	λ_1	λ_2	λ_3	$\lambda_1^2/\sum\lambda^2$
K-aggregator	+0.107	+0.078	≈ 0	0.657
BLP	+0.307	+0.003	≈ 0	0.934

Notes: Specification with FL instruments and demographics. Eigenvalues of $\zeta\hat{\Gamma}\zeta^\top$ (K-aggregator) and $\hat{\Sigma}$ (BLP), evaluated at sample-mean shares and prices. The third column provides the cumulative variance at K=1 i.e. the share of total squared eigenvalues captured by the leading eigenvalue.

We next compare the models' fit using the Rivers–Vuong test for non-nested GMM models (Rivers and Vuong 2002). Table 8 reports the matched FL specifications. The test statistic is $T_N = -28.4$ ($p < 10^{-4}$), decisively rejecting equal fit in favour of the K-aggregator. The qualitative conclusion is robust, though the gap has narrowed relative to the comparison without demographics (J_N/df ratio of 4.0× versus the earlier 23×).

TABLE 8. Rivers–Vuong comparison: FL instruments with demographics.

	K-aggregator	BLP
J_N	213.47	930.69
Overidentifying restrictions	29	32
J_N/df	7.36	29.08
Number of parameters	21	18

Neither model encompasses the other. Cross-evaluation of moment conditions—each model's residuals tested against the other's instruments—yields decisive rejections on both sides ($p < 10^{-4}$). The per-degree-of-freedom violation is an order of magnitude larger for the K-aggregator residuals tested against BLP's instruments (Wald/df = 99.8 versus 10.5 in the reverse direction), indicating that BLP's random-coefficient structure captures substantial variation—particularly through the demographic interaction instruments—that the rank-2 K-aggregator does not.

7.1.4. *Discussion.* The central finding of the model comparison is that the two approaches identify a different number of dimensions of cross-price substitution. The K-aggregator identifies two—with eigenvalues of comparable magnitude—while BLP identifies only one, statistically, concentrating nearly all cross-price effects onto a single dominant dimension. Sequential testing finds no evidence for $K > 2$ in any K-aggregator specification, and for BLP the estimated rank-one pattern is invariant across all seven instrument sets.

Adding demographic taste shifters Π remains essential for identification. Without Π, the estimated Γ conflates substitution structure with demographic heterogeneity, and the second eigenvalue is not identifiable. With Π, Γ captures the pure substitution structure, the eigenvalue separation sharpens dramatically, and J falls. The price-elasticity parameter is robustly estimated near unity: $\hat{\theta} = 1.18$ (95% CI: 1.05–1.30).

The K-aggregator fits the moment conditions substantially better (J_N/df of 7.4 versus 29.1), and the Rivers–Vuong test decisively favours it. The linearity of the K-aggregator specification—cross-price effects enter linearly once the own-price demand curve is inverted—also contributes to more transparent identification.

Conclusions

In this paper, we characterize demand systems with "simple" cross-price effects, in the sense that they can be described by a matrix of low rank K that is notably smaller than the number of products. This allows for tractability in economic modeling and identification in demand estimation, especially when a large number of products are present in the data. This property holds for most demand systems used in practice, such as logit, CES, directly-additive preferences (with a rank one), mixed and nested logit (with a rank given by the number of nests or types of consumers).

We show that such demands can be expressed as functions of the price of the corresponding good as well as K "aggregators" that are functions of the full vector of prices. We then provide the functional form restrictions for demand and utility imposed jointly by rationality (Slutsky symmetry) and the low-rank assumption. We focus on the case of homothetic preferences but also describe how to extend our results to allow for more complex income effects. Such demand system can also be interpreted as the aggregation of consumers making random choices, and has a convenient representation as a "Perturbed Utility Model".

These results provide a novel approach to understanding functional forms of demand and can be used to construct flexible yet practical demand systems for modeling and estimation. We illustrate our results by proposing a functional form of demand with flexible and potentially non-parametric own-price effects (hence flexible price elasticities, markups and pass-through when applied to firms' decisions) combined with cross-price effects captured parametrically by a single rank-K symmetric matrix to be estimated. Such specification remains flexible enough to allow for complementarity (unlike random utility models) and match the matrix of cross-price effects obtained by any mixed logit demand evaluated at a given set of prices. We also propose to project such price interactions on product attributes, following typical approaches in BLP estimations. Using NielsenIQ scanner data on ready-to-eat cereals, we find that a substitution matrix with rank $K = 2$ provides the best fit, and that adding demographic taste shifters is essential for identification: it stabilizes the price-elasticity parameter, sharpens the eigenvalue separation in the substitution matrix, and reveals interpretable heterogeneity in consumer valuations of product characteristics.

Our approach contrasts with the standard BLP in various ways: i) the functional form is simpler, and substitution effects enter linearly once we invert the demand curve capturing own-price effects; ii) at a given set of tastes and prices, the cross-price effects are more flexible than random-utility models as they allow for complementarity; iii) estimation is faster and optimal instruments are easier to compute, thanks to the linearity of the moment restrictions; iv) instruments used in BLP estimation (e.g. Gandhi and Houde 2019) are also useful with this approach; v) we provide the utility functions and price indices for a representative consumer, which simplifies the analysis of consumer welfare. Estimating both models on the same data with matched instruments and demographic controls, we find that BLP substitution effects, evaluated at sample mean, can be well approximated by an even more simple matrix of lower rank. A Rivers–Vuong test decisively favours the K-aggregator specification, though an encompassing analysis reveals that neither model fully captures the variation identified by the other's instruments.

References

Allen, Roy and John Rehbeck (2019). "Identification with additively separable heterogeneity". In: *Econometrica* 87.3, pp. 1021–1054.

Backus, Matthew, Christopher Conlon, and Michael Sinkinson (2021). "Common ownership and competition in the ready-to-eat cereal industry". In.

Bai, Jushan and Serena Ng (2002). "Determining the number of factors in approximate factor models". In: *Econometrica* 70.1, pp. 191–221.

Baker, Christopher G, P-A Absil, and Kyle A Gallivan (2008). "An implicit trust-region method on Riemannian manifolds". In: *IMA journal of numerical analysis* 28.4, pp. 665–689.

Berry (1994). "Estimating discrete-choice models of product differentiation". In: *The RAND Journal of Economics*, pp. 242–262.

Berry, James Levinsohn, and Ariel Pakes (1995). "Automobile prices in market equilibrium". In: *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Blundell, Richard and Jean-Marc Robin (2000). "Latent separability: Grouping goods without weak separability". In: *Econometrica* 68.1, pp. 53–84.

Borusyak, Kirill, Mauricio Caceres Bravo, and Peter Hull (2025). "Estimating demand with recentered instruments". In: *arXiv preprint arXiv:2504.04056*.

Boyd, J Hayden and Robert E Mellman (1980). "The effect of fuel economy standards on the US automotive market: an hedonic demand analysis". In: *Transportation Research Part A: General* 14.5-6, pp. 367–378.

Broda, Christian and David E Weinstein (2006). "Globalization and the Gains from Variety". In: *The Quarterly journal of economics* 121.2, pp. 541–585.

Chandrasekaran, Venkat et al. (2009). "Sparse and low-rank matrix decompositions". In: *IFAC Proceedings Volumes* 42.10, pp. 1493–1498.

Combettes, Patrick L (2018). "Perspective functions: Properties, constructions, and examples". In: *Set-Valued and Variational Analysis* 26, pp. 247–264.

Combettes, Patrick L and Christian L Müller (2018). "Perspective functions: Proximal calculus and applications in high-dimensional statistics". In: *Journal of Mathematical Analysis and Applications* 457.2, pp. 1283–1306.

Conlon, Christopher and Jeff Gortmaker (2020). "Best practices for differentiated products demand estimation with pyblp". In: *The RAND Journal of Economics* 51.4, pp. 1108–1161.

– (2025). "Incorporating micro data into differentiated products demand estimation with PyBLP". In: *Journal of Econometrics*, p. 105926.

Deaton, Angus and John Muellbauer (1980). "An almost ideal demand system". In: *The American economic review* 70.3, pp. 312–326.

Fally, Thibault (2022). "Generalized separability and integrability: Consumer demand with a price aggregator". In: *Journal of Economic Theory* 203, p. 105471.

Fosgerau, Mogens, Julien Monardo, and André De Palma (2024). "The inverse product differentiation logit model". In: *American Economic Journal: Microeconomics* 16.4, pp. 329–370.

Fosgerau, Mogens, Mads Paulsen, and Thomas Kjær Rasmussen (2022). "A perturbed utility route choice model". In: *Transportation Research Part C: Emerging Technologies* 136, p. 103514.

Gandhi, Amit and Jean-Francois Houde (2019). "Measuring substitution patterns in differentiated-products industries". In: *NBER Working paper* w26375.

Goldman, Steven M and Hirofumi Uzawa (1964). "A note on separability in demand analysis". In: *Econometrica: Journal of the Econometric Society*, pp. 387–398.

Greene, William H and Terry G Seaks (1991). "The restricted least squares estimator: a pedagogical note". In: *The Review of Economics and Statistics*, pp. 563–567.

Lewbel, Arthur (1991). "The rank of demand systems: theory and nonparametric estimation". In: *Econometrica* 59.3, pp. 711–730.

Lewbel, Arthur and Krishna Pendakur (2009). "Tricks with Hicks: The EASI demand system". In: *The American Economic Review* 99.3, pp. 827–863.

Ligon, Ethan (2025). `ManifoldGMM`*: GMM Estimation on Riemannian Manifolds*. DOI: `10.5281/zenodo.17508879`. URL: `https://doi.org/10.5281/zenodo.17508879`.

Matsuyama, Kiminori and Philip Ushchev (2017). "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems". In: *Working paper*.

McFadden, Daniel and Kenneth Train (2000). "Mixed MNL models for discrete response". In: *Journal of applied Econometrics* 15.5, pp. 447–470.

McFadden, Daniel L and Mogens Fosgerau (2012). "A theory of the perturbed consumer with general budgets". In.

Nevo, Aviv (2001). "Measuring market power in the ready-to-eat cereal industry". In: *Econometrica* 69.2, pp. 307–342.

Rivers, Douglas and Quang Vuong (2002). "Model selection tests for nonlinear dynamic models". In: *The Econometrics Journal* 5.1, pp. 1–39.

Salanié, Bernard and Frank A Wolak (2022). "Fast, detail-free, and approximately correct: Estimating mixed demand systems". In.

Sheng, Zhou and Gonglin Yuan (2024). "Convergence and worst-case complexity of adaptive Riemannian trust-region methods for optimization on manifolds". In: *Journal of Global Optimization*, pp. 1–26.

Townsend, James, Niklas Koep, and Sebastian Weichwald (2016). "Pymanopt: A Python Toolbox for Optimization on Manifolds using Automatic Differentiation". In: *Journal of Machine Learning Research* 17.137, pp. 1–5. URL: `http://jmlr.org/papers/v17/16-177.html`.

Appendix

Appendix A. Proposition 1

Conditions for Proposition 1. Here we state conditions for Proposition 1 more formally:

- C1 **(Rank)** At any price x, we assume that matrix $\frac{\partial F_i}{\partial x_j}(x)$ is the sum of a diagonal matrix $\sigma(x)$ with non-zero diagonal elements $\sigma_i(x)$ and a matrix $\Sigma(x)$ of rank K. Without loss of generality, we assume that $\sigma_i(x) < 0$ for each i and x.
- C1' **(Identifying goods)** Furthermore, we assume that there are K goods $k = 1, ..., K$ such that the truncated matrix with coefficients $\Sigma_{kj} = \left\{ \frac{\partial F_k}{\partial x_j} - \sigma_k \mathbb{1}(j = k) \right\}$, $k \leq K$, still has rank K for all x.
- C2 **(Stability of the rank)** We assume that the rank of the "substitution matrix" is preserved when we drop some goods. In particular, we assume that:
  - (i) The rank of the "substitution matrix" Σ remains K when we drop $K + 1$ rows, corresponding to the price of all goods $k \leq K$ and any good $i > K$.
  - (ii) The Jacobian of $(F_1(x), ..., F_K(x), x_1, ..., x_K, x_i, x_j)$ has full rank $2K + 2$, when we pick any two goods $i > K$ and $j > K$, $i \neq j$.
- C3 **(No escaping)** For any sequence $x(t)$ such that $\max_{k \leq K} |F_k(x(t))| \to \infty$ while $x_1(t)...x_K(t)$ remains bounded, some other values of F are unbounded, i.e., $\max_{j>K} |F_j(x(t))| \to \infty$.
- C4 **(Connectedness)** We assume that level sets of $(F_1(x), ..., F_K(x), x_1, ..., x_K)$ are connected, and remain also connected when we condition for the price of another good x_j.

Proposition 1. *Under assumptions [C1], [C1'], [C2], [C3] and [C4], there exist K real functions $\Lambda_1(x)$, ... , $\Lambda_K(x)$ and J functions $S_i(\Lambda, x_i)$ such that:*

$$F_i(x) \; = \; S_i\left(x_i \, , \, \Lambda_1(x) \, , ..., \, \Lambda_K(x)\right)$$

i.e., demand for each good can be reduced to a function of its own price and K "aggregators" $\Lambda_k(x)$.

Proof of Proposition 1.

Step 1: functions of K shares and prices. The first step is to show that for each good j, we can write it as a function of its own price x_j and of the first K prices x_k and demand F_k.

Denote by $\Sigma|_K$ the Σ matrix where we include only the first K rows, so that we include only the first K goods $k = 1, ..., K$. Define $\partial F|_K$ and $\sigma|_K$ similarly. We have then: $\partial F|_K \; = \; \sigma|_K + \Sigma|_K$.

The remaining rows of $\partial F - \sigma$ are then linear combinations of the first ones, denoting B_{JK} the matrix that yields the remaining rows depending on Σ_K. Thus, we can then define matrix

$$B \; = \; \begin{pmatrix} \mathbb{1}_K \\ B_{JK} \end{pmatrix}$$

where $\mathbb{1}_K$ is the identity matrix on $\mathbb{R}^K$, so that matrix B has rank K and we have then:

$$\partial F \; = \; \sigma \; + \; B \; \cdot \; \left(\partial F|_K \; - \; \sigma|_K\right)$$

This means that the gradient of each F_i is a linear combination of the gradients of x_i, the gradients of x_1, ... , x_K as well as the gradients of $F_1, ..., F_K$:

$$\frac{\partial F_i}{\partial x_j} \; = \; \sigma_i \mathbb{1}(j = i) \; + \; \sum_k B_{ik} \left[\frac{\partial F_k}{\partial x_j} - \sigma_k \mathbb{1}(j = k) \right]$$

Combined with the connectedness assumption [C4], Lemma 1 of Goldman-Uzawa (1964) [Lemma reproduced above] implies that for each good i there exists a function g_i of $2K + 1$ arguments such that we can write:

$$F_i(x) \; = \; g_i\big(x_i, F_1(x), ..., F_K(x), x_1, ..., x_K\big)$$

Step 2: σ_k and the derivatives of g_i. The rank assumption [C2-ii] on the Jacobian of the arguments of g_i implies that the coefficients in the expression above coincide with the derivatives of function g_i.

First, the derivative in the first argument is $\frac{\partial g_i}{\partial x_i} = \sigma_i$, which implies that σ_i can be written as a function of $\big(x_i, F_1(x), ..., F_K(x), x_1, ..., x_K\big)$ for any $i > K$. Regarding the terms in F_k, we must have:

$$\frac{\partial g_i}{\partial F_k} = B_{ik}$$

so that B_{ik} can be itself written as a function of $\big(x_i, F_1(x), ..., F_K(x), x_1, ..., x_K\big)$. For the derivatives in x_k, we obtain:

$$\frac{\partial g_i}{\partial x_k} = -\sigma_k B_{ik}$$

Hence $\frac{\partial g_i}{\partial x_k}$ can also be written as functions of $\big(x_i, F_1(x), ..., F_K(x), x_1, ..., x_K\big)$. Taking ratios, σ_k can in turn be written as functions of $\big(x_i, F_1(x), ..., F_K(x), x_1, ..., x_K\big)$. This can be done for all i such that $B_{ik} \neq 0$ is non-null.

Under assumption [C2-i], we must have at least two goods $i > K$ and $j > K$ such that $B_{ik} \neq 0$ and $B_{jk} \neq 0$, i.e. $\frac{\partial g_i}{\partial F_k} \neq 0$ and $\frac{\partial g_j}{\partial F_k} \neq 0$. If that wasn't the case, e.g. if all the B_{ik} cells are zero for a specific k, then one of the columns of the B_{JK} matrix would be null, and its rank would be strictly smaller than K after we remove the $k \leq K$ rows. If the B_{ik} cells are all zero for a specific k, aside from a single i good, then the same reasoning would apply if we remove the row corresponding to that good i. Hence there exist at least two non-zero entries B_{ik} and B_{jk} for any specific column k.

We then obtain that σ_k can be written both as function of $\big(x_i, F_1(x), ..., F_K(x), x_1, ..., x_K\big)$ and as a function of $\big(x_j, F_1(x), ..., F_K(x), x_1, ..., x_K\big)$. Under assumption [C2-ii], linear independent of their Jacobians implies that σ_k only depends on the common set of arguments, i.e. $\big(F_1(x)...F_K(x), x_1...x_K\big)$. We use this result in the next step.

Step 3: defining flows Ψ_k. For any good $k \leq K$, define the "flow" $\Psi_k\big(t\big|F_1, ..., F_K, x_1, ..., x_K\big)$ as a function from $\mathbb{R} \times \mathbb{R}^{2K}$ to $\mathbb{R}^{2K}$ such that:

$$\Psi_k\big(0\big|F_1, ..., F_K, x_1, ..., x_K\big) = \big(F_1, ..., F_K, x_1, ..., x_K\big)$$

and:

$$\frac{\partial \Psi_{k,xk'}}{\partial t} = 0 \quad ; \quad \frac{\partial \Psi_{k,Fk'}}{\partial t} = 0$$

whenever $k' \neq k$, i.e. t does not lead to changes in values w.r.t coordinates $F_{k'}$ and $x_{k'}$ for $k' \neq k$), while:

$$\frac{\partial \Psi_{k,xk}}{\partial t} = 1 \quad ; \quad \frac{\partial \Psi_{k,Fk}}{\partial t} = \sigma_k\big(\Psi(t)\big) < 0$$

which means that as time t increases we also increase the price x_k of good k and its corresponding demand F_k by a corresponding amount $\sigma_k < 0$. In this transformation, since the sign of σ_k does not change, there is a unique coordinate in F_k for each coordinate x_k.

This is an ordinary differential equation that admits a solution. Moreover, it is defined for all $t \in \mathbb{R}$. If the flow Ψ_k was defined only up to an upperbound T, it would leave any compact as t approaches T. Since x would remain bounded (x_i remains constant for any $i \neq k$ and x_k would increase by at most T from $t = 0$), and F_i remains constant for any $i \neq k$, it must be that $|F_k|$ goes to infinity for some k. Under assumption [C3], this leads to a contradiction as other F_i could then not remain constant.

Alternatively, note that the flows would be also globally defined under the alternative assumption that σ_k be bounded.

Step 4: invariance property. We can then check that each demand function g_j are invariant to transformations along flows Ψ_k. Take any good j and define:

$$F_j(t) = g_j\big(x_{j0}, \Psi_k\big(t\big|F_1, ..., F_K, x_1, ..., x_K\big)\big)$$

where we hold the first argument fixed to some arbitrary value x_{j0}. We obtain, using the derivatives of flow Ψ_k:

$$\frac{\partial F_j}{\partial t} = \sigma_k(\Psi_k)\frac{\partial g_i}{\partial F_k}(\Psi_k) + \frac{\partial g_i}{\partial x_k}(\Psi_k) = 0$$

which is equal to zero given the earlier finding on the derivatives $\frac{\partial g_i}{\partial x_k}$ and $\frac{\partial g_i}{\partial F_k}$.

Step 5: commutative frame. We can then show that the set of flows $\{\Psi_k\}_{k\leq K}$ commute, i.e. it is equivalent to shift first by t_k using the flow Φ_k and then shift by $t_{k'}$ using flow $\Phi_{k'}$ or vice versa:

$$\Psi_{k'}\big(+t_{k'}\big|\Psi_k\big(+t_k\big|\{F,x\}_K\big)\big) = \Psi_k\big(+t_k\big|\Psi_{k'}\big(+t_{k'}\big|\{F,x\}_K\big)\big)$$

where $\{F,x\}_K \equiv (F_1,...,F_K,x_1,...,x_K) \in \mathbb{R}^{2K}$.

The reason for this is that these flows remain on the same level sets of functions g_i's, holding constant their first argument (thus, they satisfy the integrability condition of Frobenius Theorem). Take some fixed values $\bar{x}_i$ for $i > K$ and define function G from $\mathbb{R}^{2K}$ to $\mathbb{R}^{J-K}$ as:

$$G\big(\{F,x\}_K\big) \;=\; \big\{g_i(\bar{x}_i,\{F,x\}_K\big\}_{i>K}$$

Three remarks about the Jacobian are then relevant: i) first, the Jacobian of G in F (holding x constant) coincides with matrix B_{JK} defined earlier, which has a rank K; ii) each column of the Jacobian of G in x (keeping F constant) is colinear with the k'th column of B, hence again the Jacobian of G in x; iii) given the colinearity, the rank of G in (F,x) combined is exactly K again.

Hence we can apply the "constant-rank level set theorem" (see e.g. Lee 2003, chapter 5, theorem 5.1) telling us that each level set of G in $\{F,x\}$ is a properly embedded submanifold of dimension K in $\mathbb{R}^{2K}$. Moreover, since the differential in F has rank K, locally there is a unique F on the level set, given x. Similarly, there is a unique x on the level set, conditional on F. Hence, locally, there is a one-to-one mapping between x and F such that $\{F,x\}$ remains on the level set.

Then, notice that the x coordinates of $\Psi_{k'}\big(+t_{k'}\big|\Psi_k\big(+t_k\big|\{F,x\}_K\big)\big)$ are the same as those of $\Psi_k\big(+t_k\big|\Psi_{k'}\big(+t_{k'}\big|\{F,x\}_K\big)\big)$ since we are both shifting x_k by $+t_k$ and $x_{k'}$ by $+t_{k'}$ while keeping other x's constant. Since both remain on the same level set of G, the remaining coordinates in F must be the same in both cases, showing that the two flows commute.[22]

We can then simply define $\Psi\big(t_1,...,t_K\big|\{F,x\}_K\big)$ as the combination of shifts along each flow Ψ_k by t_k without referring to the ordering. Using this "commutative frame", it is then easy to define aggregators and demand as a function of these aggregators.

Step 6: aggregators and demand function. We define aggregators Λ as the F component of Ψ, where we shift the initial $\{F,x\}_K$ by $t=-x$:

$$\big(\Lambda_1(x),...,\Lambda_K(x),0,...0\big) \;=\; \Psi\big(-x_1,...,-x_K\big|\big(F_1(x),...,F_K(x),x_1,...,x_K\big)\big)$$

or equivalently defining each Λ_k as: $\Lambda_k(x) \;=\; \Psi_{k,Fk}\big(-x_k\big|\big(F_1(x),...,F_K(x),x_1,...,x_K\big)\big)$. By pushing back by $+x_k$ (inverting the flow), notice that we have:

$$\big(F_1(x),...,F_K(x),x_1,...,x_K\big) \;=\; \Psi\big(x_1,...,x_K\big|\big(\Lambda_1(x),...,\Lambda_K(x),0,...0\big)\big)$$

so we can express each F_k as a function of x_k and the Λ's:

$$F_k(x) \;=\; \Psi_{Fk}\big(x_k\big|\big(\Lambda_1(x),...,\Lambda_K(x),0,...0\big)\big)$$

This gives Proposition 1 for goods $k \leq K$. where the D_k function is given by Ψ_{Fk}.

For goods $i > K$, using the invariance property (step 4), we can see that we can replace each demand F_k as a argument by $\Lambda_k(x)$ and each price x_k by 0:

$$\begin{aligned} F_i(x) &= g_i\big(x_i,F_1(x),...,F_K(x),x_1,...,x_K\big) \\ &= g_i\big(x_i,\Psi\big(-x_1,...,-x_K\big|\big(F_1(x),...,F_K(x),x_1,...,x_K\big)\big)\big) \\ &= g_i\big(x_i,\Lambda_1(x),...,\Lambda_K(x),0,...,0\big) \end{aligned}$$

Denoting D_i the function g_i by dropping the $0,...0$ argument, we obtain the result from Proposition 1, i.e. that we can express demand as a function of its own price x_i and K aggregators:

$$F_i = D_i\big(x_i,\Lambda_1(x),...,\Lambda_K(x)\big) \equiv g_i\big(x_i,\Lambda_1(x),...,\Lambda_K(x),0,...,0\big)$$

[22] Note that this argument is local, but flows that commute locally also commute globally.

Lemma from Goldman and Uzawa (1964). Take a smooth function f and several smooth functions $\{g_k\}_{k\leq K}$, each from $\mathbb{R}^J$ to $\mathbb{R}$. Suppose that there are scalar functions λ_k, $k \leq K$, from $\mathbb{R}^J$ to $\mathbb{R}$ such that:

$$\nabla f(x) \ = \ \sum_k \lambda_k(x) \,.\, \nabla g_k(x)$$

for any x, and suppose that each level set of $\{g_k\}_{k\leq K}$ is connected, then there exists a function G from $\mathbb{R}$ to $\mathbb{R}$ such that:

$$f(x) \ = \ G\big(g_1(x), ..., g_K(x)\big)$$

Proof of the lemma. It suffices to show that $g_k(x_0) = g_k(x_1)$ (for all k) implies $f(x_0) = f(x_1)$ for any $x_0 \in \mathbb{R}^J$ and $x_1 \in \mathbb{R}^J$, i.e. that f only takes a single value on a level set of the set of g. Take x_0 and x_1 on such a level set. Since they are on the same level set, following our assumption on connectedness, there is a smooth path $x(t)$ between these two points that remains on the level set, with $x(0) = x_0$ and $x(1) = x_1$. Along that path, define $\phi(t) = f(x(t))$. We have then:

$$\phi'(t) \ = \ \nabla f(\phi(t)) \cdot x'(t) \ = \ \lambda(\phi(t)) \,.\, \sum_k \nabla g_k(\phi(t)) \cdot x'(t) \ = \ 0$$

which is null since each $\nabla g_k(\phi(t)) \cdot x'(t) \ = \ 0$ as $x(t)$ remains on the level set of each g_k. Hence $\phi(1) = \phi(0)$, which means that $f(x_1) = f(x_0)$.

Appendix B. Corollary 1

Here we provide a generalization of Proposition 1. When we do not have a constant set of K goods with cross-price effects of rank K, the K aggregators are Euclidean only locally. Assuming that the rank assumptions hold only locally, i.e., dropping [C1'] and the constant set of goods, we can build on Proposition 1 to define aggregators on smaller subsets that are locally Euclidean. Then, by patching these subsets together, we obtain a manifold, which actually provides a more natural (but more abstract) mathematical environment for price aggregators. Instead of condition [C1'], we rely on these two conditions.

[C5] Constant rank on level sets. *If the substitution matrix has a rank K on K goods $k \in \mathcal{K}$, then the rank remains K on the level set of $\{F_k(x), x_k\}_{\in\mathcal{K}}$*

[C6] Span of projection. *Such a level set of $\{F_k(x), x_k\}_{k\in\mathcal{K}}$ spans $\mathbb{R}$ on each other coordinate $j \notin \mathcal{K}$, i.e. for each $\bar{x} \in \mathbb{R}^N$, $\bar{y}_j \in \mathbb{R}$, $\exists y \in \mathbb{R}^N$ such that $y_j = \bar{y}_j$ and $(F_k(y), y_k) = F_k(\bar{x}), \bar{x}_k)$ for all $k \in \mathcal{K}$.*

We obtain a generalization of Proposition 1, applying similar arguments locally, where aggregators Λ are now defined on a more general K-dimensional manifold instead of $\mathbb{R}^K$:

Corollary 1. *Under conditions [C1]-[C6] (without C1'), there exists a smooth map $\Lambda(x)$ to a K-dimensional manifold and some smooth functions $S_i(x, \Lambda)$ such that: $F_i(x) = S_i\big(x_i\,, \Lambda(x)\big)$.*

Proof.

Step 1: Local constructions. Take open set U on which assumptions from Proposition 1 hold locally on U for $k \in \mathcal{K}$, a subset of $\{1, ..., N\}$ with K elements. Denote $F_\mathcal{K}(x)$ the tuplet $\{F_k(x)\}_{k\in\mathcal{K}}$.

With the added assumption, level set of $(\{F_\mathcal{K}(x), x_\mathcal{K}\}$ is included in U and is connected, hence:

$$F_i(x) \;=\; g_i^U\Big(x_i, F_\mathcal{K}(x), x_\mathcal{K}\Big)$$

on $x \in U$ for any $i \notin \mathcal{K}$. As in Proposition 1, we also get that σ_k only depends on $\Big(F_\mathcal{K}(x), x_\mathcal{K}\Big)$ and that:

$$\frac{\partial g_i^U}{\partial x_k} \;=\; -\sigma_k \frac{\partial g_i^U}{\partial F_k}$$

for any $k \in \mathcal{K}$ and $i \notin \mathcal{K}$. Based on the same arguments as for Proposition 1, for each $k \in \mathcal{K}$ we can then construct the "flow" $\Psi_k\Big(t\Big|F_\mathcal{K}, x_\mathcal{K}\Big)$ as a function from $\mathbb{R} \times \mathbb{R}^{2K}$ to $\mathbb{R}^{2K}$ such that:

$$\frac{\partial \Psi_{k,xk}}{\partial t} \;=\; 1 \qquad ; \qquad \frac{\partial \Psi_{k,Fk}}{\partial t} \;=\; \sigma_k\Big(\Psi_k(t)\Big) \;<\; 0$$

and zero for $k' \neq k$, and we also obtain that these flows commute. These flows are defined as least locally, and each g_j is invariant under Ψ, as in Proposition 1.

Using our new assumption (which replaces the "no-escaping" assumption from Proposition 1), again each flow Ψ_k is global, i.e. defined for all t.

Step 2: Extended function. Using Ψ, we can now define an extended function $\widetilde{F}(\bar{x}, x)$ as a function of two arguments:

$$\widetilde{F}_j(\bar{x}, x) \;=\; g_j\big(\bar{x}_j, \Psi(-x_K|(F_K(x), x_K))\big) \;=\; g_j\big(\bar{x}_j, F_K(x), x_K)\big)$$

for $j > K$, where $\Psi(-x_K|(F_K, x_K))$ coincides to the previous definition of $\Lambda(x)$.

For $k \leq K$, we define $\widetilde{F}_k(\bar{x}, x)$ as:

$$(\widetilde{F}_K(\bar{x}, x))\,,\, \bar{x}_K) \;=\; \Psi\Big(\bar{x}_K\Big|\Psi(-x_K|(F_K, x_K))\Big) \;=\; \Psi\Big(\bar{x}_K - x_K\Big|(F_K, x_K))\Big)$$

This functions $\widetilde{F}_i(\bar{x}, x)$ correspond to $S_i(\bar{x}_i, \Lambda(x))$ in Proposition 1. Conditional on $\bar{x}$, they depend only on $\bar{x}_i$. They are defined for each $x \in U$ but also for each $\bar{x} \in \mathbb{R}^N$ since the flows are defined globally.

We have defined $\widetilde{F}_i(\bar{x}, x)$ locally on an open U. Next step is to show that it is globally defined if we combined open sets on $\mathbb{R}^N$. Specifically, suppose that $F_i(x) = S_i(x_i, \Lambda(x))$ for each $x \in U$, where $\Lambda(x) \in \mathbb{R}^K$ are K aggregators, and where we define $\widetilde{F}_i(\bar{x}, x) = S_i(\bar{x}_i, \Lambda(x))$ as above. Also, suppose that $F_i(x) = V_i(x_i, L(x))$ for each $x \in V$, where $L(x) \in \mathbb{R}^K$ are potentially some other K aggregators. Consider a non-empty intersection $U \cap V$, where we want to show that $S_i(\bar{x}_i, \Lambda(x)) = V_i(\bar{x}_i, L(x))$ for any $\bar{x} \in \mathbb{R}^N$ and $x \in U \cap V$.

Take a good j. Up to a reordering, suppose that we have K goods in which the substitution matrix has full rank K, with $j > K$. Consider the level set of $F_K(x), x_K$ based on the first K goods. If that level set spans $x_j \in \mathbb{R}$, then we can find x' in that level set such that $x'_j = \bar{x}_j + T$, $F_k(x') = F_k(\bar{x})$ and $x'_k = \bar{x}_k$ for each $k \leq K$, as well as a smooth path on that level set such that $x(T) = x'$ and $x(0) = \bar{x}$ (combining our connectedness assumption with assumption [C6]).

We have then: $F_i(x') = S_i(\bar{x}_i + T, \Lambda(x')) = V_i(\bar{x}_i + T, L(x'))$. Then the goal is to show that $\Lambda(x') = \Lambda(x)$ and $L(x') = L(x)$. By symmetry, showing only one of these equality is sufficient.

On that path, we have:

$$S_k\Big(\bar{x}_k, \Lambda(x(t))\Big) = S_k\Big(\bar{x}_k, \Lambda(\bar{x})\Big)$$

since $x_k(t) = \bar{x}_k$ and $F(x_k(t)) = F_k(\bar{x})$, for all $k \leq K$. The rank of W in Λ is K while the dimension of Λ is K, hence Λ must remain constant. This proves that we can consistently define $\widetilde{F}_i(\bar{x}, x) = S_i(\bar{x}_i, \Lambda(x))$ or $\widetilde{F}_i(\bar{x}, x) = V_i(\bar{x}_i, L(x))$ on $U \cap V$ and that the resulting function is smooth globally once we combine all open sets.

Step 3: Aggregators. Then, notice that the rank of the Jacobian of $\widetilde{F}_i(\bar{x}, x)$ in x is constant and equal to K. Hence, holding $\bar{x}$, the image of $\widetilde{\Lambda}(x) = \widetilde{F}(\bar{x}, x)$ (i.e. holding $\bar{x}$) is a K-dimensional manifold.

Conversely, we can retrieve:

$$(F_K(x)\,,\, x_K) \;=\; \Psi\Big(x_K - \bar{x}_K\Big|\Psi\Big(\bar{x}_K\Big|\Psi(-x_K|(F_K, x_K))\Big)\Big) \;=\; \Psi\Big(x_K - \bar{x}_K\Big|\widetilde{F}_K(\bar{x}, x))\Big)$$

so that for each individual $k \leq K$ we have: $F_k(x) \;=\; \Psi_{Fk}\Big(x_k - \bar{x}_k\Big|\widetilde{F}_K(\bar{x}, x))\Big)$

For $j > K$, we have:

$$F_j \;=\; g_j\big(x_j, F_K, x_K\big) \;=\; g_j\Big(x_j, \Psi\Big(x_K - \bar{x}_K\Big|\widetilde{F}_K(\bar{x}, x)), \bar{x}_K\Big)\Big) \;=\; g_j\Big(x_j, \widetilde{F}_K(\bar{x}, x), \bar{x}_K\Big)$$

Note: for the same reasons $\widetilde{F}$ is uniquely defined across open neighborhoods U and V, we can uniquely retrieve F on U and V.

Appendix C. Proof of Proposition 2

Here, to lighten the notation, we denote $x_i = \log p_i$.

Step 1: Separability. We start from the integrability assumption:

$$\frac{\partial P}{\partial x_i} = S_i(x_i, \Lambda(x)) \tag{33}$$

Denote $W_i(x_i, \Lambda) = \int_0^{x_i} S_i(t, \Lambda)dt$ a primitive of S_i in x_i, and denote:

$$W(x) = \sum_j W_j(x_i, \Lambda(x))$$

the sum of W_i's, evaluated at $\Lambda = \Lambda(x)$.

We have then:

$$\frac{\partial W}{\partial x_i} = \frac{\partial P}{\partial x_i} + \sum_k \left(\sum_j \frac{\partial W_j}{\partial \Lambda_k} \right) \frac{\partial \Lambda_k}{\partial x_i} \tag{34}$$

Hence the gradient of $W - P$ is colinear with the gradients of the aggregators Λ. Using Lemma 1, we can thus express $W - P$ as a function of the aggregators, using also the assumption that iso-Λ surfaces are connected. Hence:

$$W(x) - P(x) = M(\Lambda(x)) \tag{35}$$

for some function M, and thus:

$$\sum_j W_j(x_i, \Lambda(x)) = M(\Lambda(x)) + P(x) \tag{36}$$

Step 2: Implication of the rank of $\frac{\partial \Lambda_k}{\partial x_i}$. Differentiating the last equality above, and using $\frac{\partial P}{\partial x_i} = \frac{\partial W_i}{\partial x_i}$, we have:

$$\sum_k \left(\sum_j \frac{\partial W_j}{\partial \Lambda_k} \right) \frac{\partial \Lambda_k}{\partial x_i} = \sum_k \left(\frac{\partial M}{\partial \Lambda_k} \right) \frac{\partial \Lambda_k}{\partial x_i}$$

Given the assumption that the collection of vectors $\frac{\partial \Lambda_k}{\partial x_i}$ has rank K, it must be that:

$$\sum_j \frac{\partial W_j}{\partial \Lambda_k} = \frac{\partial M}{\partial \Lambda_k} \tag{37}$$

for each Λ_k.

Step 3: Symmetry and invertibility. Differentiating the last equality above (FOC in Λ_k) w.r.t x_i, we obtain:

$$\sum_j \frac{\partial^2 W_j}{\partial x_i \partial \Lambda_k} + \sum_{k'=0} \sum_j \frac{\partial^2 W_j}{\partial \Lambda_{k'} \partial \Lambda_k} \frac{\partial \Lambda_{k'}}{\partial x_i} = \sum_{k'=0} \frac{\partial^2 M}{\partial \Lambda_{k'} \partial \Lambda_k} \frac{\partial \Lambda_{k'}}{\partial x_i}$$

But notice that $\sum_j \frac{\partial^2 W_j}{\partial x_i \partial \Lambda_k} = \frac{\partial S_i}{\partial \Lambda_k}$ so

$$\frac{\partial S_i}{\partial \Lambda_k} + \sum_{k'=1} \sum_j \frac{\partial^2 W_j}{\partial \Lambda_{k'} \partial \Lambda_k} \frac{\partial \Lambda_{k'}}{\partial x_i} = \sum_{k'=1} \frac{\partial^2 M}{\partial \Lambda_{k'} \partial \Lambda_k} \frac{\partial \Lambda_{k'}}{\partial x_i}$$

This equality can be rewritten as:

$$\frac{\partial S_i}{\partial \Lambda_k} = \sum_{k'=1} \mathcal{H}_{kk'} \frac{\partial \Lambda_{k'}}{\partial x_i} \tag{38}$$

where $\mathcal{H}$ is a symmetric $K \times K$ matrix (given the symmetry of the cross derivatives) with coefficients:

$$\mathcal{H}_{kk'} = \frac{\partial^2 M}{\partial \Lambda_{k'} \partial \Lambda_k} - \sum_j \frac{\partial^2 W_j}{\partial \Lambda_{k'} \partial \Lambda_k} \tag{39}$$

Invertibility. Since we also assume that $\left\{\frac{\partial S_i}{\partial \Lambda_k}\right\}$ has full rank K, we obtain that $\mathcal{H}$ matrix must be invertible, with its inverse denoted by $\mathcal{H}^{-1}$ (also symmetric). It implies that the gradients of Λ must be a linear combination of the partial derivatives of S_i:

$$\frac{\partial \Lambda_k}{\partial x_i} = \sum_{k'=1} \mathcal{H}^{-1}_{kk'} \frac{\partial S_i}{\partial \Lambda_{k'}} \tag{40}$$

Step 4: Using the budget constraint. The budget constraint condition imposes $\sum_i S_i(x_i, \Lambda(x)) = 1$ when the aggregators Λ are evaluated at x. Differentiating, we find:

$$-\frac{\partial S_j}{\partial x_j} = \sum_{k=1} \left(\sum_i \frac{\partial S_i}{\partial \Lambda_k}\right) \frac{\partial \Lambda_k}{\partial x_j} \tag{41}$$

Hence $\frac{\partial S_j}{\partial x_j}$ is colinear with the set of vectors $\frac{\partial \Lambda_k}{\partial x_j}$. With $\mathcal{H}^{-1}_{kk'}$ the coefficients of the inverse of $\mathcal{H}$, we obtain:

$$-\frac{\partial S_j}{\partial x_j} = \sum_{k'} \sum_k \left(\sum_i \frac{\partial S_i}{\partial \Lambda_k}\right) \mathcal{H}^{-1}_{kk'} \frac{\partial S_j}{\partial \Lambda_{k'}}$$

Dividing by S_j, and denoting

$$v_k = \sum_{k'} \mathcal{H}^{-1}_{kk'} \left(\sum_i \frac{\partial S_i}{\partial \Lambda_{k'}}\right), \tag{42}$$

we have then:

$$-\frac{\partial \log S_j}{\partial x_j} = \sum_k v_k \frac{\partial \log S_j}{\partial \Lambda_k} \tag{43}$$

where the last derivatives are taken by holding V and x_i constant, respectively. This colinearity between derivatives of S_j is crucial to obtain the functional form for S_j. Before getting into integrating this differential equation, some additional work on v_k is needed.

Step 5: Coefficients v_k as functions of Λ. Take the derivative of the above equation with respect to p_i for a good $i \neq j$. Since $\frac{\partial \log S_j}{\partial x_j}$ and $\frac{\partial \log S_j}{\partial \Lambda_k}$ only depends on the aggregators and x_j, we obtain:

$$-\sum_{k'} \frac{\partial^2 \log S_j}{\partial x_j \partial \Lambda_{k'}} \frac{\partial \Lambda_{k'}}{\partial x_i} = \sum_k \frac{\partial v_k}{\partial x_i} \frac{\partial \log S_j}{\partial \Lambda_k} + \sum_{k,k'} v_k \frac{\partial^2 \log S_i}{\partial \Lambda_k \partial \Lambda_{k'}} \frac{\partial \Lambda_{k'}}{\partial x_i} \tag{44}$$

Rearranging, and multiplying by S_j, we get:

$$-\sum_{k'} \frac{\partial v_{k'}}{\partial x_i} \frac{\partial S_j}{\partial \Lambda_{k'}} = \sum_{k'} \frac{\partial \Lambda_{k'}}{\partial x_i} \mathcal{B}_{jk'} \tag{45}$$

where for some $\mathcal{B}_{jk'} = \sum_k v_k \frac{\partial^2 S_j}{\partial \Lambda_k \partial \Lambda_{k'}} + \frac{\partial^2 S_j}{\partial x_j \partial \Lambda_{k'}}$ (this notation $\mathcal{B}$ will not appear again).

As we assume that $\frac{\partial S_i}{\partial \Lambda_k}$ have full rank K even if we drop a good j, we obtain that the gradients $\frac{\partial v_k}{\partial x_i}$ are colinear with the collection of gradients $\left\{\frac{\partial \Lambda_{k'}}{\partial x_i}\right\}$. Lemma 1 implies that each v_k can be written as a function of Λ.

$$v_k = v_k(\Lambda) \tag{46}$$

Step 6: Flow Φ.

Functional form equations for S_i. Recall that: $\frac{\partial \log S_j}{\partial x_j} = -\sum_{k=1} v_k(\Lambda) \frac{\partial \log S_j}{\partial \Lambda_k}$ where $v_k(\Lambda)$ is a function of aggregators Λ. We obtain:

$$-\frac{\partial S_i}{\partial x_i} = \sum_k v_k(\Lambda) \frac{\partial S_i}{\partial \Lambda_k} \tag{47}$$

for each good i.

Defining the flow Φ. A solution to these equations is based on the existence of a mapping for each t from Λ into a new vector of aggregators $\Phi(t, \Lambda)$ such that $\Phi(0, \Lambda) = \Lambda$ and such that:

$$\frac{\partial \Phi_k}{\partial t} = v_k(\Phi) \tag{48}$$

Action of Φ on S_i. This flow can be used to highlight symmetries across goods and invariances in the demand functions S_i and function M defined in step 1. First, consider $S_i(x_i + t, \Phi(t, \Lambda))$ as a function of t. Its derivative in t is given by:

$$\begin{aligned} & \frac{\partial S_i}{\partial x_i}(x_i + t, \Phi(t, \Lambda)) + \sum_{k=1} \frac{\partial \Phi_k}{\partial t}(\Phi(t, \Lambda)) \frac{\partial S_i}{\partial \Lambda_k}(x_i + t, \Phi(t, \Lambda)) \\ = \; & \frac{\partial S_i}{\partial x_i}(x_i + t, \Phi(t, \Lambda)) + \sum_{k=1} v_k(\Phi(t, \Lambda)) \frac{\partial S_i}{\partial \Lambda_k}(x_i + t, \Phi(t, \Lambda)) \; = \; 0 \end{aligned}$$

hence it does not depend on t, which implies:

$$S_i(x_i + t, \Phi(t, \Lambda)) = S_i(x_i, \Lambda) \tag{49}$$

for any t, x_i and Λ. Another way to highlight the role of Φ is to see that it captures the price effects and reduces demand to a function of aggregators after adjusting that price effect:

$$S_i(x_i, \Lambda) = S_i(0, \Phi(-x_i, \Lambda)) \tag{50}$$

Note also that $S_i(x_i, \Phi(t, \Lambda)) = S_i(x_i - t, \Lambda)$ strictly increases with t since S_i decreases in x_i.

Maximal flow. We still need to check that Φ is a global flow, i.e. defined for all $t \in \mathbb{R}$ (i.e. the vector field v is "complete").

By contradiction, suppose that for some Λ_0 the flow $\Phi(t, \Lambda_0)$ is defined only up to T. We would have:

$$\lim_{t \to T} S_i(0, \Phi(t, \Lambda_0)) = \lim_{t \to T} S_i(-t, \Lambda_0) = S_i(-T, \Lambda_0) > 0$$

for all i.

However, if the flow is defined only up to T, $\Phi([0, T), \Lambda_0)$ of $[0, T)$ cannot be contained into a compact set ("Escape Lemma", see Lemma 9.19 in Lee 2013's Intro to Smooth Manifolds). Given our assumption ("no escaping"), this implies that $\max_i |\log S_i(0, \Phi(t, \Lambda_0))|$ is unbounded and cannot have a finite limit, and contradicts the results above.

Other properties. Notice that $\Phi(t, \Phi(t', \Lambda)) = \Phi(t + t', \Lambda)$ for any t and t', hence, for any given t, Φ it is invertible in Λ, with inverse $\Phi(-t, \Lambda)$, given that $\Phi(t, \Phi(-t, \Lambda)) = \Lambda$. Also, for any t, Φ is differentiable (hence a diffeomorphism).

Step 7: Projecting aggregators. Here the goal is to show that each pair Λ can be written as:

$$\Lambda \; = \; F(\lambda, \Lambda')$$

for some isomorphism $F : \mathbb{R} \times \mathcal{M}_0 \to \mathbb{R}^K$ for some submanifold $\mathcal{M}_0$, such that we have a canonical projection on flow Φ, i.e. such that:

$$\Phi(t, F(\lambda, \Lambda')) \; = \; F(\lambda + t, \Lambda') \tag{51}$$

for any t, λ and Λ'.

Total price effects. Define the function D_0 of Λ by evaluating the sum of D_i at a reference point $x_i = 0$ for each good:

$$S_0(\Lambda) = \sum_i S_i(0, \Lambda)$$

Since for each i we obtain that $S_i(0, \Phi(t, \Lambda)) = S_i(-t, \Lambda)$ strictly increases with t, we also obtain that $S_0(\Phi(t, \Lambda))$ strictly increases with t.

We define by $\mathcal{M}_0$ the set of Λ such that $S_0(\Lambda) = 1$:

$$\mathcal{M}_0 \; = \; S_0^{-1}(\{1\}) \tag{52}$$

As $S_0(\Phi(t,\Lambda))$ strictly increases with t, we can deduce that the vector field v is never tangent to $\mathcal{M}_0$. This will be useful to apply the Flowout Theorem (see below).

Defining the mapping. We then simply construct F such that

$$F(\lambda, \Lambda') \; = \; \Phi(-\lambda, \Lambda')$$

As the flow Φ is global (and differentiable), F is defined for all λ and Λ'. We have yet to show surjectivity and injectivity globally.

Surjectivity. Take $\Lambda \in \mathbb{R}^K$. We need to find $\Lambda' \in \mathcal{M}_0$ and $\lambda \in \mathbb{R}$ such that $F(\lambda, \Lambda') = \Lambda$. But notice that we have then:

$$\Lambda' = \Phi(\lambda, \Lambda)$$

So, for such a λ and Λ' to exist, we need to show that

$$S_0(\Phi(\lambda, \Lambda)) \; = \; 1$$

for some λ. Note that:

$$S_0(\Phi(\lambda, \Lambda)) \; = \; \sum_i S_i(0, \Phi(\lambda, \Lambda)) \; = \; \sum_i S_i(-\lambda, \Lambda)$$

One of our assumption (on "Total price effects") is that for any Λ and any $y > 0$, there exist a real $t \in \mathbb{R}$ such that:

$$\sum_i S_i(t, \Lambda) = 1 \tag{53}$$

We can thus find λ such that:

$$\sum_i S_i(-\lambda, \Lambda) = 1$$

which is equivalent to having $\Phi(\lambda, \Lambda)) \in \mathcal{M}_0$. Setting $\Lambda' = \Phi(\lambda, \Lambda))$, we have $\Lambda = \Phi(-\lambda, \Lambda') = F(\lambda, \Lambda')$.

Injectivity. We also need to show that F is globally injective, but this is relatively easy using function D_0. Consider two sets of aggregators, (λ_1, Λ'_1) vs. (λ_0, Λ'_0), we have then:

$$\begin{aligned} F(\lambda_1, \Lambda'_1) \; = \; F(\lambda_0, \Lambda'_0) \quad &\Longleftrightarrow \quad \Phi(-\lambda_1, \Lambda'_1) = \Phi(-\lambda_0, \Lambda'_0) \\ &\Longleftrightarrow \quad \Phi(\lambda_0 - \lambda_1, \Lambda'_1) = \Lambda'_0 \end{aligned}$$

This implies that $D_0(\Phi(\lambda_0 - \lambda_1, \Lambda'_1)) = 1$. Since D_0 is strictly monotonic in λ, and since $D_0(\Phi(0, \Lambda'_1)) = 1$, we obtain that λ_1 must be equal to λ_0. In turn, we get: $\Lambda'_0 = \Phi(\lambda_0 - \lambda_1, \Lambda'_1) = \Phi(0, \Lambda'_1) = \Lambda'_1$.

Implication for D_i. These results imply that we can write:

$$S_i(x_i, F(\lambda, \Lambda')) \; = \; S_i\big(x_i, \Phi(-\lambda, \Lambda')\big) \; = \; S_i\big(0, \Phi(-x_i - \lambda, \Lambda')\big) \; = \; S_i\big(x_i + \lambda, \Lambda'\big) \tag{54}$$

Hence, up to a isomorphic mapping of the aggregators, we can rewrite S_i as a function of the price shifter λ and a vector of aggregators that belongs to a submanifold of lower dimension.

Step 8: Implications for M and V. As described earlier, evaluating aggregators at x, we must have:

$$\sum_j W_j(x_i, \Lambda(x)) \; = \; M(\Lambda(x)) \; + \; P(x) \tag{55}$$

$$P(x) \; = \; \sum_j W_j(x_i, \Lambda(x)) \; - \; M(\Lambda(x)) \tag{56}$$

and the first order condition (37) in each aggregator Λ_k implies that we must have:

$$\sum_j \frac{\partial W_j}{\partial \Lambda_k} \; = \; \frac{\partial M}{\partial \Lambda_k} \tag{57}$$

The same sets of FOC can be applied to (λ, Λ') if we use the canonical representation of aggregators.

Note that $W_i(x_i, \Lambda) = \int_0^{x_i} S_i(t, \Lambda)dt$. Hence, using the new functional form based on D_i and λ, we obtain:

$$W_i(x_i, \lambda, \Lambda') = \int_0^{x_i+\lambda} S_i(t, \Lambda')dt$$

The first order condition in λ implies that we must have:

$$\frac{\partial M}{\partial \lambda} \;=\; \sum_j \frac{\partial W_j}{\partial \lambda} \;=\; \sum_j S_j(x_i + \lambda, \Lambda') \;=\; 1$$

Hence:

$$M(\lambda, \Lambda') \;=\; G(\Lambda') + \lambda$$

for some function G that is independent of λ.

We obtain:

$$P(x) \;= -M(\lambda(x), \Lambda'(x)) + W(x) \;=\; -G(\Lambda') - \lambda + \sum_i \int_0^{x_i+\lambda} S_i(t, \Lambda')dt$$

where the aggregators $\lambda = \lambda(x)$ and $\Lambda' = \Lambda'(x)$ are such that the derivatives of the RHS in λ and Λ' are null.

Homogeneity for aggregators. Based on the FOC for aggregators, we have then:

$$\lambda(x + a) \;=\; \lambda(x) + a \quad \text{and} \quad \Lambda'(x + a) \;=\; \Lambda'(x)$$

Note again that x captures the vector of log prices. In the main text and the remainder of this appendix, we then switch back to using $\log p$ and $\log \lambda$.

Appendix D. Proof of Lemma 1

Suppose that the function $\mathcal{P}(p,\lambda,\Lambda)$ is defined as in equation (7):

$$\log \mathcal{P}(p,\lambda,\Lambda) \;=\; \log\lambda \;-\; G(\Lambda) \;+\; \sum_j \int_{t=1}^{p_j/\lambda} D_j(t,\Lambda)\, d\log t$$

Notice that we have:

$$\left.\frac{\partial \log \mathcal{P}}{\partial \log p_i}\right|_{\lambda,\Lambda} \;=\; D_j(p_j/\lambda,\Lambda)$$

It is positive and decreasing in p_i, hence $\log\mathcal{P}$ is strictly concave in $\log p$, conditional on Λ and λ.

As described, we consider two cases:

i) If $\log\mathcal{P}(p,\lambda,\Lambda)$ is convex in Λ, define $\log\widetilde{\mathcal{P}}(p) = \{\min_\Lambda \log\mathcal{P}(p,1,\Lambda)\}$.

ii) If $\log\mathcal{P}(p,\lambda,\Lambda)$ is concave in $(\Lambda,\log p)$, define $\log\widetilde{\mathcal{P}}(p) = \max_\Lambda \log\mathcal{P}(p,1,\Lambda)$.

In each of these two cases, the max or min operations preserve the concavity property, so we obtain that $\log\widetilde{\mathcal{P}}(p)$ is strictly concave in $\log p$. We then define:

$$\log P(p) \;=\; \max_\lambda \{\log\lambda + \log\widetilde{\mathcal{P}}(p/\lambda)\}$$

which coincides with the function $P(p)$ defined in the statement of Lemma 1.

Using the fact that $\log\widetilde{\mathcal{P}}$ is strictly concave in $\log p$, we can show that $\log P$ is concave in p. We do this by examining the Hessian, and showing that it is semi-definite negative. Since the right-hand side is strictly concave in $\log\lambda$, aggregator λ is uniquely defined. The first-order condition is:

$$1 \;=\; \sum_i \frac{\partial \log\widetilde{\mathcal{P}}(p/\lambda)}{\partial \log p_i}$$

Differentiating, we get:

$$0 \;=\; -\left(\sum_{i,l} \frac{\partial^2 \log\widetilde{\mathcal{P}}}{\partial\log p_i \partial\log p_l}\right)\frac{\partial\lambda}{\partial\log p_j} \;+\; \sum_i \frac{\partial^2\log\widetilde{\mathcal{P}}}{\partial\log p_i\partial\log p_j}$$

In matrix form, denoting by $\widetilde{\mathcal{H}}$ the Hessian of $\log\widetilde{\mathcal{P}}$, this yields:

$$0 \;=\; -(\mathbb{1}^t\widetilde{\mathcal{H}}\mathbb{1})\nabla\lambda^t \;+\; \mathbb{1}^t\widetilde{\mathcal{H}}$$

and thus: $\nabla\lambda^t = \eta\mathbb{1}^t\widetilde{\mathcal{H}}$ where $\eta<0$ denotes $1/(\mathbb{1}^t\widetilde{\mathcal{H}}\mathbb{1})$.

Turning to P and using again the envelope theorem, we obtain: $\frac{\partial P}{\partial p_i} = \frac{P(p/\lambda)}{p_i}\frac{\partial\log\widetilde{\mathcal{P}}(p/\lambda)}{\partial\log p_i}$ and thus the Hessian:

$$\begin{aligned}\frac{\partial^2 P}{\partial p_i\partial p_j} \;&=\; \frac{P}{p_ip_j}\left[\frac{\partial\log\widetilde{\mathcal{P}}}{\partial\log p_i}\frac{\partial\log\widetilde{\mathcal{P}}}{\partial\log p_j} \;-\; \mathbb{1}(i=j)\frac{\partial\log\widetilde{\mathcal{P}}}{\partial\log p_i}\right]\\ &+\; \frac{P}{p_ip_j}\left[\frac{\partial^2\log\widetilde{\mathcal{P}}}{\partial\log p_i\partial\log p_j} \;-\; \sum_l \frac{\partial^2\log\widetilde{\mathcal{P}}}{\partial\log p_i\partial\log p_l}\frac{\partial\lambda}{\partial\log p_j}\right]\end{aligned}$$

In this expression, the terms in first brackets are the coefficients of a semi-definite negative matrix because the diagonal coefficients are negative and weakly "dominate" the non-diagonal coefficients, since its row sum (or column sum) is equal to zero.

Next we show that the terms in the second brackets are also the coefficients of a negative semi-definite matrix. In matrix form, the terms in the second matrix coincide with the matrix M defined as:

$$M \;=\; \widetilde{\mathcal{H}} - \widetilde{\mathcal{H}}\mathbb{1}\nabla\lambda^t \;=\; \widetilde{\mathcal{H}} - \eta\widetilde{\mathcal{H}}\mathbb{1}\mathbb{1}^t\widetilde{\mathcal{H}}$$

To prove negative semi-definiteness, we show that for any vector v, we have $v^tMv \le 0$. Indeed:

$$v^tMv \le 0 \qquad\Leftrightarrow\qquad (v^t\widetilde{\mathcal{H}}v) \le \eta(v^t\widetilde{\mathcal{H}}\mathbb{1})(\mathbb{1}^t\widetilde{\mathcal{H}}v) \qquad\Leftrightarrow\qquad (\mathbb{1}^t\widetilde{\mathcal{H}}\mathbb{1})(v^t\widetilde{\mathcal{H}}v) \ge (v^t\widetilde{\mathcal{H}}\mathbb{1})^2$$

The latter is Cauchy-Schwarz inequality, which holds for any semi-definite matrix $\widetilde{\mathcal{H}}$

Appendix E. Proof of Proposition 3

Lemma 1 proves that the demand system is rational and well-defined (under assumptions of Lemma 1). This implies that there exists a well-defined concave utility function $U(q)$ that generates such demand system. By definition, under homothetic preferences, the price index can be obtained from utility as

$$\log P(p) \;=\; -\max_{q_i}\Big\{\log U(q) \quad s.t. \quad \sum_i p_i q_i = 1\Big\}$$

But we can then show that:

$$\max_{q_i}\Big\{\log U(q) \quad s.t. \quad \sum_i p_i q_i = 1\Big\} \;=\; \max_{q}\Big\{1 \,-\, \sum_i p_i q_i \,+\, \log U(q)\Big\}$$

To see this equality, notice that in the right, the optimum in q satisfies: $p_i = \frac{\partial \log U(q)}{\partial q_i}$. On the left, we have $\mu p_i = \frac{\partial \log U(q)}{\partial q_i}$ where μ is the Lagrange multiplier. But since $U(q)$ is homogeneous of degree one, we must have:

$$1 \;=\; \sum_i p_i q_i \;=\; \sum_i q_i \frac{\partial \log U}{\partial q_i}/\mu \;=\; 1/\mu$$

So $\mu = 1$ and the optimal q are the same in both maximization problems. This implies that, on the right, $\sum_i p_i q_i = 1$ at the optimum.

Combining, ignoring a constant term (-1), this implies that the log price index is the conjugate of log utility:

$$\log P(p) \;=\; \min_q \Big\{\sum_i p_i q_i \,-\, \log U(q)\Big\}$$

Hence, applying the Legendre-Fenchel duality theorem, we can obtain log utility as the concave conjugate of the log price index. Denote $W_i(p_i, \Lambda) = \int_{t=0}^{\log p_i} D_i(t, \Lambda)\,dt$ and $u_i(q_i, \Lambda) = \min_p\{p_i q_i - W_i(p_i, \Lambda)\}$ its conjugate in p_i (i.e. conditional on Λ). In the case where $\log \mathcal{P}(p, \lambda, \Lambda)$ is convex in Λ (condition i) of Lemma 1), we get:

$$\begin{aligned}
\log U(q) \;&=\; \min_p \Big\{\sum_i p_i q_i \,-\, \log P(p)\Big\} \\
&=\; \min_p \Big\{\sum_i p_i q_i \,-\, \max_\lambda \max_\Lambda \Big\{\sum_i W_i(p_i/\lambda, \Lambda) \,+\, \log\lambda \,-\, G(\Lambda)\Big\}\Big\} \\
&=\; \min_p \min_\lambda \min_\Lambda \Big\{\sum_i p_i q_i \,-\, \sum_i W_i(p_i/\lambda, \Lambda) \,-\, \log\lambda \,+\, G(\Lambda)\Big\} \\
&=\; \min_\lambda \min_\Lambda \min_p \Big\{\sum_i p_i q_i \,-\, \sum_i W_i(p_i/\lambda, \Lambda) \,-\, \log\lambda \,+\, G(\Lambda)\Big\} \\
&=\; \min_\lambda \min_\Lambda \Big\{\sum_i \min_p \Big\{p_i q_i - W_i(p_i/\lambda, \Lambda)\Big\} \,-\, \log\lambda \,+\, G(\Lambda)\Big\} \\
&=\; \min_\lambda \min_\Lambda \Big\{\sum_i u_i(q_i\lambda, \Lambda) \,-\, \log\lambda \,+\, G(\Lambda)\Big\}
\end{aligned}$$

In the concave case (condition ii) of Lemma 1), we have:

$$\begin{aligned}
\log U(q) \;&=\; \min_p \Big\{\sum_i p_i q_i \,-\, \max_\lambda \min_\Lambda \Big\{\sum_i W_i(p_i/\lambda, \Lambda) \,+\, \log\lambda \,-\, G(\Lambda)\Big\}\Big\} \\
&=\; \min_p \min_\lambda \max_\Lambda \Big\{\sum_i p_i q_i \,-\, \sum_i W_i(p_i/\lambda, \Lambda) \,-\, \log\lambda \,+\, G(\Lambda)\Big\} \\
&=\; \min_\lambda \max_\Lambda \Big\{\sum_i \min_p \Big\{p_i q_i - W_i(p_i/\lambda, \Lambda)\Big\} \,-\, \log\lambda \,+\, G(\Lambda)\Big\} \\
&=\; \min_\lambda \max_\Lambda \Big\{\sum_i u_i(q_i\lambda, \Lambda) \,-\, \log\lambda \,+\, G(\Lambda)\Big\}
\end{aligned}$$

General case, under assumptions of Lemma 1. Again, denote $x_i = \log p_i$.

In case i) and ii) of Lemma 1, the following function:

$$F(x) \;=\; \log \widetilde{\mathcal{P}}(e^x) \;=\; \max_\Lambda\{\log \mathcal{P}(e^x, 1, \Lambda)\} \;=\; \max_\Lambda \Big\{ -G(\Lambda) \;+\; \sum_j \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt \Big\}$$

is concave in x. Define F^* the conjugate, which is again concave in y, to be used as perturbation in the PUM representation:

$$F^*(y) \;=\; \min_x \Big\{ \sum_j \pi_j x_j - F(x) \Big\}$$

Also define $W_j(x_j, \Lambda) = \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt$, which is concave in x_j (since D_j is decreasing in its first argument), and define its conjugate:

$$W_j^*(\pi_j, \Lambda) = \min_{x_j} \Big\{ \pi_j x_j - W_j(x_j, \Lambda) \Big\} = \min_{x_j} \Big\{ \pi_j x_j - \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt \Big\}$$

which is again concave in π_j.

In case ii) of Lemma 1, assuming $\log \mathcal{P}(p, \lambda, \Lambda)$ is concave in $(\Lambda, \log p)$, we obtain that:

$$F(x) \;=\; \max_\Lambda \Big\{ -G(\Lambda) \;+\; \sum_j \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt \Big\}$$

Back its conjugate F^*, we use the joint concavity in (Λ, x) to obtain:

$$\begin{aligned}
F^*(\pi) \;&=\; \min_x \Big\{ \sum_j \pi_j x_j - \max_\Lambda \Big\{ -G(\Lambda) \;+\; \sum_j \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt \Big\} \Big\} \\
&=\; \min_x \min_\Lambda \Big\{ \sum_j \pi_j x_j \;+\; G(\Lambda) \;-\; \sum_j \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt \Big\} \Big\} \\
&=\; \min_\Lambda \min_x \Big\{ \sum_j \pi_j x_j \;+\; G(\Lambda) \;-\; \sum_j \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt \Big\} \\
&=\; \min_\Lambda \Big\{ G(\Lambda) \;+ \sum_j \min_{x_j} \Big\{ \pi_j x_j \;-\; \int_{t=0}^{x_j} D_j(e^t, \Lambda)\, dt \Big\} \Big\} \\
&=\; \min_\Lambda \Big\{ G(\Lambda) \;+ \sum_j W^*(\pi_j, \Lambda) \Big\}
\end{aligned}$$

In case i) of Lemma 1, assuming $\log \mathcal{P}(p, \lambda, \Lambda)$ is convex in Λ, we obtain instead that $G(\Lambda) + \sum_j W^*(\pi_j, \Lambda)$ is concave in Λ (taking the minimum in x preserves concavity), such that:

$$F^*(\pi) \;=\; \max_\Lambda \Big\{ G(\Lambda) \;+ \sum_j W^*(\pi_j, \Lambda) \Big\}$$

Now, define the following PUM indirect utility:

$$V(p) \;=\; \max_{\pi \in \Delta} \Big\{ -\sum_j \pi_j \log p_j \;-\; F^*(\pi) \Big\}$$

We want to show that $V(p) = -\log P(p)$, i.e. that it corresponds to the indirect utility function (at least up to a constant term), and that the expenditure shares $S_i = \frac{\partial \log P}{\partial \log p_j}$ correspond to π_j.

First, using the equalities above on F^*, we can check that Λ satisfies the same condition just above as a function of π (last line) as it does as a function of x (first line) as long as x is minimizing the expression for F^* (first line), i.e. as long as:

$$\pi_j \;=\; D_j(e^{x_j}, \Lambda)$$

Next, we can see that the partial derivatives of $\log V$ and $\log P$ coincide, conditional on Λ:

$$\frac{\partial \log V}{\partial \log p_j} \;=\; \pi_j$$

where π_j satisfies:

$$\lambda - \log p_j \;=\; \frac{\partial F^*(\pi)}{\partial \pi_j}$$

where λ is the Lagrange multiplier associated with the constraint that π is in the simplex: $\pi \in \Delta$, i.e. $\sum_j \pi_j = 1$. In turn, from the definition of F^* (first line above), we obtain:

$$x_j \;=\; \frac{\partial F^*(\pi)}{\partial \pi_j} \;=\; \lambda - \log p_j$$

with x_j also such that it satisfies the FOC:

$$\pi_j \;=\; D_j(e^{x_j}, \Lambda)$$

Given that we must have $\sum_j \pi_j \;=\; 1$, the Lagrange multiplier λ must coincide with the special aggregator $\log \lambda$, and we end up with $S_j = \pi_j = D_j(p_j/\lambda, \Lambda)$.

Baseline specification. Suppose that indirect utility is given by:

$$\log V(p) \;=\; \max_{S \in \Delta} \Big\{ -\sum_i S_i \log p_i \;-\; \sum_j W_j^*\big(S_j\big) - \frac{1}{2}\sum_i \sum_j B_{ij} S_j S_i \Big\}$$

From Roy's identity and the envelope theorem, we obtain that S_i coincides with expenditure shares. The FOC in S_i is:

$$-\log p_i \;-\; W_i^{*\prime}\big(S_i\big) - \sum_j B_{ij} S_j \;=\; \lambda$$

where λ is the Lagrange multiplier associated with the constraint that W is on the simplex. With $-D_i$ denoting the inverse of $W_i^{*\prime}$, we obtain:

$$S_i \;=\; D_i\Big(\log p_i \;+\; \lambda \;+\; \sum_j B_{ij} S_j\Big)$$

PUM in prices and quantities. The above PUM representation uses log prices as "observables" and expenditure shares as the dual variables. Using results from Proposition 3, we can also obtain a simple PUM formulation as a function of price levels (instead of log prices) and quantities as the dual variables.

In fact, we have already shown earlier (see above in the proof of Proposition 3), that any homothetic preferences with strictly concave price index function $\log P(p)$ can be expressed as:

$$\log P(p) \;=\; \min_q \Big\{ \sum_i p_i q_i \;-\; \log U(q) \Big\}$$

(up to a constant term equal to one). In this case, the perturbation function coincides with log utility described in Proposition 3:

$$\log U(q) \;=\; \sum_i u_i(q_i \lambda, \Lambda) \;-\; \log \lambda \;+\; G(\Lambda)$$

Appendix G. Additional details on our baseline specification

Expenditure shares are specified as:

$$S_i \;=\; D_i\Big(\log p_i + \lambda + \sum_j \beta_{ij} S_j\Big)$$

where β is a positive semi-definite matrix with $\beta_{ij} = \sum_k b_{ik} b_{jk}$.

Here we describe how to retrieve cross-price elasticities (or semi-elasticities), and how to choose parameters to fit any mixed Logit/CES structure with K aggregators (at a given price vector).

Semi-elasticities, conditional on λ. We obtain:

$$\frac{dS_i}{d\log p_j} \;=\; \mathbb{1}(i=j)D_i' \;+\; D_i' \frac{d\lambda}{d\log p_j} \;+\; D_i' \sum_{i'} \beta_{ii'} \frac{dS_{i'}}{d\log p_j}$$

In matrix form, this yields:

$$J \;=\; \sigma[I \;+\; \mathbb{1}\,\nabla\lambda^t \;+\; \beta J]$$
$$[I - \sigma\beta]J \;=\; \sigma[I \;+\; \mathbb{1}\,\nabla\lambda^t]$$
$$J \;=\; [I - \sigma\beta]^{-1}\sigma[I \;+\; \mathbb{1}\,\nabla\lambda^t]$$

Thus, holding λ constant, the matrix of cross-price effects is given by:

$$J|_\lambda \;=\; [I - \sigma\beta]^{-1}\sigma$$

When $\rho(\sigma\beta) < 1$ (spectral radius smaller than unity), we can develop and the sum converges:

$$J|_\lambda \;=\; \sigma + \sigma\beta\sigma + (\sigma\beta)^2\sigma + (\sigma\beta)^3\sigma + ...$$

$J|_\lambda$ is symmetric and also equal to $J|_\lambda = \sigma[I - \sigma\beta]^{-1}$.

Endogenizing λ. Aggregator λ is such that $\sum_i S_i = 1$, hence: $\mathbb{1}^t J = 0$, which leads to:

$$0 \;=\; \mathbb{1}^t J \;=\; \mathbb{1}^t[I - \sigma\beta]^{-1}\sigma[I \;+\; \mathbb{1}\,\nabla\lambda^t] \;=\; \mathbb{1}^t J|_\lambda[I \;+\; \mathbb{1}\,\nabla\lambda^t]$$

So:

$$\nabla\lambda^t \;=\; -(\mathbb{1}^t J|_\lambda \mathbb{1})^{-1}\mathbb{1}^t J|_\lambda$$

So the full matrix is then:

$$J \;=\; J|_\lambda[I \;+\; \mathbb{1}\,\nabla\lambda^t] \;=\; J|_\lambda \;-\; (\mathbb{1}^t J|_\lambda \mathbb{1})^{-1} J|_\lambda \mathbb{1}\mathbb{1}^t J|_\lambda$$

which is again symmetric.

Matching cross-price effects from mixed Logit/CES. Take a mixed logit model: can we find equivalent J with baseline demand with observationally the same aggregate expenditure shares and cross-price effects, at a given vector of prices?

Consider logit cross price effects with K types of consumers that we aggregate, with income share I_h for each consumer type h, so that:

$$S_i \;=\; \sum_h I_h w_{ih}$$
$$w_{ih} \;=\; \frac{A_{ih} e^{-\theta_h \log p_i}}{\sum_j A_{ih} e^{-\theta_h \log p_j}}$$
$$\frac{\partial S_i}{\partial \log p_j} \;=\; -\sum_h I_h \theta_h \Big[w_{ih}\mathbb{1}(i=j) - w_{ih}w_{jh}\Big]$$

Take a vector of prices $\{p_i\}$. Take our baseline specification with iso-elastic demand (or another parameterization).

$$\widetilde{S}_i \;=\; \exp\Big[-\theta_i\Big(\bar{\alpha}_i + \log p_i + \lambda + \sum_j \beta_{ij} S_j\Big)\Big] \tag{58}$$

where β is a positive semi-definite matrix with $\beta_{ij} = \sum_k b_{ik} b_{jk}$.

Proposition 5. *For any Mixed Logit/CES demand at given prices p, we can find parameters θ_i, $\bar{\alpha}_i$, b_{ik} in our baseline specification such that it coincides with aggregate expenditure shares and the matrix of cross-price effects of such mixed Logit/CES evaluated at that price p, i.e.*

$$\widetilde{S}_i(p) \;=\; S_i(p) \qquad ; \qquad \frac{\partial \widetilde{S}_i(p)}{\partial \log p_j} \;=\; \frac{\partial S_i(p)}{\partial \log p_j}$$

Note that, with mixed logit, mixing across K consumers, the price effect matrix is: i) symmetric, ii) has negative diagonal coefficients, iii) has non-negative off-diagonal coefficients, iv) column sums and row sums are null, v) has rank K substitution effects, i.e. can be decomposed into a diagonal matrix plus a rank-K matrix. This imposes strong restrictions on the substitution patterns that the demand system must have. While we can choose our baseline demand to fit any such K consumer mixed logit, the reverse is not true, since our baseline demand systems allows for more flexible substitution patterns.

Proof of proposition. Recall that with mixed logit, the Jacobian is:

$$\frac{\partial S_i}{\partial \log p_j} \;=\; -\sum_h I_h \theta_h \Big[w_{ih} \mathbb{1}(i=j) - w_{ih} w_{jh} \Big]$$

With our baseline demand, the Jacobian is:

$$J \;=\; J|_\lambda [I \,+\, \mathbb{1}\, \nabla \lambda^t] \;=\; J|_\lambda \,-\, (\mathbb{1}^t J|_\lambda \mathbb{1})^{-1} J|_\lambda \mathbb{1}\mathbb{1}^t J|_\lambda$$

where $J|_\lambda$ is the matrix of cross-price effects holding λ constant, and given by:

$$J|_\lambda \;=\; [I - \sigma\beta]^{-1}\sigma$$

with $\beta = bb^t$ positive semi-definite. Note that $J|_\lambda$ is symmetric and also equal to $J|_\lambda = \sigma[I-\sigma\beta]^{-1}$.

The goal is to construct demand such that the two Jacobian coincides at a given price (as well as the level of aggregate expenditure shares at that price). We first define the own and cross price effect matrices, holding the "special aggregator" constant, then we check that the two coincides once we bring in that special aggregator, then we recover the underlying principal components, b.

Step 1: own and substitution matrices without special aggregator.

We first focus on $J|_\lambda$ is the matrix of cross-price effects holding λ constant.

First, construct σ as a diagonal matrix such that:

$$\sigma \mathbb{1} \;=\; -\sum_h I_h \theta_h w_h$$

where w_h denotes the vector of expenditure shares for individual h, and construct the matrix of cross-price effects Σ (without special aggregator λ) such that:

$$\Sigma \;=\; \sum_{h \neq 1} I_h \theta_h w_h w_h^t$$

where w_h^t is the transpose of the vector of expenditure shares for consumer type h, and where the sum is taken by excluding one type (here $h = 1$ but it can be any type).

Step 2: adding the special aggregator.

Then, we show that the added component related to the special aggregator exactly fills in the role of that missing type ($h = 1$).

The special aggregator adjusts such that the budget constraint holds, i.e. such that total expenditure shares always add up to unity. This implies that the column sum of the Jacobian is null: $\mathbb{1}^t J$. Given $J|_\lambda$, as shown earlier, it implies that

$$J \;=\; J|_\lambda [I \,+\, \mathbb{1}\, \nabla \lambda^t] \;=\; J|_\lambda \,-\, (\mathbb{1}^t J|_\lambda \mathbb{1})^{-1} J|_\lambda \mathbb{1}\mathbb{1}^t J|_\lambda$$

With $J|_\lambda = \sigma + \Sigma$ defined above, we obtain:

$$J|_\lambda \mathbb{1} \;=\; \sigma\mathbb{1} + \Sigma\mathbb{1} \;=\; -\sum_h I_h \theta_h w_h \;+\; \sum_{h\neq 1} I_h \theta_h w_h w_h^t \mathbb{1} \;=\; -\sum_h I_h \theta_h w_h \;+\; \sum_{h \neq 1} I_h \theta_h w_h \;=\; -I_1 \theta_1 w_1$$

and thus:

$$\mathbb{1}^t J|_\lambda \mathbb{1} \;=\; -I_1\theta_1$$
$$J|_\lambda \mathbb{1}\mathbb{1}^t J|_\lambda \;=\; (I_1\theta_1 w_1)(I_1\theta_1 w_1^t)$$

so the full Jacobian (adding the special aggregator) is:

$$J_{ij} \;=\; -\sum_h I_h\theta_h w_{ih}\mathbb{1}(i=j) \;+\; \sum_h I_h\theta_h w_{ih}w_{jh}$$

and coincides with the mixed logit Jacobian.

Step 3: spectral radius. We have yet to construct the b matrix. To do so, the next step is to show that such a substitution matrix Σ as defined above as a spectral radius smaller than one when we divide by elasticities θ.

Define $\varepsilon_i = \sqrt{\sum_h I_h\theta_h w_{ih}}$ and denote by ε the diagonal matrix with coefficients ε_i on its diagonal, so that $\sigma = -\varepsilon^2$. Note again that $\Sigma = \sum_{h\neq 1} I_h\theta_h w_h w_h^t$. To show that the spectral radius of $\varepsilon^{-1}\Sigma\varepsilon^{-1}$ is smaller than one, we need to have:

$$x^t\varepsilon^{-1}\Sigma\varepsilon^{-1}x < x^t x$$

for any non-zero vector x. Up to a rescaling, this is equivalent to having:

$$x^t\Sigma x < x^t\varepsilon^2 x$$

for any $x \neq 0$. Splitting up across coefficients, this is equivalent to:

$$\sum_{h\neq 1} I_h \big(\sum_i x_i w_{ih}\big)^2 \;<\; \sum_i x_i^2 \,\big(\sum_h I_h w_{ih}\big)$$
$$\Leftrightarrow \sum_{h\neq 1} I_h \Big[\big(\sum_i x_i w_{ih}\big)^2 \;-\; \big(\sum_i x_i^2 w_{ih}\big)\Big] \;<\; \big(\sum_i x_i^2 w_{i1}\big) I_1$$

On the left, each term in brackets is non-positive, and strictly negative unless x is uniform across all goods i. The term on the right is positive, with a strict inequality if x is uniform and non-zero. Hence this inequality holds for any $x \neq 0$.

Step 4: Inversion. Finally, we want to find b and $\beta = bb^t$ such that:

$$J|_\lambda \;=\; [I - \sigma\beta]^{-1}\sigma \;=\; \sigma + \Sigma$$

knowing that $\sigma = -\varepsilon^2$ and that $\varepsilon^{-1}\Sigma\varepsilon^{-1}$ has a spectral radius smaller than unity.

With this property of the spectral radius, note that $I - \varepsilon^{-1}\Sigma\varepsilon^{-1}$ is invertible, we can then define:

$$\beta \;=\; \varepsilon^{-1}[I - \varepsilon^{-1}\Sigma\varepsilon^{-1}]^{-1}\varepsilon^{-1} - \varepsilon^{-2}$$

With a spectral radius smaller than unity and $\sigma = -\varepsilon^2$, we can see this inversion as the sum:

$$\varepsilon\beta\varepsilon \;=\; \varepsilon^{-1}\Sigma\varepsilon^{-1} \;+\; \varepsilon^{-1}\Sigma\varepsilon^{-2}\Sigma\varepsilon^{-1} + \varepsilon^{-1}\Sigma\varepsilon^{-2}\Sigma\varepsilon^{-2}\Sigma\varepsilon^{-1}...$$

This is a symmetric positive semi-definite matrix, so we can write

$$\beta \;=\; bb^t$$

where b has as many columns as the rank as β. Furthermore, this expression is equivalent to:

$$\varepsilon\beta\varepsilon \;=\; \varepsilon^{-1}\Sigma\Big[I + \varepsilon^{-2}\Sigma + \varepsilon^{-2}\Sigma\varepsilon^{-2}\Sigma + ...\Big]\varepsilon^{-1}$$
$$\Leftrightarrow \;\; \varepsilon\beta\varepsilon \;=\; \varepsilon^{-1}\Sigma\Big[I - \varepsilon^{-2}\Sigma\Big]^{-1}\varepsilon^{-1}$$
$$\Leftrightarrow \;\; \beta \;=\; \sigma^{-1}\Sigma\Big[I + \sigma^{-1}\Sigma\Big]^{-1}\sigma^{-1}$$

In this formula we can see that Σ has the same rank as β and b, hence b can be defined with just $K-1$ columns.

Appendix H. Variations on Cereal Demand

This appendix collects estimation results across the full range of instrument sets and demographic settings for both the K-aggregator and BLP models. The main text reports the preferred specification (FL with demographic taste shifters); the tables below document robustness and identify the patterns that motivate the preferred choice.

H.1. K-Aggregator Variations.

Baseline two-step results. Table 9 reports two-step GMM estimates across all instrument sets at $K = 1$ and $K = 2$, with and without demographic interaction instruments (but without estimating Π). With demographic instruments, the price-elasticity parameter θ is remarkably stable at 1.1–1.4. Without demographics, θ exhibits much greater sensitivity, ranging from 0.18 (Nevo) to 6.46 (GH-poly).

Table 9. Two-step GMM results across instrument sets and demographic settings (without Π).

Spec	Demog	K	J	dof	$\hat{\theta}$	λ_1	λ_2
FL	demog	1	236.01	44	1.13	3.99	≈ 0
	demog	2	236.92	39	1.11	3.94	0.062
	nodemog	1	167.92	30	0.67	4.24	≈ 0
	nodemog	2	136.77	25	0.68	3.65	0.184
FL-level	demog	1	221.46	44	1.18	4.70	≈ 0
	demog	2	217.62	39	1.18	4.59	0.005
	nodemog	1	133.65	30	2.06	10.26	≈ 0
	nodemog	2	117.97	25	1.75	8.77	3.37
FL-poly	demog	1	296.96	44	1.10	0.009	≈ 0
	demog	2	283.23	39	1.12	1.90	0.002
	nodemog	1	239.68	30	1.04	0.009	≈ 0
	nodemog	2	208.59	25	1.07	4.46	0.002
GH	demog	1	59.58	14	1.37	39.5	≈ 0
	demog	2	21.61	9	2.29	92.4	3.13
	nodemog	1	50.35	0	0.96	0.78	≈ 0
	nodemog	2	0.00	0	3.00	67.8	0.357
Nevo	demog	1	262.80	28	1.15	7.82	≈ 0
	demog	2	260.64	23	1.15	7.95	0.003
	nodemog	1	192.19	14	0.18	6.31	≈ 0
	nodemog	2	182.15	9	0.18	6.25	0.013
Combined	demog	1	784.91	92	1.12	0.020	≈ 0
	demog	2	458.15	87	1.22	1.29	0.021
	nodemog	1	691.22	78	1.08	0.022	≈ 0
	nodemog	2	349.24	73	1.21	0.779	0.035

Results with demographic taste shifters (Π). Table 10 reports two-step GMM results with demographic taste shifters estimated. Adding Π lowers J across all overidentified specifications and stabilizes θ.

Table 10. Two-step GMM with demographic taste shifters (Π).

Spec	K	J	dof	$\hat{\theta}$	λ_1	λ_2
FL	1	220.47	34	1.16	1.13	≈ 0
	2	213.47	29	1.18	5.84	0.706
	3	212.98	24	1.18	5.82	0.817
FL-level	1	180.23	34	1.24	1.76	≈ 0
	2	148.75	29	1.27	10.36	2.14
	3	147.12	24	1.27	10.23	2.04
FL-poly	1	238.21	34	1.08	0.004	≈ 0
	2	203.48	29	1.11	2.56	0.002
GH	1	3.62	4	1.53	33.3	≈ 0
	2	0.24	0	2.17	51.5	30.2
Nevo	1	184.35	18	1.20	0.116	≈ 0
	2	162.09	13	1.16	6.22	0.053
Combined	1	503.86	82	1.19	0.019	≈ 0
	2	449.17	77	1.23	0.99	0.037

Estimated Π *matrices at* $K = 2$. Table 11 compares the estimated demographic taste shifters across instrument sets. The patterns are qualitatively robust: fiber coefficients are consistently positive, fat coefficients are negative or near zero, and sodium is negligible throughout.

Table 11. $\hat{\Pi}$ at $K = 2$ for selected instrument sets.

Spec	Demographic	Sugar	Fiber	Protein	Fat	Sodium
FL	Income	0.002	0.094	0.030	0.020	−0.001
	Age	0.003	0.126	0.042	0.004	−0.001
FL-level	Income	0.005	0.099	0.014	−0.081	−0.002
	Age	0.007	0.124	0.006	−0.131	−0.002
GH	Income	−0.113	0.411	−0.458	1.030	−0.002
	Age	−0.140	0.587	−0.647	1.319	−0.002
Nevo	Income	0.002	−0.003	−0.429	0.030	−0.000
	Age	0.006	−0.015	−0.569	0.015	0.000

Notes: Each entry gives the effect of a one-unit increase in market-mean demographics on the valuation of the corresponding characteristic. The FL-level and FL estimates are qualitatively similar; the GH and Nevo estimates show larger magnitudes and some sign differences, reflecting limited overidentification (dof = 0 for GH at $K = 2$).

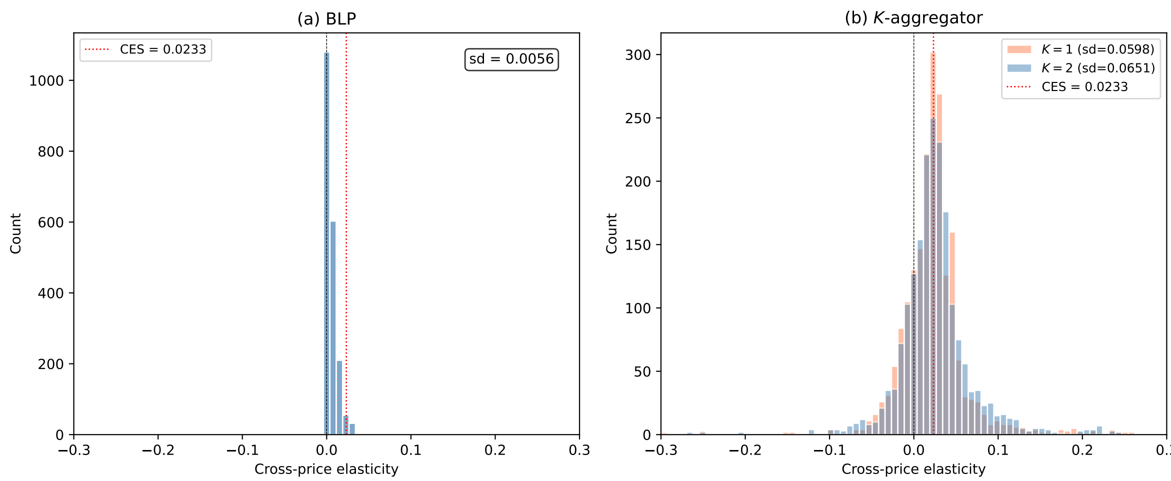
Bootstrap inference. Table 12 reports percentile confidence intervals from the moment wild bootstrap for the two Fally–Ligon instrument sets.

In both specifications, θ is precisely estimated near unity and the first eigenvalue is clearly nonzero. The second eigenvalue has 95% confidence intervals that barely exclude zero, indicating that the second factor is present but borderline significant.

Table 12. Bootstrap percentile confidence intervals (demog Π, $K = 2$).

	FL-level (512 reps)		FL (796 reps)	
Parameter	Estimate	95% CI	Estimate	95% CI
θ	1.27	(1.15, 1.39)	1.18	(1.05, 1.30)
λ_1	10.36	(2.77, 20.29)	5.84	(1.08, 14.26)
λ_2	2.14	(0.002, 3.80)	0.71	(0.003, 2.56)

Figure 2. Distribution of cross-price elasticities for BLP and K-Agg (FL instruments and demographic taste shifters)



The bootstrap Wald test of $H_0 \colon \Gamma_{ll'} = 0$ for all $l \neq l'$ (diagonality) yields $W = 9.15$ ($p = 0.52$) for FL-level and $W = 4.69$ ($p = 0.91$) for FL. We cannot reject diagonality with either instrument set.

CUE pathology. Continuously-updating estimation was run for all specifications but produces degenerate results. In every case, the CUE criterion converges to approximately $J \approx 21$, with Γ eigenvalues collapsing to near zero. This behavior is consistent across instrument sets and K values, suggesting a pathological attractor in the CUE objective surface rather than a genuine economic finding. The two-step estimator does not exhibit this behavior.

H.2. **BLP Variations.** Table 13 reports BLP estimates across instrument sets under three demographic settings: (i) no demographics (200 Halton draws), (ii) empirical demographic draws without π, and (iii) full demographics with π taste shifters (initialized at $0.01 \times \text{randn}(7, 2)$, seed 42).

Key patterns across specifications:

(1) σ_1 *revival.* With π taste shifters, every specification shows $\sigma_1 \approx 0.10$, a remarkably consistent estimate of the constant's random coefficient. Without demographics, 4 of 7 specifications collapse σ_1 to zero.

Table 13. BLP three-way demographic comparison.

Spec	Mode	Objective	β_{price}	σ_1	σ_{sugar}
FL-level	No-demog	753.1	−3.548	0.000	0.000
	Emp-draws	741.6	−3.564	0.170	0.010
	Demog	756.3	−3.435	0.097	0.009
FL	No-demog	913.0	−5.307	0.000	0.000
	Demog	930.7	−3.412	0.101	0.007
Nevo	No-demog	341.0	−0.980	0.101	0.010
	Demog	358.5	−3.321	0.099	0.012
GH	No-demog	161.9	−1.215	0.000	0.003
	Demog	232.8	−3.686	0.100	0.002
GH-poly	No-demog	140.8	−7.851	0.000	0.002
FL-poly	No-demog	719.5	−4.337	0.159	0.000
	Demog	221.7	−4.098	0.102	0.000
Combined	No-demog	2071.2	−3.466	0.156	0.001
	Demog	1695.5	−3.495	0.096	0.000

Notes: $\sigma_{\text{price}} = 0$ in all 21 estimation runs and is omitted. No-demog uses 200 Halton draws; Emp-draws uses the empirical demographic distribution without π; Demog uses empirical draws plus π taste shifters.

(2) β_{price} *stabilization.* The demog estimates cluster around −3.3 to −3.7, much more tightly than the no-demog range (−0.98 to −7.85).
(3) $\sigma_{\text{price}} = 0$ *universally.* The price random coefficient is zero in all 21 estimation runs (3 modes × 7 specs). The data do not support heterogeneity in price sensitivity beyond what is captured by demographics.
(4) π *initialization matters.* Small random initialization avoids the degeneracy of the earlier ones-initialization, which collapsed Σ and inflated Π.

BLP π matrices. Two distinct π patterns emerge across instruments:

Pattern A (Nevo, FL-level, FL-poly): The constant row shows a large negative income coefficient ($\pi_{1,\text{inc}} \approx -0.10$) and moderate negative age coefficient ($\pi_{1,\text{age}} \approx -0.03$), implying higher-income and older households have lower baseline utility for inside goods. Protein is consistently positive.

Pattern B (GH, FL, Combined): The constant row is much smaller ($|\pi_{1,\text{inc}}| < 0.02$), suggesting these instrument sets do not identify the demographic baseline shift. Other rows are broadly similar to Pattern A but noisier.

Across all specifications, $\pi_{\text{fat,income}} \approx -0.012$ consistently: higher income is associated with lower fat preference. Sodium entries are negligible (< 0.001).

Rank of the BLP substitution matrix. Table 14 reports the eigenvalue decomposition of BLP's implied cross-price matrix Σ across all seven instrument specifications, using the full demographic specification with π taste shifters. The leading eigenvalue captures 93.4–93.5% of the squared eigenvalue magnitude in every case, and the ratio λ_1/λ_2 ranges from 67 to 125. The effectively rank-one substitution structure is invariant to the choice of instruments.

Table 14. Eigenvalue decomposition of the BLP structural cross-price matrix Σ across instrument specifications. All estimates use the full demographic specification with π taste shifters.

Instruments	λ_1	λ_2	λ_3	λ_1/λ_2	$\lambda_1^2/\sum\lambda^2$
FL	+0.307	+0.003	−0.002	98	0.934
FL-level	+0.347	+0.004	−0.002	85	0.935
FL-poly	+0.391	+0.005	−0.002	83	0.935
GH	+0.335	+0.005	−0.002	67	0.934
GH-poly	+0.390	+0.005	+0.000	82	0.934
Nevo	+0.341	+0.005	−0.002	73	0.934
Combined	+0.314	+0.003	−0.002	125	0.935

Notes: λ_1/λ_2 is the ratio of the two largest eigenvalues. The last column provides the cumulative variance at K=1 i.e. the share of total squared eigenvalue magnitude captured by the leading eigenvalue.

Figure 3. Cross-price effects for BLP and K-Agg (FL, Π, $K = 2$) against CES