The Consumer-Product Characteristic Project*

Laurence Ales **Rebecca Lessem** **Christoper I. Telmer**
Ariel Zetlin-Jones
Tepper School of Business, Carnegie Mellon University

This Version: June 25, 2025

Abstract

We develop a framework for representing the very high dimensional set of consumer goods and services in terms of a lower-dimensional set of *characteristics*. The approach is analogous to O*NET (a database of occupational characteristics) for occupations. The framework expands the traditional taxonomy of durable goods, non-durable goods and services to incorporate many more characteristics of the consumer's consumption bundle. The framework consists of 25 characteristics selected from considerations based on economic theory, existing empirical work and expert analysis. Within this framework any good or service can then be represented as a 25-dimensional vector in the space of these characteristics. We use the framework to explain patterns for aggregate consumption. We first look at secular changes for the U.S. between 1959 and 2023. This time period is characterized by a broad movement from durables towards services. With our approach we show that goods and services that are characterized by a high level of *keeping up with the Joneses'* features experience a significant growth in consumption share. We also look at the behavior of the consumption basket during the 2020 pandemic. In this case the behavior of consumption during the contraction and recovery can be explained by looking at characteristics that center on the interaction between humans during the consumption phase.

1 Introduction

A class of differentiated products is completely described by a vector of objectively measured characteristics.

S. Rosen

In analyzing consumer behavior economists typically take one of two approaches. They either analyze very specific goods in isolation — *e.g.*, automobiles, housing, education — or they analyze broad categories of goods, the totality of which captures the consumer's overall *consumption*

*E-mail: ales@cmu.edu (Ales), rlessem@andrew.cmu.edu (Lessem), chris.telmer@cmu.edu (Telmer), azj@andrew.cmu.edu (Zetlin-Jones). We thank Jeff Galak, Gianluca Violante for helpful discussions. This research is supported by grants from CMU's Block Center for Technology and Society and the Richard King Mellon Foundation.

bundle. The most common example of the latter approach is the categorization of durable goods, non-durable goods, and services. Our research attempts to find a middle ground between fine detail and broad categories. We develop a model of the *characteristics* of goods and services that goes far beyond the 'traditional' taxonomy of durables, non-durables and services, yet is also sufficiently parsimonious so as to be useful for the study of consumption. An example of the limitation of the traditional approach might be helpful in justifying our methodology.

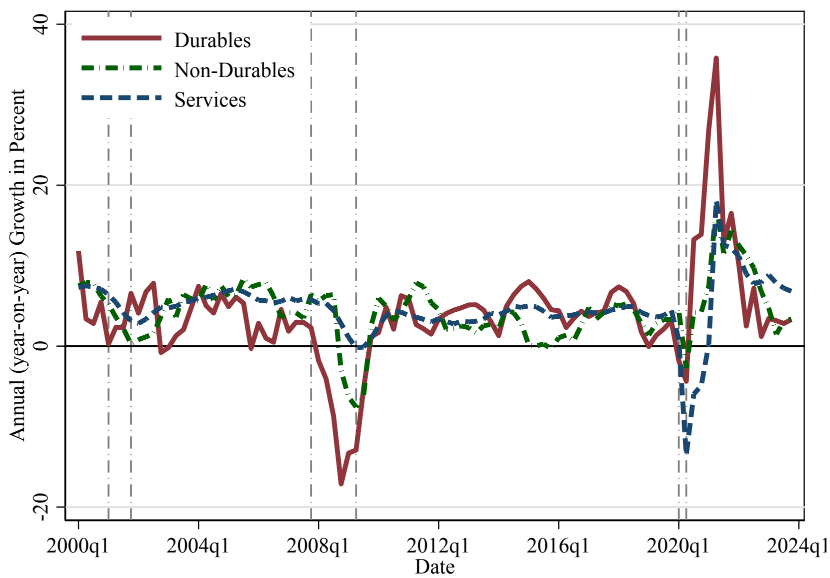

Figure 1: Growth rates of Durables, Non-Durables and Services 2000-2024. Source: Bureau of Economic Analysis.

Figure 1 shows that the traditional categorization taxonomy broke down during the COVID-19 pandemic in terms of its ability to predict how the components of the consumption bundle behave during a macroeconomic downturn. Consumption expenditure on durable goods, for instance, has been strongly pro-cyclical in the past, yet it behaved in the opposite manner throughout much of 2020. A similar phenomenon is manifest in the 2020 collapse in expenditure on services, a category that has been relatively stable throughout expansions and downturns. Within broad categories we also see disparate behavior. For example in the 2020 recession we saw a contraction for some durables such as autos but expansion for some durables such as furniture. Likewise, some services behaved as services whereas others behaved like durables. Spending on services such as meals in restaurants and accommodation collapsed whereas that on communication services was largely unaffected.

Figure 2 takes a longer time perspective and looks at the expenditures and prices of some durable goods since 1959. The left panel focuses on expenditure shares, for the selected goods we noticed that cars & trucks are roughly stable over the time period but appliances and sporting equipment have (respectively) a large decline and increase. On the right panel of Figure 2 we look at prices. The large increase of prices for cars & trucks together with a (roughly) constant

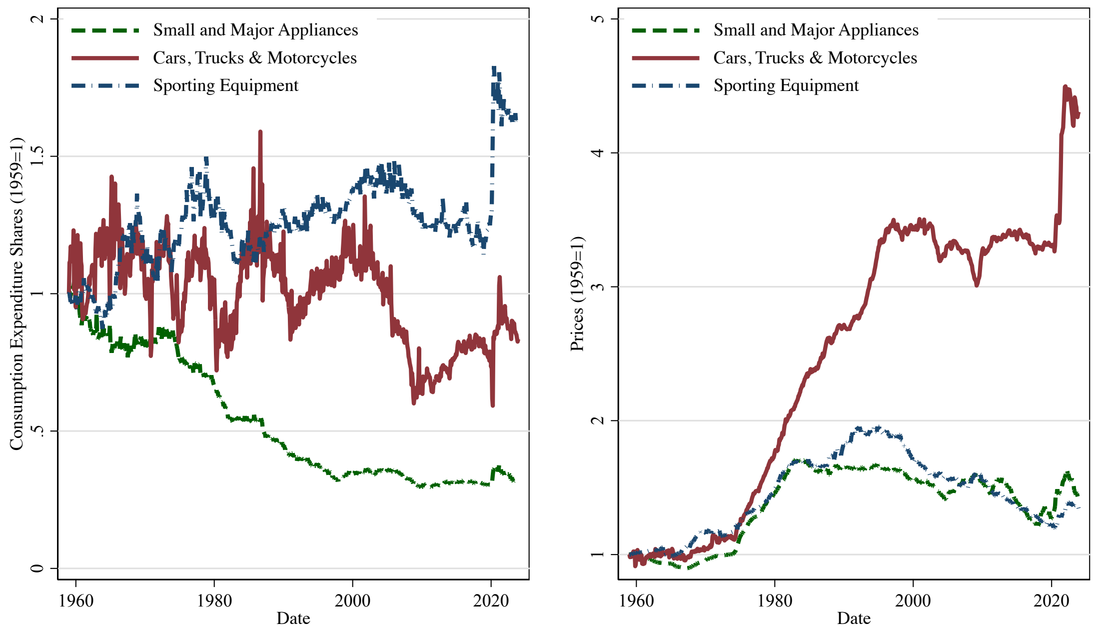

Figure 2: Expenditures and Prices of Some Durable Goods Categories Over Time. Source: Bureau of Economic Analysis.

expenditure shares points to an elasticity of substitution roughly equal to one. On the other hand the similar increase in price for both appliances and sporting equipment, paired with shares that evolve in opposite directions, points to two very different degrees of complementarity and substitutability for these two goods. While any expenditure data can be rationalized with sufficient flexibility in terms of elasticity of substitution this approach is ad-hoc and unsatisfactory. In this paper we propose taking a closer look at what makes goods like the ones described in Figure 2 similar or different trying to explain why a different behavior might be observed over time.

In this paper we describe each consumer good in terms of an N-dimensional vector of *characteristics*. The idea is that N (equal to 25 in the paper) is much smaller than the large number of goods, M, and that knowledge of a good's particular vector of characteristics (and prices) are sufficient to describe consumer behavior. Going back to the behavior described in Figure 1, we will show that the characteristic *Physical Human Interaction Required to Use or Consume*[1] is an important characteristic to explain the behavior during the time period 2020-2021. Looking at the time period 1959-2023 we show that the characteristic *Keeping Up With the Joneses*[2] is an important characteristic to explain the changing nature of the consumption basket.

[1] A product requires physical human interaction for use if a person experiences close physical contact with at least one individual (outside their family) during the typical use of the product. Products that require physical human interaction may feature more (or less) intensive interactions, meaning that the interactions are with a greater (fewer) number of persons or for a longer (shorter) period of time

[2] A *Keeping Up With the Joneses* good is one for which other people's level of consumption is clearly known, and for which the consumer suffers a utility penalty for being far from (above or below) a thresh- old level such as the average. The marginal penalty is increasing as the difference gets larger.

In what follows, Section 2 introduces the theoretical methodology of the paper as well as the structure of the taxonomy. Section 3 provides descriptions of the characteristics that constitute our goods taxonomy. The section concludes describing how the characteristics are measured for each good. Section 4 analyzes the set of characteristics and provides an analysis of the consumption bundle during 2019-2022 and 1959-2023 time periods. Section 5 concludes.

2 Methodology

2.1 Goods as Bundles of Characteristics

Consider an environment with M goods.[3] Consumers are heterogenous and described by a set of characteristics Ψ. A consumer, with income Y chooses goods, q, taking prices, p, as given. As notation denote with boldface the vector of M goods or prices. The standard version of the consumer's choice problem is given by:

$$\max_{\mathbf{q}} \ \tilde{U}(\mathbf{q}; \Psi), \quad s.t. \ \ \mathbf{p} \cdot \mathbf{q} \leq Y. \tag{1}$$

In this specification note that: (i) $\mathbf{q}$ is a high dimensional object, representing each and every good that is available in goods markets, and (ii) the consumer has a specific preference for each good in this vast set. In an influential paper, Lancaster (1966) makes a compelling argument for an alternative approach aimed explaining key drivers behind consumption choices and in the process at lowering the dimensionality of the problem. The key novelty of the approach is to assume that the utility function and the budget constrain are not defined in the same space. The consumer problem is re-written as follows:

$$\max_{\mathbf{o},\mathbf{q}} \ \widehat{U}(\mathbf{o}; \Psi) \quad s.t. \quad \mathbf{o} = B \cdot \mathbf{q}, \qquad \mathbf{p} \cdot \mathbf{q} \leq Y, \tag{2}$$

where B is an $N \times M$ matrix and $\mathbf{o}$ is a O-dimensional vector of consumption objectives.[4] A key difference in the above is that while in (1) $\tilde{U} : \mathbb{R}^M \to \mathbb{R}$ in the above, $\widehat{U} : \mathbb{R}^O \to R$ with $O << M$. In other words, in Lancaster's model consumers do not value goods *per se*. Rather, goods are objects that one acquires in the marketplace which then serve as inputs into a process that yields the objects (objectives) that consumers really care about. A meal in a restaurant, for example, might be an element of $\mathbf{q}$. The consumer's budget constraint and goods prices determine the monetary cost of a meal purchase while a good's consumption value–the utility it provides–comes from the fact a meal is used to deliver to the consumer specific objectives, such as 'nutrition,' 'aesthetic appeal,' and 'social interaction.' The matrix B describes quantities of a given consumption objective each specific good provides to the consumer.

Lancaster's framework and results are a significant theoretical contribution. Applying the

[3] From here onwards the word 'goods' is to be understood as a stand-in for the more common 'goods and services.'

[4] Lancaster (1966) refers to these objectives as *characteristics*. In this paper we prefer to save the term *characteristics* for something more apropos.

framework, however, runs into difficulties. For example, several features of goods seem at best indirectly related to specific consumer objectives and more related to the nature of the good. Durability, for example, is obviously an important feature of many consumer goods, but is better thought of as a characteristic of the good itself rather than an object that the consumer derives utility directly from. Likewise some aspect of the goods itself might drive direct utility to the consumers. With this in mind we construct a consumer problem that is a hybrid combination of the traditional consumer problem (1) and the Lancaster formulation (7).[5]

We consider goods as heterogenous objects characterized by a set of characteristics and by allowing more flexibility on how a consumption goods impacts the utility of a consumer. In our formulation some of the characteristics consumers care directly about while others consumers care about because they influence the relationship between goods, their prices, and how they deliver specific consumption objectives. This departure from Lancaster's framework allows us to distinguish between features of goods themselves and features of goods that consumers care about. Formally, consumers derive utility directly from consumption objectives $\mathbf{o}$ as well as from the good purchased $\mathbf{q}$. The consumer maximization problem is written as:

$$\max_{\mathbf{o},\mathbf{q}} U(\mathbf{o}; f_I(\mathbf{q};\phi);\Psi) \quad s.t. \quad \begin{aligned} \mathbf{o} &= f_{II}(\mathbf{q};\delta), \\ \mathbf{p}\cdot\mathbf{q} &\leq Y. \end{aligned} \tag{3}$$

The major differences relative to the standard model described in (1) is that: (i) the set of consumption objectives is much smaller than the set of consumption goods; (ii) goods purchased enter directly in the utility function only via the function $f_I : \mathbb{R}^M \to \mathbb{R}^N$ and indirectly impacting consumption objectives by the function $f_{II}(\mathbf{q};\delta)$ using purchased goods as inputs.[6]

We can now define more precisely *characteristics* as anything that affects the solution to (3) that is neither price nor income. Broadly we consider three types of characteristics. The vector of characteristics Ψ are characteristics that are directly associated with the consumer, they captures how aspects of the consumer such as age, gender, location. The vector δ represents characteristics (like durability) that are not valued by consumers per se but only because they impact the creation of consumption objectives. Finally the vector of characteristics ϕ represents characteristics (like conspicuousness) that the consumer directly cares about. We assume a total of N characteristics, described by a vector $\mu \in \mathbb{R}^N$. The N-dimensional vector of characteristics is given by combined vectors $\mu = \{\phi, \delta, \Psi\}$, so the optimal solution for each good q_i^* with $i \in \{1, \ldots, M\}$, can be written as: $q(\mathbf{p}, Y, \mu)$.

Example 1. *Two elements of* $\mathbf{q}$ *might be expenditures for* new and used autos and motorcycles *(say*

[5]The original work of Lancaster (1966) and Lancaster (1971) aimed at establishing an "efficiency frontier"for consumer goods. The existence of such frontier would allow empirical researchers to determine the substitutability across goods without knowledge of consumer preferences. To establish this result the Lancaster formulation requires strong modeling assumption including assuming goods are a linear combination of characteristics and that each characteristic associated with a positive marginal effect on utility. Our model does not feature such restrictions. For further details, refer to Hendler (1975).

[6]We refer to f_{II} as the unbundling function since un-bundles purchased goods into consumption objectives.

an automobile) and for Public land transportation *(a train ticket). In our notation the function* $f_{II}(\mathbf{q};\delta)$ *maps each of these goods into a consumption objective,* o = *'transportation services,' from which the consumer derives utility. From this perspective, the two goods are the same. But from a broader perspective, they are obviously not. They produce the transportation services in a different manner. This is captured by the characteristic* δ*, one aspect of which measures durability. A car produces transportation over a longer period of time, so its* δ *is different from that of a train ticket. Continuing on this example, one of the characteristics* ϕ *might be 'conspicuousness.' Perhaps the car is a luxury car and the consumer derives conspicuous consumption services from it? Unlike* δ*,* ϕ *arises from the consumer's preference for the good itself — as opposed to consumption objectives that the good generates — hence* ϕ *and the associated* q *appear directly in the utility function. Similarly,* Ψ *also appears in the utility function. It captures how aspects of the consumer —* e.g.*, age, gender, location — affect that solution for* q*.*

The model described in (3) provides a low-dimensional representation of the goods space: goods can be described by the low-dimensional number of characteristics they possess. In this regard the model provides an organizational device — a taxonomy — with which we can house the set of characteristics that constitute a basis for the goods space.[7] We expand on this next.

2.2 A Taxonomy for Goods

To understand the approach of introducing a taxonomy for goods, it is helpful to relate it with to the O*NET Database, a taxonomy of the characteristics of occupations. For the O*NET the development of the characteristics for occupation evolved organically over time.[8] Our approach in developing the taxonomy relies on the model described in the previous section. Following the O*NET, we refer to our taxonomy as our *Content Model*. It is depicted in Figure 3. Its skeleton is defined along two dimensions. First, on the vertical dimension, we differentiate between characteristics that directly impact the consumer (aspects connected with Ψ and f_I) and characteristics that indirectly impact the consumer (the unbundling function f_{II}). Second, on the horizontal dimension, the distinction between how a good is purchased and how it consumed/used. That is asking if a particular characteristic describes the act of shopping for a good or the act of consuming/using the good. These dimensions are formulated in terms of six *domains*, each of which is represented as a square in Figure 3. Each domain is populated with a set of characteristics. Figure 3 illustrates a subset of the characteristics that constitute our model, Table 1 provides the full set. We next elaborate on each of the domains in Figure 3 and provide some representative examples. We begin with the northwest square and proceed in a clockwise direction. In each case we label the domain and state which of the characteristics from our model it represents.

[7] Lancaster (1966) highlights how the standard model is not a useful starting point for building a product taxonomy as each good is different only with respect to the preferences. While Lancaster (1966) recognizes the possibility of using to build a taxonomy it never operationalized this step.

[8] The precursor of the O*NET, the *Dictionary of Occupational Titles* (DOT), was initiated in 1939. By the 1990s the number of titles in the DOT grew to roughly 12,000. In 1996 the O*NET was born as the offspring of the DOT. The number of occupational titles was condensed to a list of 1,102, which was linked to a taxonomy of occupational characteristics that now numbers 277.

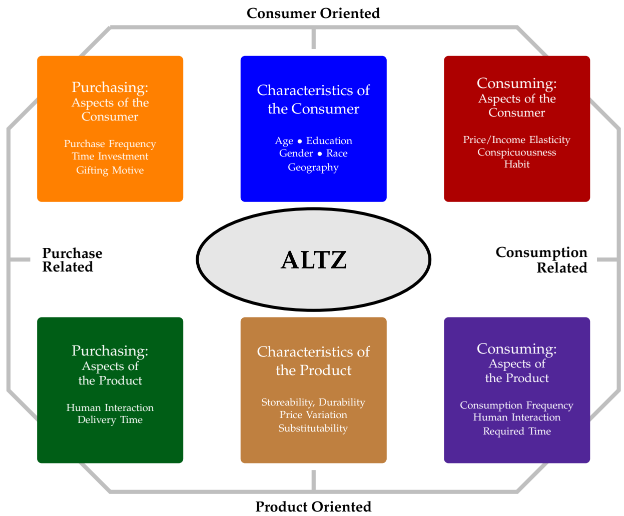

Figure 3: The Ales, Lessem, Telmer, Zetlin-Jones (2025) Content Model.

- *Characteristics of the Consumer* (Ψ). Consumer demand, $q(\mathbf{p}, Y; \phi, \Psi, \delta)$, depends on characteristics of the consumer herself: age, education, gender, race, geographic location and so on. Given this the vector Ψ are not just a characteristics of the consumer, they are also characteristics of the goods.

- *Direct Impact on the Consumer* (ϕ) These characteristics directly impact the consumer and are indeed associated with the consumer herself than the good itself. We can further split the characteristics depending if they are associated with the purchase or consumption of the good.

  - *Purchasing Related* These are characteristics associated with the purchase of the good directly. (ϕ). An example are characteristics associated with the likelihood that the good is purchased for gift-giving. Another set of characteristics are related to whether goods are *Search, Experience or Credence* goods.
  - *Consuming Related* These characteristics describe consumption behavior that is primarily associated with how the consumer derives utility from consuming the good *itself*. For example, an automobile may give the consumer utility that derives from the car itself, rather than the bundle of consumption objectives that constitute a car. A luxury car, for example, may yield utility in the form of conspicuous consumption. *Conspicuousness*, therefore, would be an element of the set of ϕ characteristics. It is associated with the good, not the bundle delivered by the good, and it is a modifier that appears in the utility function, not the unbundling function. Another characteristic in this domain that will be important later on is *Keeping Up With the Joneses*.

- *Indirect Impact on the Consumer* (δ) These characteristics impact the consumer only indirectly as they affect the unbundling of consumption goods into consumption objectives via the function f_{II}. Continuing with the above example of the automobile. The good gets "unbundled"into consumption objectives such as transportation services and safety. The same is true of a train ticket. In both cases, it is the transportation and safety activities — elements of c — that give rise to utility. But *how* these different goods get unbundled into the same consumption objectives is different. This is captured by δ, a modifier in the unbundling function. The characteristic *Durability*, for example, is part of the unbundling function that maps the car into transportation services. The same is true of the lack of durability of the train ticket. Other characteristics that are included in δ include *Excludability*, *Rivalrousness* and *With-Product-Heterogeneity*. As before we can further differentiate if the characteristics, while related to the product, they are more relevant during the purchase or consumption of the product.

- *Consuming Related* One example relates to the COVID pandemic. Using a car posed a smaller risk of infection than using a train ticket. The same was true of a restaurant meal compared to groceries. The characteristic *Physical Human Interaction Required to Use/Consume* captures the associated generation of activities connected to human interaction.
- *Purchasing Related* Continuing on the previous example we can put in this category characteristics such as *Physical Human Interaction Required to Purchase/Receive*. *Delivery Time* is another characteristic that is clearly in this category.

3 Measurement

To implement our content model in practice, we proceed in three steps: (I) define the set of characteristics $\{\phi, \delta, \Psi\}$; (II) define the set of goods to be represented in terms of the characteristics; (III) obtain quantitative estimates of the characteristics for each of the goods: assigning a value between 0 and 1 for each good and for each characteristic.

Example 2. *Consider one particular good: a meal in a restaurant. One of the consumer-driven characteristic, ϕ, might be* Conspicuousness, *the extent to which consumer demand is driven by a desire to have conspicuous consumption. We represent this as a number between 0 and 1, say $\phi = 0.60$, indicating that the consumer does indeed derive utility from being seen in a restaurant, but less than the utility from being seen wearing designer jewelry, which, for example, has $\phi = 0.90$. A different characteristic might be* Physical Interaction Required to Purchase or Receive, *an element of the vector δ. A good that can be acquired without any human interaction at all* (e.g., *through online shopping*) *is represented by $\delta = 0$, whereas $\delta = 1$ indicates the opposite. Returning to the restaurant meal, it might have $\delta = 0.7$, indicating that human interaction is hard to avoid, but less so than, say, for a haircut, which has $\delta = 0.95$.*

We now provide more details on each step.

3.1 Defining the Set of Characteristics

To generate the list and definitions for characteristics we proceed by using a combination of prior knowledge, expert analysis and survey data. In detail:

1. We assembled a focus group of Carnegie Mellon University undergraduates as a part of a course. The students, supervised by us, formulated some of the elements of our characteristics set based on both personal experience and their research. In addition, they conducted online surveys (having taken the requisite amount of *Institutional Review Board* training) using their personal networks as respondents to determine any gaps in the set of characteristics.

2. We incorporate insights from academic sources ranging from detailed studies of price and income elasticities (e.g., Taylor and Houthakker (2009)), to the literature on 'memory goods' (e.g., Gilboa et al. (2016), Hai et al. (2020)), to the framework based on search, experience and credence goods (e.g., Ford et al. (1988)).

3. We employed a professional surveying firm to conduct consumer surveys in order to map several of our characteristics into quantitative terms for a set of Bureau of Economic Analysis (BEA) goods and services. A helpful by-product of this surveying exercise was to provide insights used to both expand and to prune our set of characteristics.

4. Finally, we used our own expertise in order to define certain characteristics. For example, the COVID-19 pandemic and its ensuing disruption of a number of consumption regularities motivated us to include characteristics such as *Human Interaction Required to Consume*.

This process resulted in a final set of 25 characteristics, not including demographic characteristics that we directly associate with the consumer herself. The characteristics are listed in Table 1.

Our companion technical document provides granular details for how each characteristic is defined and measured.[9] Overall, we don't view our set of characteristics as being in any way final. We view our characteristics in much the same way as the profession has come to view the O*NET's occupational characteristics, as an evolving set that errs on the side of casting a wide net and which, ultimately, relies on empirical verification to define its elements.

3.2 Defining the Set of Goods

Our set of goods derives from the BEA table, *Table 2.4.5U. Personal Consumption Expenditures by Type of Product*. The goal is to settle on a number of broad consumption categories that represents a balance between providing dis-aggregation and parsimony. The traditional BEA taxonomy of *three* goods — Durables, Non-Durables and Services — leans too far toward parsimony. The highest level of disaggregation in the BEA table features 366 goods, thus leaning too far the other way. We settled on 69 distinct goods and services and aggregated the remaining goods accordingly. The final list of consumption goods is in Appendix C.

[9]The technical document is available here: CPCPTechnicalDocument.pdf.

Table 1: Twenty Five Characteristics of the ALTZ Content Model.

Domain and Characteristic	Abbreviation
Consuming: Aspects of the Consumer	
Conspicuousness	Conspic
Keeping Up With the Joneses	KUWtJ
Brandedness	Brand
Income Sensitivity	IncElast
Price Sensitivity	PriceElast
Subsistence Level	Subsist
Susceptibility to Habit Formation	Habit
Predictable Time Variation: Preferences	TVPrefs
Purchasing: Aspects of the Consumer	
Gifting Products	Gift
Philanthropy and the Provision of Public Goods	Charity
Shopping Frequency	ShopFreq
Time-Investment Required for Purchase	PurchInv
Search, Experience, Credence	SEC
Purchasing: Aspects of the Product/Market	
Physical Human Interaction Required to Purchase or Receive	PurchHuman
Time Required Between Purchase and Receipt	PurchTimeLag
Characteristics of the Product	
Tangibility	Tangible
Excludability	Exclude
Rivalrousness	Rival
Durability	Durable
Storability	Storable
Within-Product Heterogeneity	ProdHet
Predictable Time Variation: Production	TVProdcn
Consuming: Aspects of the Product	
Frequency of Consumption/Use	ConsFreq
Physical Human Interaction Required to Use or Consume	ConsHuman
Time-Investment Required for Use	ConsInv

3.3 Estimating the Characteristic Vector for Each of the Goods

We follow several different approaches in order to associate numerical values with each of our 25 characteristics, for each of the 69 goods described above. For the two elasticity-based characteristics, *Price Sensitivity* and *Income Sensitivity*, we obtained direct estimates from well-known sources: Taylor and Houthakker (2009). For *Time-Investment Required for Purchase* and *Shopping Frequency* we used the consumption surveys in the *Survey of Consumer Finances*. For *Durability* we began with the traditional BEA taxonomy and then used our own judgment to expand the scope to include, for instance, some BEA-defined services such as health care, which are obviously durable-good expenditures. Finally, we used what O*NET calls 'expert opinion' on the remainder of our characteristics, where we, the authors, served as the experts. In each case we began by outlining, *a priori*, an algorithm for arriving at our estimates. Then we took 2-3 iterations through our algorithm with an increasing amount of discussion and feedback meant to refine our estimates. Our technical document describes this approach, for each characteristic, in detail.

4 Results

In this section we provide some quantitative applications and analysis of the content model introduced in previous sections. We first take a closer look at the matrix of characteristics; we then analyze the changing consumption bundle of households during the 2020-2022 pandemic and recovery; we then conclude by looking at the historical evolution of goods and services purchased in the U.S.

4.1 Initial Analysis Of Characteristics

As a first step we look at the relationship between the characteristics of our taxonomy. To do so we use the information contained in the content model that evaluates our 25 characteristics over the 69 aggregated BEA products. As a first step we standardize the values for each characteristic. Overall, we observe cross-correlations between 0.76 and -0.54. While we observe no perfect correlation across characteristics, certain characteristics are closely related.[10] For example, not surprisingly, *Durability* and *Storability* are the most correlated characteristics while the most negatively correlated characteristics are (also not surprisingly) *Time-Investment Required for Purchase* and *Shopping Frequency*. Figure 8 in Appendix B displays the entire correlation matrix.

[10] Of course the degree of correlation depends critically on the number and type of products. It is clear that working with a larger number of more detailed product categories would lower the correlation between characteristics while working at a more aggregated level would raise the correlation.

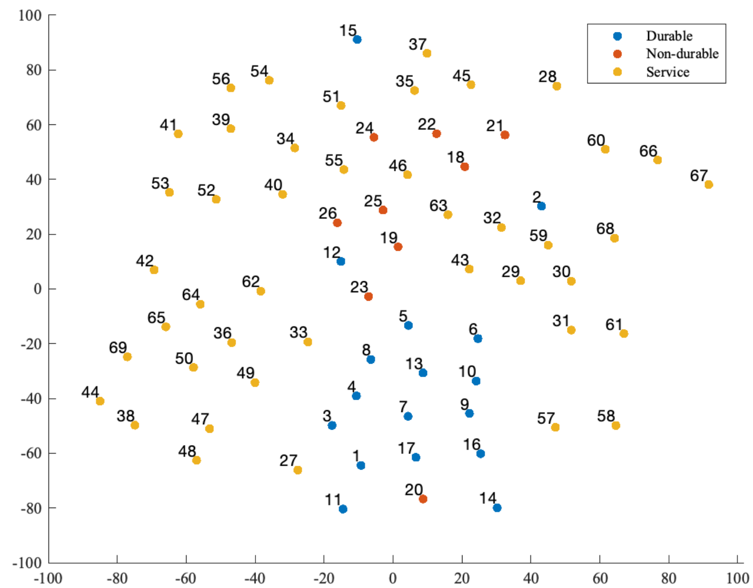

Figure 4: T-distributed Stochastic Neighbor Embedding of ALTZ Products and Characteristics. Numbers in the figure represent the ALTZ product numbers.

We next look at how the characteristics in our content model relate to the commonly used classification of the Bureau of Economic Analysis (BEA). The standard BEA classification usually considers products divided in one of three broad categories: durables, non-durables, and services. The definition from the BEA states that durables are: *"Tangible products that can be stored or inventoried and that have an average life of at least three years."* It is clear that the BEA definition, from the perspective of the ALTZ model, overlaps several characteristics of products that our content model identifies. For example, the definition includes criteria based on *Tangibility*, *Storability*, and *Durability*. Similarly, services are defined by the BEA as *"Products that cannot be stored and are consumed at the place and time of their purchase."* thus overlapping product characteristics such as *Storability* and *Time Required Between Purchase and Receipt*.

We can use the characteristics in the taxonomy to better understand the classification among durables, non-durables and services of the BEA. We look at the 69 aggregate products. Each of the aggregated products has a natural relationship to the underlying BEA product lists.[11] A difficulty in expressing the relationship is that each product is described by a point in a 25-dimensional space (each dimension representing each characteristic). To aid in the visualization we use a t-distributed stochastic neighbor embedding algorithm (t-SNE) to reduce the dimensionality of the dataset to two dimensions. The results are displayed in Figure 4; in the diagram products close to each other are "close" in the 25-dimensional space. Each product is labeled with a number signifying the ALTZ product number. In the figure we also overlay the BEA classification with

[11]Refer to the online appendix for a detailed crosswalk between BEA products and the aggregated products used in this project.

different colors. While information on BEA classification is not used in the algorithm, the t-SNE plot seems to generate clusters of similar products. For example, *durables* appear close to each other. The diagram highlights the high heterogeneity of *services*. Indeed, some *services* appear to be more closely related to *durables* than to other types of *services*. Compare as example *Accommodations in hotels and motels*" (product number 49) and *Insurance* (product 53).[12] To formally

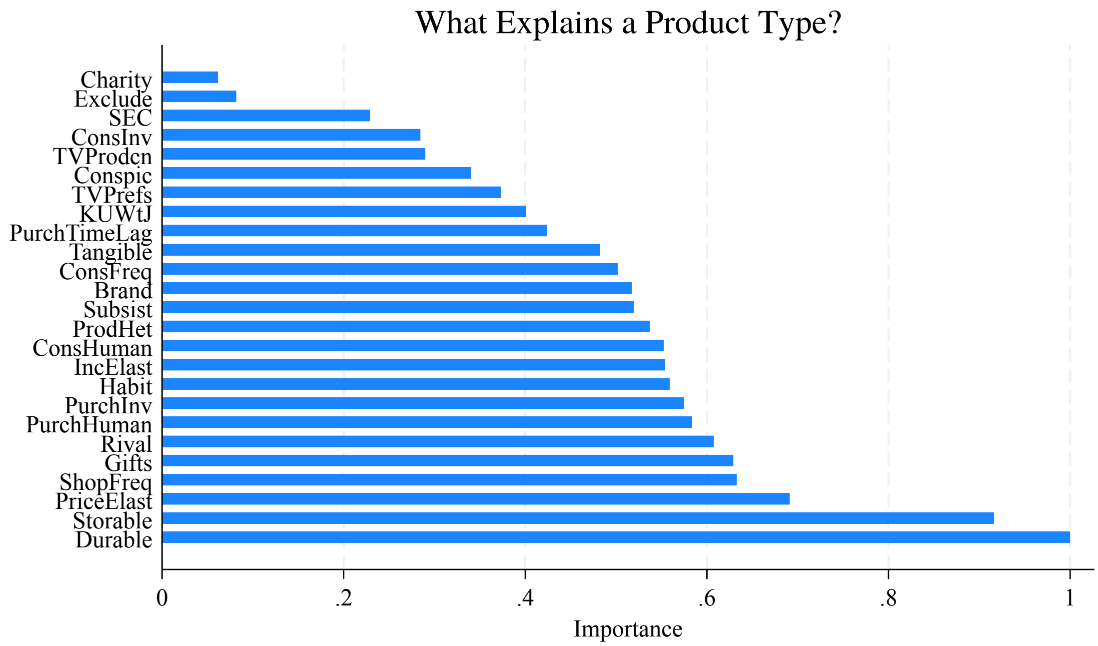

Figure 5: Random Forest Importance of Characteristics in Determining BEA Product Classification.

relate our characteristics to the ones used by the BEA, we use a decision-tree, random forest classifier (Breiman (2001)). The idea is that in a decision tree we can use our characteristics to classify a particular product. as a *durable*, *non-durable* or *service*. Ideally only a few of the characteristics should be sufficient to classify products. The random forest algorithm provides a measure of importance of each variable (in our case the characteristics) when trying to build a classifier for the type of BEA products. The results are displayed in Figure 5 (refer to Appendix **??** for reference on the abbreviations used in the Figure and elsewhere in this paper). We see here that the algorithm recognizes *Durability* and *Storability* as key variables in a decision tree to determine whether a product is a durable, non-durable and service based on the BEA definition. Indeed above we highlighted how these two characteristics are implicit in the definition of types of goods used by BEA. However, from Figure 5, we also see that a large number of variables are

[12]Care should be used in interpreting t-SNE plots as they try to represent high dimensional objects in lower dimension. As an example of potential distortion induced in this process, consider that the closest products (in terms of Euclidean distance in the original 25-dimensional space) are *Alcohol at Eating/Drinking Places* (48) and *Meals at full-service restaurants* (47); this seems to be consistent with the diagram in Figure 4. On the other hand, the furthest apart goods are *Household utilities* (28) and *Jewelry and watches* (14) these are not the furthest apart products in Figure 4.

needed for a correct classification of the goods. This points to the large heterogeneity in products in each of the three classifications.

4.2 Aggregate Consumption Analysis: The 2019-2021 Pandemic

The framework developed in this paper can be of help to understand the evolution of consumption patterns over time. We begin by looking at the behavior of aggregate consumption during the 2019-2021 COVID pandemic. As anticipated in the introduction, the 2019-2021 period saw a large collapse and quick recovery of durables and, differently than previous recessions, a large collapse and slow recovery for services. Our starting point is the model described in (3) substituting out c in the utility function. In what follows q_{it} is real expenditures on the i-th product at time t. Our parametrization of the aggregate consumer problem at time t is given by:

$$\max_{q_{1t},\ldots,q_{Mt}} \left[\sum_{i=1}^{M} a_{it}^{\frac{1}{\sigma}} q_{it}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

subject to:

$$\sum_{i=1}^{M} p_{it} q_{it} = Y_t, \forall t.$$

Where p_{it} is the price for consumption good i in period t and Y_t are financial resources available in period t. Relative to the original problem in (3) we now allow the utility function U and the function governing the relationship between good and activities to vary over time. In a simple way, in the objective function, the weights a_{it} now represent the preference shifter for good i at time t. We next show how to recover weights a_{it} from aggregate consumption data. From first order conditions of the above problem we get:

$$s_{it} = \frac{a_{it} p_{it}^{1-\sigma}}{\sum_j a_{jt} p_{jt}^{1-\sigma}}, \quad \forall i = 1, \ldots, M, \forall t. \tag{4}$$

Where $s_{it} = \frac{p_{it} q_{it}}{W_t}$ is the share of total expenditures incurred in good i. Taking logs and differentiating between two time periods t and t' we get:

$$\log \frac{s_{it'}}{s_{it}} = \log \frac{a_{it'}}{a_{it}} + (1-\sigma) \log \frac{p_{it'}}{p_{it}} + \Gamma_{t',t}, \quad \forall i = 1, \ldots, M \tag{5}$$

With $\Gamma_{t',t}$ a constant depending on t and t'. The above equation can be used to recover the relative weights for consumption over time. As a final step we connect the recovered preference shifters to our 25-characteristics by estimating the following:

$$\log \frac{a_{it'}}{a_{it}} + \Gamma_{t',t} = \beta_0 + \sum_{j \in J} \beta_j^c \cdot X_{i,j}, \quad \forall\ i = 1, \ldots, M. \tag{6}$$

In the above matrix $X_{i,j}$ is the dataset in our content model denoting the value of characteristic j for good i. In (6) the coefficients β_j^c denote the role that the $j-$th characteristic has in explaining movement in the preference shifters.

To study the COVID pandemic we are interested in both the initial drop and recovery. For the "drop" we set t' as April 2020 and t as April 2019. For the "recovery" we set t' as April 2021 and t as April 2020. Our data source for expenditures and prices is the BEA.[13] Finally we use a value of $\sigma = 0.89$ taken from Herrendorf et al. (2013) (in Appendix B we consider the Cobb-Douglas case with $\sigma = 1$). To start, Table 2 and Table 3 look at an estimation of equation (6) on fewer chosen characteristics. The selection of the first set of characteristics is motivated by the different behavior of durables, non-durables and services during the pandemic. The characteristics, given the results in Figure (5), represent the four most important characteristics in explaining the BEA product classification. The remaining two characteristics, motivated by the nature of the pandemic itself, relate to the interaction with other humans during the consumption (*ConsHuman*) and purchase (*PurchHuman*) of products. Table 2 looks at the beginning of the pandemic,

Table 2: Estimates of β_j^c in equation (6) between April 2019 and April 2020. Estimates based on the 69 products of the taxonomy.

	(1)		(2)	
Durable	0.162	(0.111)	0.193*	(0.0913)
Storable	0.0422	(0.112)	-0.0443	(0.0916)
PriceElast	0.0436	(0.0744)	-0.0407	(0.0611)
ConsHuman			-0.299**	(0.0868)
PurchHuman			-0.103	(0.0849)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the "Drop". In column (1) of Table 2 we see how the four characteristics used to explain the BEA good classification fail at providing any significant explanatory power in the behavior of products during the pandemic. In column (2) we add the characteristics relating to interaction with humans. In this case the characteristic *ConsHuman* is the only significant variable (notably interaction in purchase, *PurchHuman* is also not significant). The estimates for *ConsHuman* are intuitive: a higher degree of human interaction in consumption is associated with a decrease in expenditure shares during the "drop"-phase of the pandemic. Table 3 looks at later stages in the pandemic, the "Recovery". Similarly to the previous case here too *ConsHuman* plays an important role in explaining how different products recovered after the initial set of pandemic lock-downs.

In the previous regressions we chose specific characteristics motivated by the nature of the events in the 2019-2021 pandemic. We next deploy a stepwise regression model to automatically

[13]Specifically we use table 2.4.4U for Price Indexes and Table 2.4.5U for Personal Consumption Expenditures by Type of Product. We look at data reported at a monthly frequency

Table 3: Estimates of β_j^c in equation (6) between April 2020 and April 2021. Estimates based on the 69 products of the taxonomy.

	(1)		(2)	
Durable	-0.0517	(0.0812)	-0.0596	(0.0716)
Storable	0.00507	(0.0821)	0.0516	(0.0718)
PriceElast	-0.0420	(0.0544)	0.0158	(0.0479)
ConsHuman			0.227**	(0.0681)
PurchHuman			0.0223	(0.0666)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

select, now looking at all the characteristics in the taxonomy, which ones are the most relevant when explaining the decline and recovery of consumption during the pandemic. In Table 4

Table 4: Stepwise regression estimates of β_j^c in equation (6) between April 2019 and April 2020. Estimates based on the 69 Products of the taxonomy. Column (1): forward selection with threshold significance level of 0.01. Column (2): forward selection with threshold significance level of 0.05.

	(1)		(2)	
ConsHuman	-0.386***	(0.000)	-0.227**	(0.001)
Conspic			-0.188**	(0.005)
Durable			0.208***	(0.000)
Charity			0.202***	(0.000)
ConsFreq			0.157*	(0.011)

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

we look at the initial drop in the pandemic. Specifically the estimates are derived with a from a forward selection criterion, in this case the algorithm starts with only a constant and progressively adds characteristics if the p-value is below a threshold (0.01 and 0.05 in our case). From Table 4 it emerges that *ConsHuman* is selected in both cases. In column (2), additional variables are selected as highly significant: for example *Charity* and *Durable*. Table 5 now looks at the recovery period. As for the previous case, the characteristic *ConsHuman* is selected together with the characteristic describing rivalrousness *Rival*. The previous two tables look at the case where the elasticity of substitution is set exogenously at 0.89. In Appendix we consider the Cobb-Douglas case ($\sigma = 1$) similar results hold in this case too.

Table 5: Stepwise regression estimates of β_j^c in equation (6) between April 2020 and April 2021. Estimates based on the 69 Products of the taxonomy. Column (1): backward selection with threshold significance level of 0.01. Column (2): backward selection with threshold significance level of 0.05.

	(1)		(2)	
ConsHuman	0.244***	(0.000)	0.233***	(0.000)
Rival	0.139**	(0.003)	0.125**	(0.005)
Subsist			-0.105*	(0.019)

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.3 Aggregate Consumption Analysis: Historical Trends

We next take a longer time-horizon perspective. As before we look at BEA consumption and price data but now consider the time period between 1959 to 2023. Similar to the previous section we are interested in extracting information on preference shifts as described in equation (5) and project this information on characteristics as described in equation (6). In equation (5) we now consider t the average of the time period 1959-1964 and t' the average during 2018-2023.

In Figure 6 we display the changes over time of prices and consumption shares. In the figure we observe certain known facts. For example, most non-durables products exhibit a decline over time of the share of total consumption. We also see that most (but not all) services become more important in the consumption basket over time. As a next step, similar to what we have done

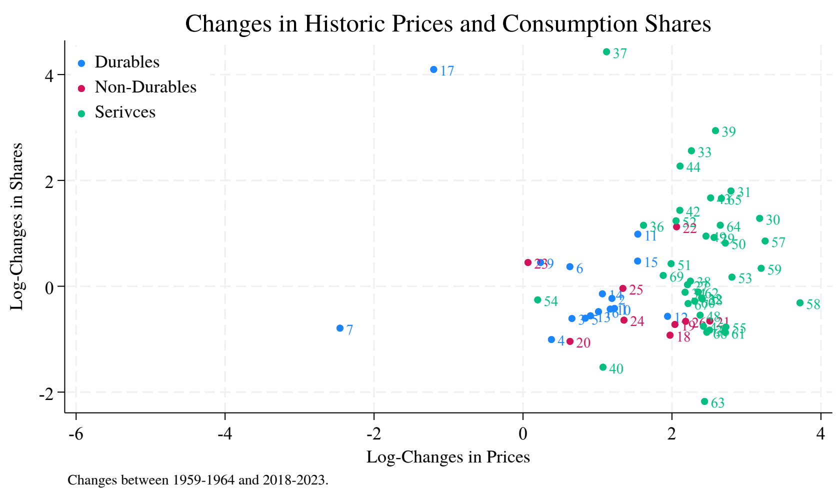

Figure 6: Historical changes in Prices and Consumption Shares. Numbers represent Product numbers for our content model.

in the previous section, we look at the relationship between the log $\frac{a_{it'}}{a_{it}}$ and the 25 characteristics. Table 6 looks at stepwise regression estimates. In the table we consider different combination

Table 6: Stepwise regression estimates of β_j^c in Equation(6) between 1959 and 2023. Estimates based on the 69 Products of the taxonomy. Column (1): fixed characteristics $\sigma = 0.89$; Column (2): fixed characteristics, $\sigma = 0.96$; Column (3): backward selection (0.1 threshold), $\sigma = 0.89$; Column (4): backward selection (0.1 threshold), $\sigma = 0.96$.

	(1)	(2)	(3)	(4)
Durable	-0.303	-0.294		-0.279
	(0.249)	(0.261)		(0.077)
Storable	0.220	0.180		
	(0.401)	(0.488)		
PriceElast	-0.126	-0.133		
	(0.486)	(0.461)		
Charity			-0.460*	-0.495*
			(0.027)	(0.015)
KUWtJ			0.481**	0.524**
			(0.008)	(0.004)
ConsFreq			-0.386*	-0.424**
			(0.020)	(0.009)
TVPrefs			-0.380*	-0.372*
			(0.035)	(0.034)

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

of algorithm (both forward and backward selection procedure) as well as different elasticity of substitution: we use both a value of $\sigma = 0.89$ as in the previous section then we follow the methodology of Herrendorf et al. (2013) and estimate a value of $\sigma = 0.96(0.076)$ using our own data. In the table we first consider characteristics that following our prior discussion best describe the BEA classification of goods between *durables*, *non-durables*, and *services*. As anticipated from Figure 6, these characteristics are unable to explain any trends. We then, as in the previous section, allow all of our characteristics to determine the evolution of $\frac{a_{it'}}{a_{it}}$. From the table now some characteristics become statistically significant. In particular the characteristic "Keeping Up With the Jones" (*KUWtJ*) emerges as a significant one.That is, goods that tend to score high in this characteristic, perhaps not surprisingly, see their share of consumption grow over time. In most specifications consumption frequency (*ConsFreq*) is selected with the opposite effect on shares of consumption over time.

5 Conclusion

The paper introduces a comprehensive framework for characterizing consumer goods and services through a reduced set of 25 characteristics, offering a refined lens for analyzing consumption patterns. This approach bridges the gap between detailed product-level analysis and broad consumption categories, providing a more granular understanding of consumer behavior. Our empirical findings underscore the importance of characteristics such as "Keeping Up With the Joneses" and "Physical Human Interaction Required to Use or Consume" in driving consumption trends, particularly during the COVID-19 pandemic and over the extended period from 1959 to 2023. This framework not only elucidates historical consumption dynamics but also holds significant implications for future research and policy formulation aimed at addressing the evolving landscape of consumer demand.

References

Breiman, L. (2001). Random forests. *Machine learning 45*, 5–32.

Ford, G. T., D. B. Smith, and J. L. Swasy (1988). An empirical test of the search, experience and credence attributes framework. *Advances in Consumer Research 15*(1).

Gilboa, I., A. Postlewaite, and L. Samuelson (2016). Memorable consumption. *Journal of Economic Theory 165*, 414–455.

Hai, R., D. Krueger, and A. Postlewaite (2020). On the welfare cost of consumption fluctuations in the presence of memorable goods. *Quantitative Economics 11*(4), 1177–1214.

Hendler, R. (1975). Lancaster's new approach to consumer demand and its limitations. *The American Economic Review 65*(1), 194–199.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2013). Two perspectives on preferences and structural transformation. *American Economic Review 103*(7), 2752–2789.

Lancaster, K. (1971). Consumer demand: A new approach.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy 74*(2), 132–157.

Taylor, L. D. and H. S. Houthakker (2009). *Consumer demand in the United States: Prices, income, and consumption behavior*. Springer Science & Business Media.

Tippins, N. T. and M. L. Hilton (2010). A database for a changing economy: Review of the occupational information network (o*net).

Appendix

A The O*NET database

In this section we provide a brief summary of the O*NET content model and its historical development.[14] The O*NET content model is a hierarchical taxonomy of occupational characteristics comprised of five levels. The highest level is a set of six *domains*. These six domains are organized using a two-dimensional conceptual framework that classifies characteristics in terms of their worker-versus-job orientation and in terms of their degree of occupational specificity. Figure 7 provides a visualization.

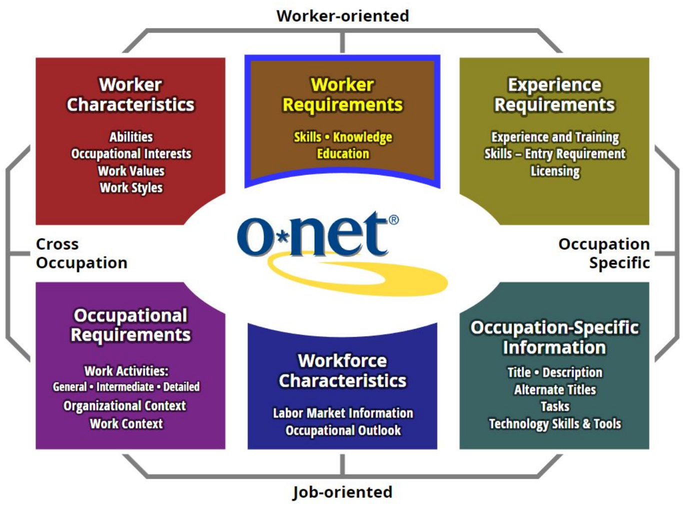

Figure 7: The O*NET content model.

Level-1 domains are at the top of each box, followed by Level-2 Domains. Below Level-2 there are three levels of characteristics, the first two of which are parents of families of sub-characteristics. In total the model is comprised of 277 descriptors, in addition to detailed occupational information collected by the Bureau of Labor Statistics (BLS).

The development of the O*NET content model followed a bottom-up approach. Its roots lie in the *Dictionary of Occupational Titles* (DOT) which first appeared in 1939. By 1990 the DOT was comprised of over 12,000 job titles. The DOT wasn't a content model *per se*, but rather an exhaustive list of job titles and descriptions. What it lacked was a language and a framework connecting the job titles to one another and to some sort of a low-dimensional representation of what each job *is*. It lacked a content model. The first attempt at developing one coincided with the 1990 revision of the DOT in which the small army of 'trained occupational analysts' that maintained and developed it were instructed to append to each job description a list of attributes such as aptitude and temperament requirements, vocational preparation, physical demands, and several other job 'descriptors.' These descriptors were an attempt to develop a cross-occupational

[14] For additional details refer to Tippins and Hilton (2010).

language that could be used to describe a wide variety of occupations, both in absolute terms and relative to one another. They were in essence a content model without a formally-defined taxonomy to house them nor a conceptual framework to restrict them in a manner analogous to an economics model.

In 1990 the U.S. secretary of labor convened the Advisory Board for the Dictionary of Occupational Titles (APDOT). The APDOT set forth a process to transform the 1990 DOT revision into a more formal content model. This led to the 1996 O*NET prototype, which was 'field tested,' revised further, and then published in 1998 as O*NET 98. Among many other important changes relative to the DOT, the number of occupational titles was reduced from 12,000 to 1,102. Alongside this consolidation was a increase in the complexity of the O*NET taxonomy and the number of characteristics that it contained, which now stands at 277. The O*NET was revised 13 times between 1998 and 2009, at which point it was reviewed by a National Academy of Sciences panel. The panel made a number of recommendations, many of which are work-in-progress. The most recent version of the O*NET is version 28.3.

B Additional Results

The following tables replicate the results in Table 4 and Table 5 in the case of a Cobb-Douglas specification for preferences.

Table 7: Stepwise Regression estimates of β_j^c in equation (6) between April 2019 and April 2020. Cobb-Douglas case. Estimates based on the 69 Products of the taxonomy. Column (1): backward selection with threshold significance level of 0.01. Column (2): backward selection with threshold significance level of 0.05.

	(1)		(2)	
ConsHuman	-0.386***	(0.0586)	-0.227**	(0.0690)
Conspic			-0.189**	(0.0653)
Durable			0.209***	(0.0562)
Charity			0.203***	(0.0550)
ConsFreq			0.157*	(0.0601)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The resulting correlation matrix across characteristics is displayed in Figure 8.

C BEA Product List

Table 8: Stepwise Regression Estimates of β_j^c in Equation (6) between April 2020 and April 2021. Cobb-Douglas case. Estimates Based on the 69 Products of the Taxonomy. Column (1): backward selection with threshold significance level of 0.01. Column (2): backward selection with threshold significance level of 0.05.

	(1)		(2)	
ConsHuman	0.244***	(0.000)	0.233***	(0.000)
Rival	0.140**	(0.003)	0.126**	(0.006)
Subsist			-0.107*	(0.018)

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Product List: Part I (Durables: 17 Categories).

BEA Classification	Product Name
Durable Goods	New and used autos and motorcycles
Durable Goods	Motor vehicle parts and accessories
Durable Goods	Furniture and household furnishings
Durable Goods	Household appliances
Durable Goods	Glassware, tableware, and household utensils
Durable Goods	Tools and equipment for home and garden
Durable Goods	Video and audio equipment
Durable Goods	Computers/tablets/peripherals, computer software, other
Durable Goods	Sporting equipment
Durable Goods	Bicylcles and bicycle accessories
Durable Goods	Pleasure boats, aircraft, and other recreational vehicles
Durable Goods	Recreational and Educational Books
Durable Goods	Musical instruments
Durable Goods	Jewelry and watches
Durable Goods	Personal medical equipment, eyeglasses
Durable Goods	Luggage
Durable Goods	Telephones and communications

Table 10: Product List: Part II (Nondurables: 9 Categories).

BEA Classification	Product Name
Nondurable Goods	Food and non-alc. beverages purch. from non-dining places
Nondurable Goods	Alcoholic beverages for home, tobacco
Nondurable Goods	Clothing and footwear
Nondurable Goods	Gasoline, lubricants, fuel and other energy products
Nondurable Goods	Pharmaceutical drugs (prescription and non-prescription)
Nondurable Goods	Recreational items
Nondurable Goods	Household supplies, disposable and somewhat durable
Nondurable Goods	Personal care products
Nondurable Goods	Magazines, newspapers, and stationary

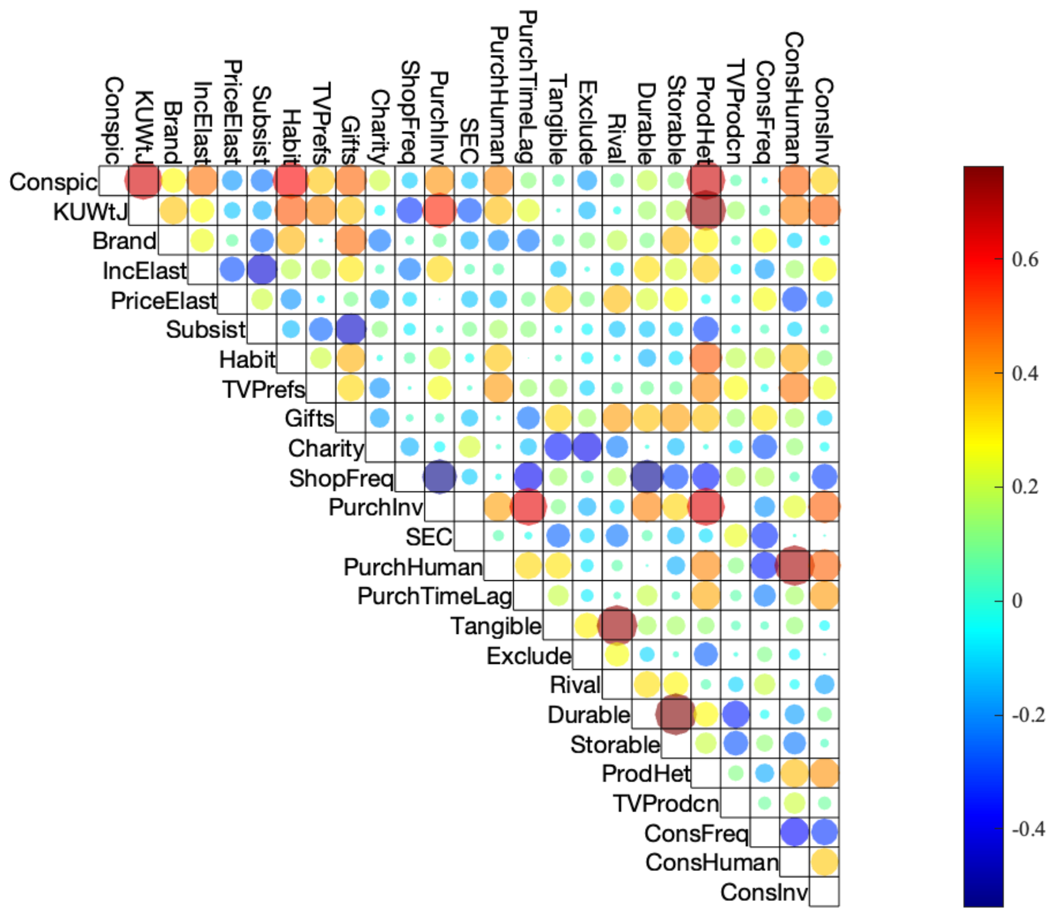

Figure 8: Cross Correlation of Characteristics in ALTZ content model.

Table 11: Product List: Part III (Services: 43 Categories).

BEA Classification	Product Name
Services	Housing Services
Services	Household utilities
Services	Outpatient Health Care Services
Services	Hospital Health Care Services
Services	Nursing homes
Services	Motor vehicle maintenance and repair
Services	Motor vehicle leasing and rental
Services	Parking fees and driving tolls
Services	Public land transportation
Services	Air transportation
Services	Water transportation
Services	Recreation services
Services	Television services
Services	Photo and audio visual services
Services	Video and audio streaming and rental
Services	Gambling
Services	Pet services
Services	Package tours
Services	School lunches and food supplied by government
Services	Meals at limited service eating places
Services	Meals at full-service restaurants, trains, movies, perf. arts, sports arenas
Services	Alcohol at Eating/Drinking Places
Services	Accomodations in hotels and motels
Services	Accomodations at schools
Services	Retail banking services
Services	Financial services other than retail banking
Services	Insurance
Services	Telecommunication services
Services	Postal and delivery services
Services	Internet access
Services	Educational services, higher education and vocational
Services	Educational services, elementary and secondary schools
Services	Legal services, accounting services, other business services
Services	Labor and Professional Organization Dues
Services	Funeral and burial services
Services	Personal care services
Services	Clothing and footwear services
Services	Child care
Services	Day care and nursery schools
Services	Social assistance
Services	Giving to social advocacy and religious institutions
Services	Household maintenance
Services	Foreign travel by US residents