

On the distribution of product price and quality

Alex Coad

Published online: 5 May 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract We investigate the structure of demand by focusing on the distribution of prices within narrowly-defined classes of goods. We observe considerable heterogeneity—products that are functionally similar but presumably of different ‘quality’ may sell at very different prices. We analyze distribution of prices for bottles of wine, used cars, houses in London and week-long holidays in Majorca, and observe in each case that the the resulting distribution is more skewed than the lognormal but less skewed than a Pareto distribution. We then present a theoretical model whereby products can distinguish themselves along multiple hedonic dimensions of ‘performance’, with these product attributes being random variables subject to multiplicative interactions. Variations of this model can reproduce a lognormal price distribution and a Pareto distribution as lower and upper bound benchmarks (respectively).

Keywords Price distribution · Quality · Consumption · Stochastic modelling · Pareto distribution

JEL Classification L15 · D12 · D11 · C16 · D31

1 Introduction

In our modern economy, household consumption patterns are characterized by demand heterogeneity. Consumers seek to distinguish themselves through their consumption activities. Firms respond to this heterogeneity by creating

A. Coad (✉)
Max Planck Institute of Economics, Evolutionary Economics Group,
Kahlaische Strasse 10, 07745 Jena, Germany
e-mail: coad@econ.mpg.de

niches, and the associated processes of variety generation and attempts at product differentiation are important drivers of industrial innovation. This research aims to improve our understanding of the structure and composition of consumer demand through an analysis of the heterogeneity of consumer products. More specifically, we focus on the price distribution, which presumably also sheds light on the underlying distribution of product quality. We collect data on prices of bottles of wine, prices of used automobiles, prices of houses in the London area and prices of week-long holiday packages in Majorca. These goods have similar ‘functions’, in the sense that cars transport people from A to B, wine is a beverage made from grapes with a loosely-specified alcoholic content, houses provide shelter, and the holidays we analyze offer one week of leisure time in a similar geographical environment. However, the prices of these goods vary dramatically, due to differences in quality, design, or even due to brand names and Veblenian ‘snob effects’.

Our analysis applies both non-parametric and parametric techniques. A non-parametric kernel density analysis allows a visual appreciation of the underlying distribution. This analysis offers little support to the view that consumer groups can be neatly separated into a small number of discrete strata. Instead, we observe a smooth distribution of prices, the smooth form indicating that there is a continuum of consumer groups and corresponding market niches, ranging from the very cheapest goods to the most extravagantly-priced luxury items. A parametric analysis provides more structure to our results, by comparing the empirical distributions to the log-normal and Pareto distributions. We observe that the price distribution is right-skewed, having fat tails that fall between the lower bound case of a lognormal and an upper bound case of a Pareto distribution.

In a theoretical model we build upon a standard hedonic model of product attributes to generate a right-skewed product quality distribution, which exhibits a fat tail that lies between the lognormal and Pareto benchmarks.

The layout of the paper is as follows. In Section 2 we discuss the theoretical background in which this analysis of consumption behaviour is framed. In Section 3 we present the databases along with some summary statistics. Section 4 contains our analysis of the price distributions, and in Section 5 we propose a statistical explanation of the observed distributions. We conclude in Section 6.

2 Framing the research question

In this section we argue that a skewed income distribution, as well as consumer heterogeneity, offers a basis for conjectures of a skewed distribution of product prices. Indeed, moving from poorer to richer individuals within a cross-section, we can expect that similar goods will sell at quite different prices as firms try to create niches and cater for heterogeneous consumers. It has long been known that the distribution of income is skewed to the right and has heavy tails—signifying that whilst most people may earn a modest income there are a few

individuals with an extremely high income (for a survey of research into the income distribution, see Kleiber and Kotz 2003). It follows that heterogeneous individuals can utilize their income for many different purposes. How do richer individuals go about spending their higher income? This is the context in which we place our investigation of the price distribution. Figure 1 shows four possible reactions to increases in income, which will be discussed in the following. A discussion of the four categories in Fig. 1 is helpful to situate the focus of our present investigation.

The first category, labelled A in Fig. 1, concerns changes in the propensity to save. Do richer people save more? The answer appears to be affirmative. It has long been known that, within a cross-section, individuals with higher income save more (see the survey in Dynan et al. 2004). In addition, evidence in Dynan et al. (2004) suggests that saving rates rise with lifetime ‘permanent’ income. Saving may be a residual category, in the sense that wealthy individuals may end up saving what they didn’t have the time or imagination to spend in the present period. It should also be recognized that some types of investment spending undertaken by individuals, such as the purchase of consumer durables, may share some features with saving behaviour.

The second category (B in Fig. 1) concerns the propensity to spend on services as income increases. Because income rises but the 24-hour-day time constraint is fixed, some wealthier individuals may prefer to ‘save time’ and pay other people to help with specific tasks. For example, poor people may have to cook for themselves whereas richer individuals may prefer to eat in restaurants where their food is prepared for them by others. Schettkat and Yocarini (2006)

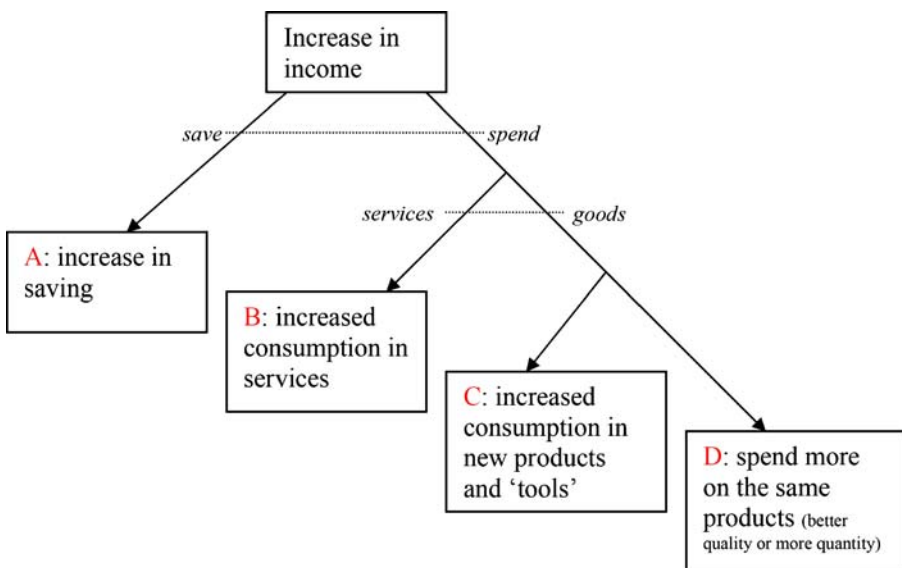


Fig. 1 Changes in behaviour following increases in income

present evidence from a number of countries suggesting that budget elasticities are lower than one for manufactured goods but above one for services (see in particular their Table 8). At the macroeconomic level, this qualitative shift in expenditure patterns would be visible even in aggregate statistics on industrial production, as consumption moves from industrial classifications based on manufactured goods to those related to services. Indeed, economists have observed a secular decrease in manufacturing and the rise of services industries in the composition of many economies.¹ As a result, it is meaningful to consider that increases in income will lead to increases in consumption in services.

Category C concerns the relationship between income and levels of consumption on other manufactured goods. A key component of this category relates to the development of individual consumption patterns along existing paths, or more qualitative shifts in consumption behaviour, made possible by the use of complementary ‘tools’, the services of which can assist an individual in the pursuit of increasing levels of ‘satisfaction’. Tools cannot be consumed themselves in the literal sense, but they are instrumental in the satisfaction of basic wants, yielding utility only indirectly through the services they provide. For example, a television set can address the want for cognitive arousal, or an elegant bow-tie may satisfy the want for social status. The increasing use of tools in consumption behaviour is discussed at length in Witt (2001), who goes as far as saying that “the demand for tools is the predominant feature of modern consumption expenditures” (pp. 27–28) and that “[t]he lion’s share of the long term growth of per capita consumption seems to be due to the increasing importance of these tools” (p. 33).

The fourth category (D in Fig. 1) concerns cases of ‘within-the-industry’ shifts in which richer individuals spend more on better quality² products (or perhaps a larger quantity of the same products) without necessarily changing the composition of their expenditure. (Note however that the distinction between the use of ‘tools’ and the choice of better quality products may be rather murky.) Whilst early contributions to consumption theory assumed goods to be of constant quality and sold in perfectly competitive markets, more recent research has made efforts to account for diversity of products within specific industries. For example, Caves and Porter (1977) suggest that firms within a given industry arrange themselves in different ‘strategic groups’ in order to fill different niches and to cater for different customers. Shaked and Sutton (1987) introduce the concept of ‘vertical differentiation’ to acknowledge the heterogeneity of consumer demand. Their concept of vertical differentiation contrasts with the case of ‘horizontal differentiation’ in the well-known model of Hotelling (1929). The authors define vertical differentiation in this way:

¹ Although the increasing outsourcing of contemporary economic activities and general measurement error may lead to an overstatement of the relative growth of the service sector, we should acknowledge that it cannot account for all of the growth.

² It could be that richer individuals pay higher prices for products of the same quality. This is the case, for example, if they cut down on ‘search costs’ associated with finding a bargain.

“given any two distinct products, if they were sold at the same price, then all consumers would choose the same one (the ‘higher quality’ product)” (p. 134).

One might begin to speculate to what extent the skewed price distribution emerges from the skewed shape of the income distribution. For example, is income dispersion a *sufficient* condition for price dispersion? Is it a *necessary* condition? These interesting questions cannot be addressed in this paper, although they would merit investigation in future work. In this paper we allow for price dispersion to stem from heterogeneous income as well as heterogeneous preferences within income groups, without distinguishing between these two effects.

Traditionally, the fourth category mentioned above has been empirically investigated using Engel curves (see *inter alia* Banks et al. 1997 and Chai and Moneta 2008). These curves focus on how expenditure share on a certain class of goods changes over the cross-sectional income distribution. Engel curves analyses are typically made at a relatively high degree of aggregation over industries, however, even though results obtained from Engel’s curve analyses are highly dependent upon the level of aggregation of the industries. Furthermore, there is implicit confusion about the degree to which consumption takes the form of better quality products or higher quantities consumed.

Pioneering work by Bils and Klenow (2001) investigates ‘quality Engel curves’ for 66 classes of durable goods bought by US households: their analysis decomposes the structure of expenditure patterns into quality effects and quantity effects. They observe that, in general, consumers respond to changes in income by focusing on higher quality products rather than seeking larger quantities of these products.

This paper seeks to complement the existing literature by focusing on cross-sectional heterogeneity among products in very narrowly-defined markets. The datasets we analyze are novel in that they have a relatively large amount of different products selling at very different prices. These datasets contain enough observations for a detailed parametric analysis of the price distributions (Section 4), and the results from this analysis in turn spur us on to propose a statistical explanation for the observed distribution (Section 5).

The stream of literature that appears to be the closest to our chosen topic is perhaps the literature on price dispersion. This body of work generally seeks to investigate departures from the ‘law of one price’ that one might expect to govern the sale of homogenous goods. While Engel curve analyses are typically undertaken at quite an aggregated level, the price dispersion literature is situated at such a disaggregated level that every attempt is made to obtain exactly identical commodities. Some early theoretical models sought to explain price dispersion in equilibrium by acknowledging that information is imperfect and that search is costly (Varian 1980; Salop and Stiglitz 1982). These models were also able to explain why some retail outlets might distinguish between their everyday prices and temporary discounts. Kirman and Vignes (1991) find a considerable degree of price dispersion in their empirical analysis of the peculiar case of the Marseilles fish market—a wholesale market for perishable goods where sellers deal with fishmongers or restaurant owners rather than

with individual consumers. Considerable price heterogeneity is observed in this market, largely because information on prices is imperfect (prices are not posted) and prices vary with the time of day. Other researchers have focused on the emergence of price dispersion in other markets relating to homogenous goods, although the degree of dispersion they find is typically far less than that reported by Kirman and Vignes. For instance, Clay et al. (2001) observe some systematic deviations from the ‘law of one price’ in their investigation of online book sales, where books can be seen as homogenous goods with unique ISBN identification numbers. Syverson (2007) also observes price dispersion in his analysis of the ready-mixed concrete industry. This paper distinguishes itself from the price dispersion literature because imperfect information is assumed to play only a minor role in explaining the tremendous heterogeneity that we observe between the prices of goods. We consider goods that have similar ‘functionalities’ although we allow for considerable heterogeneity in terms of the quality of the good. As such, we do not investigate differences in the prices of homogenous goods. Furthermore, in this paper we do not consider how prices may change over time.

3 Databases and summary statistics

We begin by describing how the datasets were collected before presenting some summary statistics.

3.1 Databases

In this section we present our four databases. Each dataset can be seen as a cross-section for the purposes of our statistical investigations. The prices listed are not wholesale prices, but they are the prices faced by individual consumers. Unlike the case of the fish market in Kirman and Vignes (1991), the prices are clearly communicated to consumers. With the exception of the house prices data, the prices are advertised to consumers on the same commercial platform, and so theories based on imperfect information and costly search are of little use in explaining the enormous heterogeneity we observe in the prices of comparable goods. In two of the cases (i.e. the data on wine bottles and holiday packages in Majorca) all of the products are sold by the same firm.

3.1.1 Wine data

The wine data come from the website www.foodandfinewine.com, the website of a British specialist food and wine supplier (prices expressed in GBP). This website contained the most comprehensive catalogue of wine prices I could find on the internet. The data were collected on 16th August 2007, and an unrestricted search for wine bottles returned 490 responses. 3 observations were removed because they corresponded to half-bottles or magnums. One bottle had no price available. We thus end up with 486 observations.

3.1.2 Used car data

Data on used cars were collected on the website www.autotrader.com on the 16th August 2007. This website is a platform allowing private individuals to advertise their used cars.

Our motivations for using used cars (as opposed to new cars) are that the data contains information not only on the prices of the cars but (like the data on house prices in Section 3.1.3) also some information on the quantity of cars available at each price. For example, whilst a wine supplier might sell different quantities of different wines (probably selling lots of cheaper wines and rarely selling great vintages), each individual selling a used car only sells one car at the advertised price. Our different datasets thus aim to complement each other by shedding light on different aspects of price distributions.

Our search restrictions were as follows—we selected used cars only, within 500 miles of US Zip code 10001 (which corresponds to Manhattan, New York, NY). We are interested in vehicles of all body styles, all makes and models, all prices, and of different ages.³ This returned 577,841 observations, which is too large a number to work with given that the data needs to be entered manually. We therefore limit ourselves to the upper tail of the distribution, which will allow us to test the hypotheses that the price distribution follows a Pareto distribution. We consider the top 1000 observations. However, in some of the cases the cars advertised on the website were obvious data entry mistakes.⁴ Such data entry mistakes due to which cars have implausibly large prices seemed to concentrate at the extreme upper tail of the distribution (as one might expect), and so they were not too difficult to identify and subsequently remove. We end up with 996 observations.

Our focus on the upper tail of the distribution implies that it would not be sensible to use this database to test for lognormality. Whereas a Pareto distribution is a ‘scale-free’ distribution that has an identical (unconditional) parametric characterization at any point of the distribution, this latter statistical property is not a feature of the log-normal distribution. As such, we do not test for lognormality using this database on used cars.

3.1.3 Real estate data

Our investigations also analyze the distribution of house prices. Similar to the data on used cars, data on house prices will also reflect both price and quantity consumed (i.e. each price corresponds to one house only). These data were collected from the website www.houseprices.co.uk, which is the UK Land Registry database of houses sold in England and Wales since 2000. The search was restricted to ‘London’, and over 1 million house prices were listed, sorted by date of sale. Data were taken on 30th August 2007, and data on the 500

³We restrict ourselves to the year range 1981–2008, which is the largest range available.

⁴For example, we excluded a 2004 Jeep Grand Cherokee being sold at \$1,699,500.

most recent transactions were taken. These transactions took place between the 15th and the 30th August 2007.

3.1.4 Holiday data

We continue our analysis with the dataset in Chai and Guerzoni (2007). This dataset contains prices of 1 week package holidays in Majorca organized by TUI over the period 1970–1998, as advertised in holiday brochures. Although different one-week holidays have a similar function (i.e. time spent in a leisurely environment), we can expect there to be a tremendous diversity in quality within such a class of holidays. The fact that these holidays can be seen as packages or ‘bundles’ is not inconsistent with the hedonic model developed later on (in Section 5).

Since we are not interested in a time series analysis in this particular paper, we restrict our attention to the 2 years where we have a relatively large number of observations (i.e. the years 1995 and 1998, where we have 1637 and 1131 observations respectively).

3.2 Summary statistics

Summary statistics are shown in Table 1. In all cases, we observe a positive skewness, a large kurtosis (relative to the Gaussian case), and a mean that is larger than the median. This indicates that the price distributions are right-skewed with heavy tails (or ‘leptokurtic’).

4 Analysis

4.1 Testing for lognormality

We begin with some kernel density plots of the price distributions in Figs. 2 and 3. (We do not test for lognormality in the used cars dataset for reasons mentioned in Section 3.1.2.) Casual observation of these distributions reveals that the distribution of log prices is not symmetric, but instead there appears to be a rather fat right tail. (The left tail seems especially underdeveloped for the wine data, perhaps because there are transportation costs that discourage the sale of very cheap wine.) Formal statistical tests for (log)normality are presented in Tables 2 and 3.⁵ We can convincingly reject lognormality in each case using both an ‘eyeball test’ and formal statistical tests. The distribution

⁵The Shapiro-Wilk W test for normality can be used with $7 \leq n \leq 2000$ observations, and the Shapiro-Francia W' test can be used with $5 \leq n \leq 5000$ observations. The skewness-kurtosis test presents a test for normality based on skewness and another based on kurtosis and then combines the two tests into an overall test statistic. This latter test requires a minimum of 8 observations to make its calculations. All tests are implemented using Stata 9.0.

Table 1 Summary statistics

	Mean	Std dev	Skew	Kurt.	1%	10%	25%	Median	75%	90%	99%	Obs
Wine	109.23	256.64	5.02	31.56	5.95	9.95	14.95	29.95	79	275	1500	489
Used cars	71152	21975	6.00	73.95	49995	53900	57925	65988	77993	95900	144500	996
Holidays 1995	120683	37789	2.68	13.77	73900	87400	98900	112900	129900	159400	288700	1637
Holidays 1998	138702	47691	2.18	10.03	79164	95330	108889	127037	152382	198483	327398	1131
Real estate	396332	396210	4.76	34.30	101250	172500	218550	285625	412500	695500	2436500	500

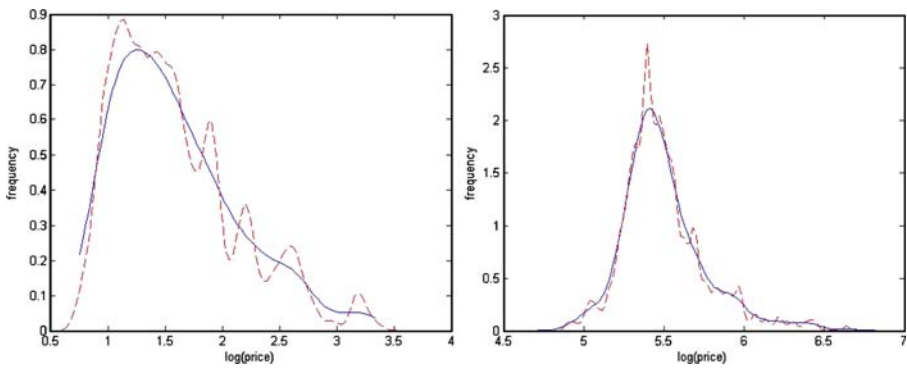


Fig. 2 Kernel densities of the distribution of prices using the wine data (*left*) and real estate data (*right*). Kernel densities obtained using the normal kernel function. The *smoother line* is the Matlab 7 default for estimating normal densities. The *dotted line* is obtained using a kernel bandwidth that is three times smaller than this default value

of prices appears to have a fatter right-tail than we would expect under a lognormal distribution.

4.2 Testing for a Pareto distribution

We investigate the possible existence of a Pareto structure in the price distributions by examining Zipf plots (see Figs. 4 and 5). Zipf plots have the log of rank on the y-axis, where the highest-priced product has rank one. If the series follow a Pareto distribution, we would expect a straight line on a Zipf plot (see Mitzenmacher 2003). Instead, the Zipf plots look slightly concave to the origin (except perhaps for the three very largest observations for the used cars data). It appears that the right-tails decay faster than a Zipf law would predict.

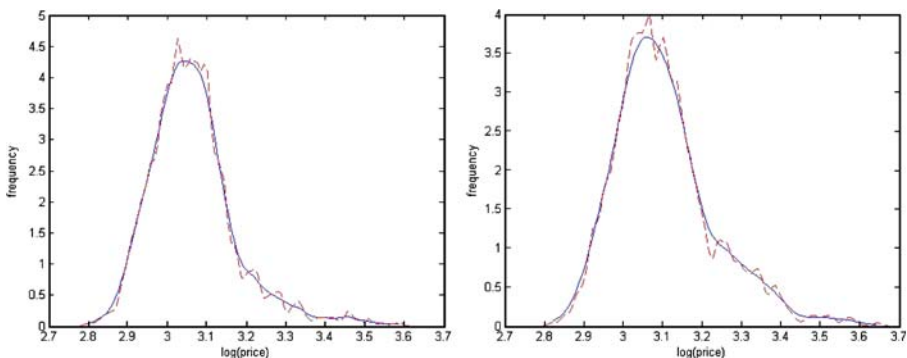


Fig. 3 Kernel densities of the distribution of prices using data for holidays in 1995 (*left*) and 1998 (*right*). Kernel densities obtained using the normal kernel function. The *smoother line* is the Matlab 7 default for estimating normal densities. The *dotted line* is obtained using a kernel bandwidth that is three times smaller than this default value

Table 2 Testing for lognormality of the price distributions

	W/W'	V/V'	Z	Prob > z	Obs
Shapiro-Wilk test for normality					
Wine	0.9339	21.807	7.402	0.000	489
Real estate	0.93358	22.343	7.468	0.000	500
Holidays 1995	0.93391	65.32	10.552	0.000	1637
Holidays 1998	0.95232	33.643	8.75	0.000	1131
Shapiro-Francia test for normality					
Wine	0.93575	22.487	6.536	0.00001	486
Real estate	0.93288	24.103	6.675	0.00001	500
Holidays 1995	0.93375	63.527	7.534	0.00001	1637
Holidays 1998	0.95232	35.251	7.137	0.00001	1131

We complement this ‘eyeball testing’ by looking for non-linear components in regressions. Whilst the ‘eyeball test’ is biased in favour of the observations with rank close to 1 (because observations with higher ranks are superimposed upon each other and less ‘visible’), our regression analysis gives equal weights to all observations. The first regression, shown in Eq. 1, includes a quadratic term for log price that will inform us if the Zipf plot is concave.

$$\text{Log}(\text{rank}) = \alpha_0 + \alpha_1 \text{Log}(\text{price}) + \alpha_2 \text{Log}(\text{price})^2 + \epsilon \tag{1}$$

The second regression is non-linear least squares estimation of the following equation:

$$\text{Log}(\text{rank}) = \beta_0 + \beta_1 \text{Log}(\text{price})^{\beta_2} + \epsilon \tag{2}$$

The regression results are presented in Tables 4 and 5. The regression results in Table 4 indicate that there is indeed a concave nature to the empirical Zipf plots. This is indicated by the negative and significant coefficients on the quadratic term in estimation of Eq. 1 in Table 4. Further evidence on the concavity is to be found in the interaction between the negative values for β_1 and the values of β_2 greater than unity in Table 5. Although the values for the R^2 coefficient are very high, note that this may be a spurious statistical result that stems from the fact that the y-axis of a Zipf plot is merely a transformation of the x-axis (see Gan et al. 2006). To sum up, the right tails of the empirical price distributions experience a faster decay than we would expect from a Pareto distribution.

Table 3 Skewness-kurtosis test for lognormality of the price distributions

	Pr (skewness)	Pr (kurtosis)	Joint	
			adj $\chi^2(2)$	prob > χ^2
Wine	0.0000	0.1000	43.43	0.0000
Real estate	0.000	0.000	.	0.0000
Holidays 1995	0.0000	0.0000	.	0.0000
Holidays 1998	0.0000	0.0000	.	0.0000

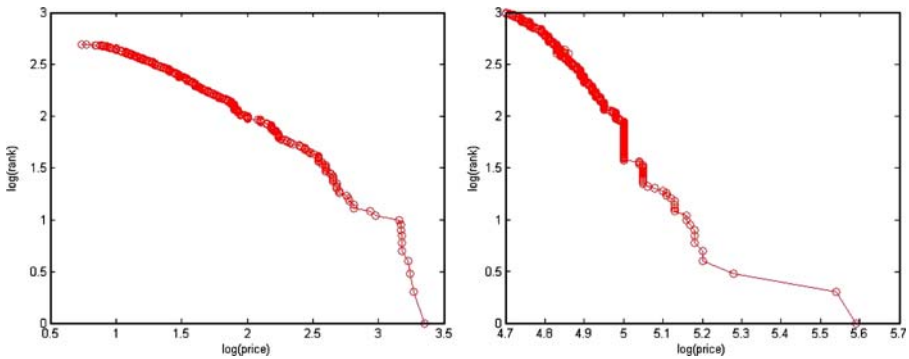


Fig. 4 Zipf plots of the distribution of prices using data for wine (*left*) and used cars (*right*)

5 Explaining the skewed nature of the price distribution

In some industries comprised of standardized goods (such as cement, one might think), there may be very little variation in quality or price. In contrast, our analysis has focused on some goods and services that display a wide leeway for differences in quality. There are indeed many dimensions in which these goods can distinguish themselves. Furthermore, it is reasonable to suppose that the overall quality of the product depends on the interactions of the product's performance in these individual dimensions. If a product scores highly (badly) in all dimensions, it will be a top (low) quality product. If, however, the product scores well in all dimensions except one in which it performs terribly, then the product will be attributed an overall low quality, and as a result will sell at a low price. This is the intuition behind the multiplicative specification in the following hedonic function—the multiplicative specification allows for a proportional effect of each attribute on the overall final quality (whereas a linear additive specification would not).

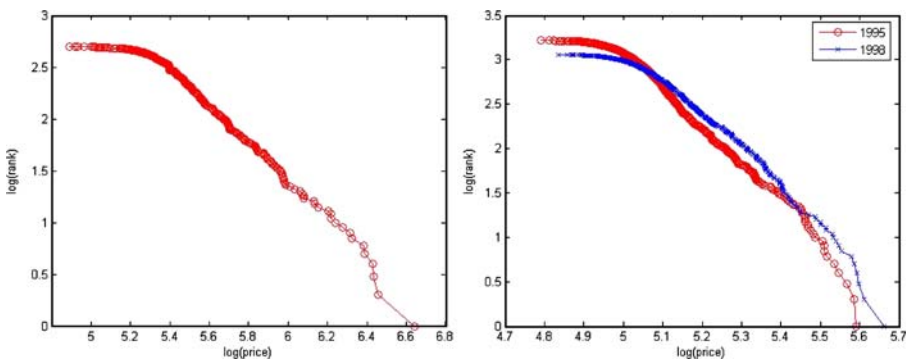


Fig. 5 Zipf plots of the distribution of prices using data for real estate prices (*left*) and holidays in 1995 and 1998 (*right*)

Table 4 Zipf plot regression results (OLS corrected for heteroskedasticity)

Estimation of Eq. 1								
	α_0	Std err	α_1	Std err	α_2	Std err	R^2	Obs
Wine	3.4656	0.0264	-0.7600	0.0182			0.9543	486
Used cars	22.8769	0.3920	-4.1974	0.0813			0.9589	996
Real estate	11.0853	0.1855	-1.6032	0.0335			0.9548	500
Holidays 1995	50.2811	0.4934	-3.7618	0.0422			0.9531	1637
Holidays 1998	45.1334	0.6213	-3.3152	0.0528			0.9506	1131
Wine	2.7988	0.0459	0.0507	0.0619	-0.2204	0.0191	0.9911	486
Used cars	10.1377	21.2979	0.9879	8.7544	-0.5273	0.8993	0.9597	996
Real estate	-8.6245	1.6393	5.4028	0.5811	-0.6209	0.0514	0.9827	500
Holidays 1995	-134.0757	9.1793	27.4646	1.5609	-1.3214	0.0663	0.9781	1637
Holidays 1998	-182.5090	6.0398	34.8774	1.0183	-1.6008	0.0429	0.9941	1131

To illustrate, we can consider that the difference between a low price holiday and a high price holiday is due to a wide range of factors such as size, comfort, luminosity, modernity, cleanliness and location of the hotel room. A very low score in any of these dimensions will have a considerable effect on the overall quality (and hence price) of the holiday package. Differences in the prices of cars can be due to the quality of the engine, the brakes, the steering, the electronics, or the general design. We can also see how an otherwise excellent car but with serious problems in any particular dimensions will not be successful on the market. Similarly, differences in the prices of wine may be due to interactions between the microclimate, humidity, mineral content of the soil, the processes of harvesting and vinification, how the wine was stored and even the quality of the cork. A low score in any one of these dimensions may be enough to completely ruin the wine. In the following model we propose an explanation of the observed price distribution that is based on a multidimensional vision of product quality, in which the overall quality depends on the multiplicative interaction of the scores in individual dimensions of product quality space.

Let us now introduce the model. There is a tradition in the literature on consumption to decompose a consumer's willingness to pay for a product into a 'hedonic function' such that the overall quality depends on a number N of

Table 5 Zipf plot regression results (non-linear least squares)

Estimation of Eq. 2								
	β_0	Std err	β_1	Std err	β_2	Std err	R^2	Obs
Wine	2.8146	0.0790	-0.1867	0.0054	2.0880	0.0246	0.9911	486
Used cars	12.1732	1.2728	-0.3620	0.1998	2.0791	0.2660	0.9597	996
Real estate	3.9390	0.0649	-0.0004	0.0001	4.8191	0.1602	0.9799	500
Holidays 1995	11.3367	0.1324	-0.0000	0.0000	8.1162	0.1896	0.9755	1637
Holidays 1998	9.1826	0.0474	-0.0000	0.0000	11.1177	0.1402	0.9911	1131

product characteristics (Lancaster 1966).⁶ Hedonic functions have in fact been estimated for all of the products examined here: wine (e.g. Combris et al. 1997), cars (e.g. Adelman and Griliches 1961), houses (e.g. Goodman and Thibodeau 1995) and holidays (e.g. Chai and Guerzoni 2007). Hedonic functions have also been estimated in the log-log form suggested here (see e.g. Goodman and Thibodeau 1995 and references therein). The hedonic function can be written in algebraic form as:

$$\text{Price} = f(x_1, x_2, \dots, x_N) \quad (3)$$

where x_i is a dimension of product quality in which products can distinguish themselves, where $i \in (1 \dots N)$. As discussed above, it seems appropriate to introduce a multiplicative functional form for the nature of the interactions of the performance in the individual product dimensions.

Suppose now that N is large and that the product characteristics are i.i.d. distributed (although we need not assume that the characteristics are normally distributed variables). The overall product price could be written as follows:

$$\text{Price} = x_1 \times x_2 \times \dots \times x_N = \prod_{i=1}^N x_i \quad (4)$$

Hence we obtain a lognormal distribution of prices:

$$\text{Log}(\text{Price}) = \log \left(\prod_{i=1}^N x_i \right) = (\log(x_1)) + (\log(x_2)) + \dots + (\log(x_N)) = \sum_{i=1}^N (\log(x_i)) \quad (5)$$

Since the $(\log(x_i))$ are i.i.d. distributed, central limit theorem implies that $\text{Log}(\text{Price})$ is normally distributed, and hence that Price is lognormally distributed.

Two modifications to this baseline model can be undertaken to explain the emergence of a Pareto distribution. A first candidate modification might be to allow for the possibility that the price distribution has a lower bound, below which prices cannot fall, such that the distribution can be modeled as a Kesten process (Kesten 1973, for applications see Gabaix 1999 and Axtell 2001). Such a model might correspond to the case where goods below a certain quality threshold are no longer sold or permitted for sale. This process has been shown to produce a Pareto distribution from a variety of possible distributions of x_i . A second possible modification would be to make the number of product attributes a random variable itself. This would correspond to the case where firms try to distinguish themselves by including additional features and characteristics to their products. For example, if N itself is exponentially distributed, the price distribution can be shown to follow a Pareto distribution (Huberman and Adamic 1999; Adamic and Huberman 1999).

⁶See also Saviotti and Metcalfe (1984) for a characteristics-based framework for describing technologies and technological change.

Our theoretical model is therefore able to reproduce both the lognormal and the Pareto distributions, which correspond to the lower bound and the upper bound (respectively) of the empirically-observed distribution. Our baseline model yields a lognormal, and some modifications to this baseline model are suggested that move the distribution away from this lower bound towards the Paretian upper bound. Our theoretical model is thus in line with the empirical regularities presented in Section 4.

6 Conclusion

This paper focuses on the cross-sectional structure of consumption behaviour within four particular industries—wine, used cars, houses in the London area, and 1 week holidays in Majorca. Empirical analyses reveal that the prices of these goods vary dramatically even within narrowly-defined markets, presumably due to differences in quality. A non-parametric kernel density analysis allows a visual appreciation of the underlying distribution whilst a parametric analysis (comparing the empirical distributions to the log-normal and Zipf distributions) provides more structure to our results. We observe that the price distribution is right-skewed, having fatter tails than the lognormal case but less fat than a Pareto distribution. In a theoretical model we build upon a standard hedonic model of product attributes and we introduce a multiplicative specification governing the interactions of the different dimensions of product quality. This generates a right-skewed product quality distribution, and variations of this model include the lognormal and Pareto benchmarks.

Acknowledgements Thanks go to Alex Frenzel-Baudisch, Guido Buenstorf, Marco Guerzoni, Corinna Manig, Alessandro Nuvolari, Pier Paolo Saviotti, Frank van Rijnsoever, Ulrich Witt, participants at the Max Planck Institute of Economics and the DIME-RAL2 WP 2.2 Conference on ‘Demand, Product Characteristics and Innovation’ held in Jena, October 18–19, 2007, as well as to the editors (Roberto Fontana and Alessandro Nuvolari) and two anonymous referees. Thanks are also due to Andreas Chai for providing the holiday data. Zlata Jakubovic provided excellent research assistance. The usual caveat applies.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Adamic LA, Huberman BA (1999) The nature of markets in the world wide web. SSRN working paper IEA5. doi:10.2139/ssrn.166108
- Adelman I, Griliches Z (1961) On an index of quality change. *J Am Stat Assoc* 56(295):535–548
- Axtell RL (2001) Zipf distribution of US firm sizes. *Science* 293:1818–1820
- Banks J, Blundell R, Lewbel A (1997) Quadratic Engel curves and consumer demand. *Rev Econ Stat* 79(4):527–539
- Bils M, Klenow PJ (2001) Quantifying quality growth. *Am Econ Rev* 91(4):1006–1030
- Caves RE, Porter ME (1977) From entry barriers to mobility barriers: conjectural decisions and the contrived deterrence to new competition. *Q J Econ* 91(2):241–262

- Chai A, Guerzoni M (2007) When differentiation does not pay: a note on the development of hotel characteristics in the Majorcan accommodation sector, 1971–1998. In: Paper presented at the DIME conference on 'demand, product characteristics and innovation, Jena, 2007
- Chai A, Moneta A (2008) Satiation, escaping satiation, and structural change: some evidence from the evolution of Engel curves. In: Papers on economics and evolution 08–18. Max Planck Institute of Economics, Evolutionary Economics Group
- Clay K, Krishnan R, Wolff E (2001) Prices and price dispersion on the web: evidence from the online book industry. *J Ind Econ* 49(4):521–539
- Combris P, Lecocq S, Visser M (1997) Estimation of a hedonic price equation for Bordeaux wine: does quality matter? *Econ J* 107:390–402
- Dynan KE, Skinner J, Zeldes SP (2004) Do the rich save more? *J Polit Econ* 112(2):397–444
- Gabaix X (1999) Zipf's law for cities: an explanation. *Q J Econ* 114:739–767
- Gan L, Li D, Song S (2006) Is the Zipf law spurious in explaining city-size distributions? *Econ Lett* 92:256–262
- Goodman A, Thibodeau T (1995) Age-related heteroskedasticity in hedonic house price equations. *J Hous Res* 6(1):25–42
- Hotelling H (1929) Stability in competition. *Econ J* 39:41–57
- Huberman BA, Adamic LA (1999) Growth dynamics of the world-wide web. *Nature* 401:131
- Kesten H (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math* 131(1):207–248
- Kirman AP, Vignes A (1991) Price dispersion. Theoretical considerations and empirical evidence from the marseilles fish market. In: Arrow KG (ed) *Issues in contemporary economics*, chapter 10. Macmillan, London, pp 160–185
- Kleiber C, Kotz S (2003) *Statistical size distributions in economics and actuarial sciences*. Wiley-Interscience, Hoboken
- Lancaster KJ (1966) A new approach to consumer theory. *J Polit Econ* 14:132–157
- Mitzenmacher M (2003) A brief history of generative models for power law and lognormal distributions. *Internet Math* 1(2):226–251
- Salop S, Stiglitz JE (1982) The theory of sales: a simple model of equilibrium price dispersion with identical agents. *Am Econ Rev* 72(5):1121–1130
- Saviotti PP, Metcalfe JS (1984) A theoretical approach to the construction of technological output indicators. *Res Policy* 13:141–151
- Schettkat R, Yocarini L (2006) The shift to services employment: a review of the literature. *Struct Chang Econ Dyn* 17:127–147
- Shaked A, Sutton J (1987) Product differentiation and industrial structure. *J Ind Econ* 36(2):131–146
- Syverson C (2007) Prices, spatial competition and heterogeneous producers: an empirical test. *J Ind Econ* 55(2):197–222
- Varian H (1980) A model of sales. *Am Econ Rev* 70(4):651–659
- Witt U (2001) Learning to consume—a theory of wants and the growth of demand. *J Evol Econ* 11:23–36