Do the Poor Pay More?
An Investigation of British Grocery Purchase Prices

Laura Blow
Andrew Leicester

Institute for Fiscal Studies

Copy-edited by Judith Payne

The Institute for Fiscal Studies
7 Ridgmount Street
London WC1E 7AE

Published by

The Institute for Fiscal Studies
7 Ridgmount Street
London WC1E 7AE
Tel: +44 (0)20 7291 4800
Fax: +44 (0)20 7323 4780
Email: mailbox@ifs.org.uk
Website: http://www.ifs.org.uk

ISBN: 978-1-903274-81-1

Preface

Funding from the Joseph Rowntree Foundation (JRF) under the 'New Insights and Innovation: Responsive Mode Funding' call is gratefully acknowledged, as is co-funding from the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS (RES-544-28-5001). The authors would like to thank Kantar Worldpanel for supplying the data used in this report, Walter Beckert, Mike Brewer, Ian Crawford, Tom Crossley, Rachel Griffith, Lars Nesheim and Ian Preston for comments and advice on earlier drafts, and participants at conferences at the Institute for Fiscal Studies and Hitotsubashi University, Tokyo, for helpful thoughts. All errors are the responsibility of the authors. Data from the 2007 Expenditure and Food Survey were collected by the Office for National Statistics and distributed by the Economic and Social Data Service. Crown Copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

Address for correspondence: Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE.
Telephone: +44 (0)20 7291 4800.
Email: laura_b@ifs.org.uk, andrew_l@ifs.org.uk.

Contents

Executive Summary

Introduction

- Food prices can have a significant effect on household welfare. By 2007, households spent on average less than 18 per cent of their budget on food prepared at home, but vulnerable groups spent more: 21 per cent for lone parents, 22 per cent for pensioners and 26 per cent for those aged 80 and over.

- This study is the first large-scale analysis for Great Britain of whether there is systematic variation in the prices paid for food across household groups, and in particular whether low-income households pay higher prices than richer households. The theoretical relationship between prices paid and income is unclear, which makes an empirical analysis essential.

- Relative to previous studies, our analysis uses extraordinarily disaggregate data, where goods are defined at the bar-code level, which allows us to compare prices paid for *identical* products. This means we do not confuse price variation with variation in product quality, which is a danger when relying on more aggregated expenditure data. We also explore whether different methods of constructing the price indices required to compare prices across household groups give a consistent story as to who pays more.

- Although it is essential not to group products that vary in quality, defining products as finely as at the bar-code level may impose too much disaggregation. First, it ignores an important way in which households may be able to economise without necessarily affecting product quality, which is through buying larger package sizes and so paying a lower price per kilo or litre. Second, whilst branded products are almost certainly distinct goods, supermarket own-brand versions of these goods might usefully be grouped together to increase the overlap of the products we are able to compare. We explore how our results change when we modify our analysis along these lines.

Previous literature

- Based on a number of small, localised studies, policy in the mid-1990s was concerned with the notion of 'food poverty' – the idea that some households may forgo meals to save money – which could be partly driven by 'food deserts': areas (particularly urban areas) where cheap, healthy food was inaccessible to those without private transport.

- However, more recent studies both of the US and of the UK found little evidence that the urban poor had less access to fresh foods or faced a less competitive, higher-priced retail environment. In terms of food prices, these studies tended to conclude that prices vary across store type (multiples, local stores, discounters and so on, which may be more or less likely to locate in rich and poor areas) but that for a given store type and location there was no evidence that prices varied between low-income and high-income areas.

- These studies tended to compare the price of a common 'basket' of goods, which may not reflect the goods bought by rich or poor, across stores and locations. Other papers have examined household-level data to explore the relationship between food prices and income. Two studies using US market research data to compare prices paid for identical products both found that richer households paid higher prices. Typically the differences were quite small but statistically significant, with the richest paying 2–3 per cent more than the poorest.

Theoretical background

- Prices could differ across consumers if sellers engage in 'price discrimination', selling at a higher price to groups of consumers who are less price-sensitive. In terms of groceries, this may be geographical (different prices in different areas) or take indirect forms such as varying prices according to package size, offering targeted discounts and operating multiple formats such as 'metro' stores which charge higher prices in inner-city areas.

- If finding low prices requires consumers to engage in costly search, economic theory shows that in equilibrium it can be optimal for different stores to charge different prices for the same good. The costs and benefits of search vary across households. It is not clear whether rich households would be expected to search more, and pay less, than poor households: richer households have higher consumption and so stand to make higher absolute savings from search, but their opportunity costs of time are likely to be higher than poorer households'.

- Variation in prices paid can arise because of temporary special offers. Consumers able to buy during the discount period could stockpile for future consumption and so pay less than consumers who cannot stock up. The willingness to stockpile will depend on storage and transport costs and the size of the discount. Again, it is not clear whether poorer households would be expected to stockpile more: they may face higher storage costs and may not be able to pay the up-front costs required to buy sufficient volume to stockpile.

- Prices could vary for identical products because of cost differences across stores and areas (for example, higher store rental prices in urban areas or the inability of small local stores to exploit economies of scale). The same good sold in two different stores could also be differentiated by the retail environment in which it is sold, with 'high-quality' stores charging more.

Data and methodology

- We use data from the Worldpanel of market research firm Kantar. The data set is a representative panel of British households' grocery purchases, recorded using an in-home bar-code scanner. Details of what was bought, the store, the price paid and household demographic information are recorded. Our main results focus on purchases during 2006.

- The data are extremely detailed. Our main analysis covers almost 18,000 households, typically making a million purchases per week in total. Purchases are recorded at the bar-code level so that we know we are comparing completely identical products. Purchases of non-bar-coded products are also recorded.

- We construct four different price indices and explore whether they tell the same story as to which groups pay higher prices. Three of these indices – Laspeyres, EKS and CPD – give one price index per demographic group (income group, age group and so on). A final index – the Paasche or 'own-basket' index – gives one index per household.

- To explore other ways in which consumers can economise on the prices they pay, we use the data to redefine 'identical product' and assess the impact on our price indices. First, we group together all goods that are the same other than the package size, to examine bulk discounts. Second, we group together all store-own-brand versions of the same product.

Results

- Based on data from June 2006, we find that different price indices tell largely the same story as to which groups face higher prices. On average, richer households pay more than poorer households but the price differences across income groups are small – typically only 1–2 per cent between the highest and lowest. All indices show that the richest group, with equivalised incomes above £60,000 per year, pay the highest prices. Those with incomes below £10,000 pay slightly more than those with incomes between £10,000 and £20,000.

- Our results are not sensitive to the particular time period chosen. Taking average monthly prices over calendar year 2006, we find those with incomes below £10,000 pay on average 0.9 per cent less than those with incomes over £60,000.

- Whilst average indices for different income groups are similar, there is variation across households in the own-basket index. Households at the 90th percentile of the price distribution pay around 8 per cent more than those at the 10th percentile, but this variation is not closely related to income.

- There is as much variation in the price index over time within a household as there is in the price index across households. Across-household variation is larger for poorer groups. Thus our results suggest that poor households pay slightly lower prices but with more variation around the average.

- Having children reduces the prices paid for food: those with three or more children pay 0.7 per cent less on average than those without children. Household size in general appears to be negatively related to prices: households with three or more adults also pay less on average than single people.

- Prices do not vary much with age, except at older ages. Compared with those aged 25–29, those aged over 80 pay around 1 per cent more on average.

- Using data on local area deprivation, we find both that richer households pay more than poorer households even conditional on living in similarly deprived areas, and that for households with similar incomes, those living in more deprived areas pay less. The richest households in the least deprived areas pay on average 1.4 per cent more than the poorest households in the most deprived areas.

- Changing the definition of a product to explore the impact of bulk discounting and buying cheaper own brands tends to reinforce rather than offset the differences across household groups. At the median, the effect of bulk discounts is to reduce the price index for the poorest households by 0.2 percentage points but to increase that for the richest households by a similar amount. The effect of cheaper own brands is larger, with a fall of 0.5 percentage points at the median for the poorest and a rise of 0.6 percentage points for the richest.

Explaining the variation in prices

Spatial search

- Households with access to a car make fewer main shopping trips per month (6.6 on average) than those without (9.4); the latter are concentrated amongst poorer households. Although those without cars make more trips, they pay, on average, slightly higher prices, so number of trips alone is not a good measure of a household's search intensity.

- Controlling for access to a car, we find that from a baseline of 10 trips per month, making one additional trip reduces prices paid by around 0.05 per cent, which is a small but statistically significant decrease. This holds within similarly densely populated areas, suggesting that the result is not just driven by rural households making more trips and facing higher prices because of less local competition.

Temporal search

- Conditional on the average number of trips made per store, increasing the number of stores visited by one reduces the price index by around 0.1–0.3 per cent. Conditional on the number of stores visited, making additional trips *increases* the price index: an extra two trips per store increases it by around 0.15 per cent.

- This is consistent with stockpiling: shopping more frequently implies buying less at each trip and taking less advantage of temporary low prices. Once we control for how much households use sales, this result disappears – making more trips per store now has no effect on prices, but visiting more stores still reduces prices, suggesting search is not simply about finding temporary special offers.

- Households that make greater use of sales also benefit most from bulk discounting. However, there is no relationship between savings from bulk buying and the number of stores visited – households do not appear to 'search' for bulk discounts.

Store choice

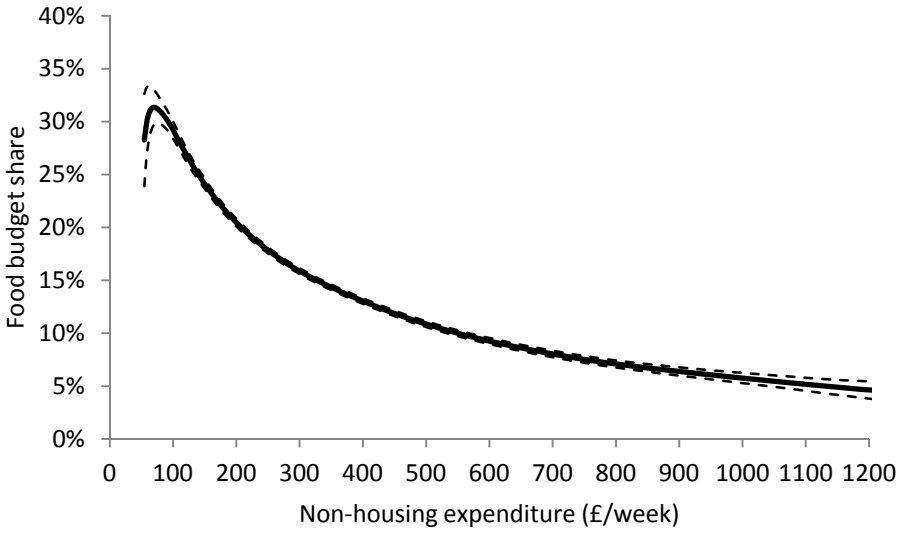
- Allowing store own brands to be grouped together as identical products, we find evidence of considerable variation in prices across store types, with 'quality' retailers and local shops charging substantially more for identical products than large retailers and discounters. But even with this product definition, there is a relatively small overlap of products bought across all store types.
- This suggests that some price variation could be driven by store choice, but the evidence is mixed: we find that richer households (which pay more on average) shop more in quality stores and expensive supermarkets and less in discounters; however, poorer households (which pay less) shop more in local stores where prices are high. Controlling for store choice reduces but does not eliminate the difference in prices across income groups.

CHAPTER 1
Introduction

The key objective of this study is to assess whether poor households pay systematically higher prices than other households for identical food products. A growing body of research has shown that in different markets, some households face higher prices than others. A survey by the National Consumer Council (2004) suggested that poor households may pay higher prices for energy, telecommunications and financial services such as banking and credit. However, food prices have not been studied in detail in the UK. Much of the existing evidence has relied on localised, small-scale studies (see Chapter 2 for a summary).

Food purchased for preparing and eating at home is a classic economic 'necessity', representing a larger part of the budget of poorer households. Using data from the 2007 Expenditure and Food Survey, Figure 1.1 shows the relationship between total expenditure (excluding housing costs) and the share of spending devoted to food at home.[1] Except for the very lowest spenders (where the estimates are much less precise), there is a clear negative relationship: those spending around £100 per week in total devote 30 per cent or so of their budget to food, falling to 5 per cent or less for those spending more than £1,000 per week.

Figure 1.1. Relationship between total spending and share spent on food, 2007

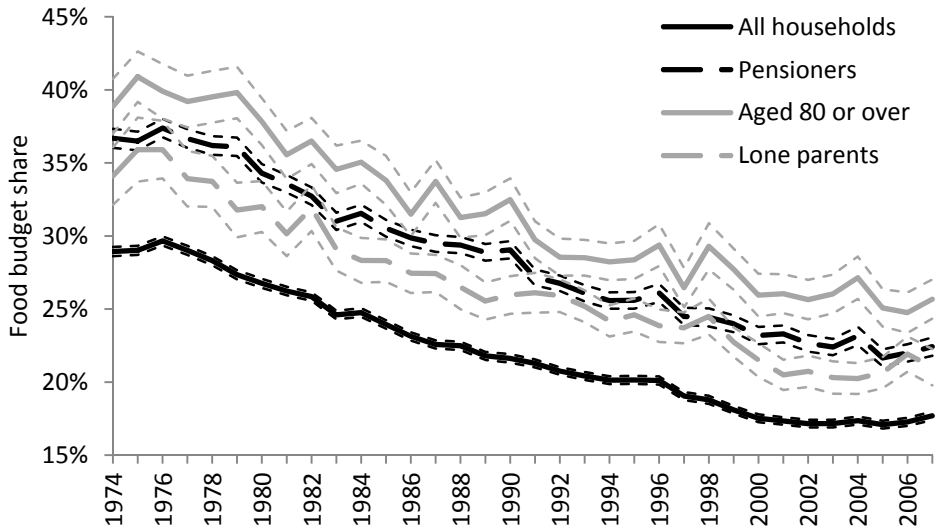


Notes: The relationship is estimated by a local linear regression of the food share on the log of the total budget, equivalised by the modified OECD equivalence scale; total spending is then re-expressed in cash terms. Dotted lines show 95 per cent confidence intervals; this means that the true value of the food share lies within these bounds 95 per cent of the time. The highest- and lowest-spending 1 per cent of households are excluded.
Source: Calculated from the 2007 Expenditure and Food Survey.

[1] We will simply call this food from now on.

Over time, food has become a less important part of the budget. This is partly because average incomes have risen, but also because average food prices (as measured by the Office for National Statistics when calculating national inflation figures) have fallen relative to economy-wide average prices. Food, though, still represents a large component of spending for many poorer households. Figure 1.2 shows average food budget shares over time for different household groups. Across all households, the food share has fallen from around 30 per cent in the mid-1970s to less than 18 per cent by 2007, though the budget share appears to have levelled off since around 2000. In 2007, the average food share for lone parents was 21.0 per cent, for households headed by a pensioner it was 22.4 per cent and for households headed by someone aged 80 or over it was still more than a quarter of their total budget, at 25.7 per cent.

Figure 1.2. Food budget shares over time, by household group



Notes: Dotted lines represent 95 per cent confidence intervals. Households spending more than 75 per cent of their budget on food are excluded.
Source: Calculated from the Family Expenditure Survey / Expenditure and Food Survey, 1974–2007.

Food prices could therefore have a significant impact on the welfare of poorer households in particular, and even more so if it turned out that poorer households pay higher prices than richer households for identical goods.

Using data on food and grocery purchases collected by market researchers, this study will assess in detail whether there is in fact any difference between households in the price they pay for food. If poor households do pay more, their real purchasing power and living standards will be lower than is implied by national price data. Using the data, we can also examine possible reasons as to *why* prices differ; we explore some of the theoretical background for differing prices in Chapter 3 and offer empirical evidence in Chapter 6.

This study represents the first large-scale analysis using British data of the question of whether and why there is variation in prices paid for food consumed at home. Our key questions are:

- Do different methodologies for comparing the prices paid by different households for identical products give the same results as to which household groups pay more?

- What are the key characteristics associated with paying higher prices, and do poor or vulnerable households pay more?
- What evidence can we provide on different theories as to why prices paid vary?

One of the unique aspects of our work is the data, which we describe fully in Section 4.1. We use a household scanner data set recording the food and grocery items purchased by a large, representative panel of British households. The data include precise details of the items that were purchased and their characteristics, the prices paid, the stores of purchase and demographic information about the households. Compared with earlier studies, we are therefore able to control precisely for product bought, and compare prices paid for identical items by different households. This avoids the problem of confounding variation in the price of identical products with price variation driven by quality differences in the purchases of different households, which is common to work that uses aggregate expenditure data. We are also able to use the data to redefine the idea of a 'product'. For example, we can group together items that differ only in their package size in order to assess how important bulk discounting is in affecting the prices that different households pay for effectively the same good. Unlike US studies that exploit similar data, we also make use of a range of methods to assess how prices vary across households, which allows us to be confident that our results are not driven by a particular methodological choice.

The rest of the study is organised as follows. Chapter 2 describes the key literature and empirical evidence on variation in food prices across households, and provides some background in terms of the evidence from studies of the food retail sector by government competition authorities. Chapter 3 outlines the theoretical economic background of price variation, the extent to which the theory suggests the poor would pay more, and ways in which we can assess the competing theories using our data. Chapter 4 describes our data and methodological approach in detail. Chapter 5 reports the main results and Chapter 6 examines different potential explanations for our findings. Chapter 7 concludes.

CHAPTER 2
Literature

2.1 Previous studies of food prices across households

In a 1995 report by the Low Income Project Team of the Nutrition Task Force (Beaumont et al.), 'food deserts' were defined as poor, urban areas where it was difficult to buy healthy food without access to private transport. A number of small-scale studies assessed how prices varied across rich and poor areas by pricing a fixed 'basket' of foods in different stores and different areas. Sooman, Macintyre and Anderson (1993) carried out a small exploratory review of the price of baskets of healthy and of unhealthy foods in two areas of Glasgow and found the healthy basket cost more in the poorer area. Mooney (1990) looked at nine stores in affluent and deprived areas of Hampstead. He found that both 'healthy' and 'unhealthy' baskets were cheaper in the deprived area, but that the relative cost of the healthy basket was higher in the poor area. Piachaud and Webb (1996) did not compare rich and poor areas but rather looked across six different shop types in Northampton and compared a corner shop and a supermarket in five locations across Britain. They found that small shops offer a narrower range and higher prices than larger shops.

Partly based on these findings, a UK Department of Health inquiry (Acheson, 1998) called for policies to reduce 'food poverty' – the idea that some poor households may forgo food because of a lack of money – and suggested that retail provision in different areas could be a contributory factor:

> Economies of scale allow food sold in supermarkets to be cheaper and to cover a wider range than that in smaller 'high street' stores … The increasing tendency to out of town supermarkets has led to the creation of 'food deserts' where cheap and varied food is only accessible to those who have private transport or are able to pay the costs of public transport, if this is available.

However, the Competition Commission (2000) found no evidence that major supermarkets were systematically avoiding low-income urban areas, or that these areas had less competition in grocery retail or consistently higher prices than other areas.[2] Based on a study of different areas in Scotland, Smith et al. (2010) also concluded that there was little evidence that deprived urban areas had poorer access to grocery stores or fresh fruit and vegetables. A recent large-scale American study (US Department of Agriculture, 2009) found that around 41 per cent of Americans lived more than a mile from the nearest supermarket, but that less than 2 per cent of Americans lived more than a mile away and had no access to a car.

[2] See, in particular, paragraphs 13.108–13.136 of Competition Commission (2000).

Some larger-scale, more systematic attempts to look at price differences in rich and poor areas have been made. Cummins and Macintyre (2002) looked at a random sample of 325 food retailers in the Glasgow area in 1997 and collected data on the price and availability of 57 common food items. They found that shop type (multiple, discounter, independent and so on) was the most important factor associated with price and availability, and that cheaper, wider-range stores were more often located in poorer than richer areas. Conditional on shop type, local deprivation levels had very little effect on price or availability. Using US data, MacDonald and Nelson (1991) examined the price of a basket of foods typically bought by low-income households in a random sample of stores in 10 large metropolitan areas. They found prices were, on average, around 2 per cent higher in stores in poorer areas, and inner-city prices around 4 per cent higher than suburban prices. However, prices were no higher in inner-city poverty areas than inner-city non-poverty areas, suggesting that location rather than local income matters most for prices. Chung and Myers (1999) used a sample of more than 50 stores in two counties in an urban area of Minnesota to examine variation in price for the most popular brands of 49 food and grocery items. They found prices were lower in chain stores than other stores, but no significant difference in overall prices between poor and non-poor areas even before taking store type into account. These studies therefore suggest that local area deprivation alone has little independent effect on prices. A similar conclusion was reached in a review of the US literature at the time by Kaufman et al. (1997). Their key findings were that supermarkets charged lower prices than small stores (of the order of 10 per cent or so on average) and that small stores were more often located in low-income inner-city areas and in rural areas whilst supermarkets were more often located in suburbs. They also suggested that inner-city supermarkets charged higher prices than suburban supermarkets, which they argued could reflect both cost and local competition differences. However, they found no evidence that low-income areas faced higher prices conditional on the store type and location.

A key problem with these studies is that they are generally based on calculating the price of a fixed basket of products in different areas or different stores. This basket may be constructed on the basis of nutritional need or on 'typical' shopping patterns, say, but it is unlikely to reflect the products that any individual household actually buys and so would not answer the question of whether poorer households pay more for the products that they actually purchase.

Indeed, Kaufman et al. (1997) found evidence of very different food spending habits in the US across income groups based on food consumption survey data. Across 65 food categories, using data from the late 1970s and late 1980s, they suggested that low-income households have unit costs (dollars per pound of produce) that are typically 10–20 per cent lower than average. This, of course, need not imply that poorer households pay less for identical products. Instead, it at least partly reflects a different mix of products within each relatively broad food category. For example, poorer households may buy lower-quality products (minced beef versus fresh steak, 'value' brands versus 'premium' brands) that lead them to pay lower unit prices. In the UK, Crawford (2003) used more disaggregate data from the 2000 National Food Survey to look at variation in unit prices for almost 250 food groups. He found that almost two-thirds of the foods have a statistically significant positive correlation between income and the unit price, and less

than 1 per cent have a significant negative correlation. Again, the large amount of aggregation within groups in the National Food Survey means there will be a great deal of variation in quality within each food group as well as variation in price.

A second problem of many earlier studies is that they are based on comparisons across rich and poor *areas* rather than households. It is likely, particularly when areas are defined very narrowly, that area-level measures of deprivation will represent the circumstances of the majority of households that live in them, but even within relatively small geographic areas there may be very localised pockets of richer and poorer households, which will not be picked up by this kind of analysis.

More recent studies have used household-level data from in-home bar-code scanners to look at the differences in prices paid, as is our approach. These studies are able to analyse prices at the individual product level, rather than relying on aggregated expenditure categories, which eliminates the potential for quality variation influencing the price paid. The paper closest to our general approach is Aguiar and Hurst (2007), who used US data from AC Nielsen Homescan covering purchases of packaged groceries[3] by over 2,000 households in the Denver area between January 1993 and March 1995. Similar to Crawford (2003), they calculated the ratio of a household's spending over a year to how much that household would have spent on its own basket of groceries had it paid average prices for it. They found that, conditional on 'shopping needs' (defined as the number of different products and product categories purchased and total expenditure), households earning more than $70,000 per year pay around 2.1 per cent more than those earning less than $30,000 per year, a relatively small but highly statistically significant difference. They also found that older households pay significantly lower prices than middle-aged households, and attributed this to older households making much more use of coupons and promotions, and shopping more intensively (making more trips to a given store, though they did not find older households shopping in a larger number of different stores). This, they argued, reflects the lower opportunity cost of time for older households. They concluded that doubling the shopping frequency reduces the price of the basket by around 7–10 per cent.

A recent paper by Broda, Leibtag and Weinstein (2009) also used AC Nielsen Homescan data, this time covering almost 40,000 households from across the US in 2005.[4] Their methodological approach was different from that of Aguiar and Hurst (2007), and used linear regression of the product price on a number of covariates including dummy

[3] Their data excluded fresh items such as fruit, vegetables and meat, all of which are observed in the data we use in our study. Aguiar and Hurst found their data capture only around 20 per cent of the spending on food at home as reported in the contemporaneous US Panel Study of Income Dynamics (PSID). By contrast, Leicester and Oldfield (2009a), in a study of the UK data, found that around 75 per cent of the spending on food at home in the contemporaneous Expenditure and Food Survey is captured by the scanner data used in this work.

[4] The paper highlighted two possible concerns in using the Nielsen data for this analysis. First, prices in chain stores are taken as the average weekly price for that product in that chain, which reduces the variation in the prices households are observed to pay (for example, if some households use loyalty cards or promotions to obtain different prices for the same good within a particular chain/week observation; for more, see Einav, Leibtag and Nevo (2008)). This is not true of our data, where prices are recorded from till receipts and so should accurately reflect the actual price paid. Second, the Nielsen data appear to under-represent the very poorest parts of the income distribution, where problems of access to low-price supermarkets may be most acute. Leicester and Oldfield (2009b) suggested that, if anything, the UK data used here tend to under-represent the very *richest* households.

variables for each product and family income.[5] In common with almost all previous studies, they found higher-income households pay more: those earning more than $100,000 per year pay around 3 per cent more than those earning less than $5,000 per year. However, they found that those earning $10,000–$30,000 pay around 1 per cent less than those earning less than $5,000. Their main result that the rich pay more was robust to including store controls and controls for local area in their specification.

The studies highlighted in this section range from small-scale, exploratory studies of a small number of stores in one location to large-scale, national studies using data on tens of thousands of households. Some look at variation in prices across stores, others across local areas and others across households. Typically, the results from them are qualitatively similar. There appears to be relatively little evidence that the poor 'pay more' for food and grocery products – if anything, richer households seem to pay higher prices, though the very poorest may pay slightly more than those a little further up the income distribution. Prices may well be higher in poor areas, reflecting the fact that poor households are more likely to locate in inner cities and very rural areas, but there is little evidence that income is related to price conditional on local area.

2.2 Regulatory background

The 2000 investigation of the UK supermarket sector by the Competition Commission (CC) was the first major investigation of the UK supermarket sector. At that point, the market share of the five largest supermarkets (Tesco, Sainbury's, Asda, Safeway and Morrisons) was around 75 per cent, with Tesco alone having almost 25 per cent of the market. In respect of pricing practices, the main focuses of the investigation were below-cost selling and 'price-flexing'. Price-flexing was defined as 'the practice of setting retail prices across different stores in different geographical areas in the light of local competitive conditions, such variation not being related to costs'.[6]

The CC concluded that all of the largest supermarkets and some other supermarkets priced below cost on some products, and that in some cases this could harm smaller retail outlets, leading to adverse effects on competition. It also concluded that there was some evidence of geographical price variations that were not cost related but pursued in response to local competitive conditions. Specifically, it found that seven retailers[7] engaged in price-flexing and that when the practice was carried out by stores with market power – Safeway, Sainsbury's and Tesco – it operated against the public interest because customers tended to pay more at stores that did not face particular competitors than they would if those competitors were present in the area.

Since the CC's 2000 investigation, the geographical pricing strategies of the main supermarkets have changed somewhat, even though the CC did not recommend any

[5] This approach is similar to the Country–Product Dummy method we use in this study, though Broda et al. did not weight their regression. See Appendix A for details of the CPD method.

[6] This is geographical price discrimination under another name (see Chapter 3).

[7] Budgens, the Co-ops, Netto, Safeway, Sainsbury's, Somerfield and Tesco.

remedial action against price-flexing. Tesco and Sainsbury's have moved voluntarily to a uniform national pricing policy, and when Morrisons acquired Safeway in 2004 these stores were converted to the Morrisons national pricing format.

However, in 2006, the Office of Fair Trading (OFT) again referred the grocery market to the CC, and another investigation was initiated. Part of the reason for the referral was continued evidence of price-flexing and concern over below-cost selling. In the final report to this second investigation (Competition Commission, 2008) and in the accompanying Pricing Practices Working Paper,[8] the CC concluded that, although most large grocery retailers set prices centrally and uniformly, or nearly uniformly, for all their stores, there was evidence that they vary some prices according to local competition, although not extensively. This is achieved mainly by the use of local vouchering. Two retailers – the Co-ops and Somerfield – continue to employ localised pricing practices that allow the level of prices at each store to respond to the degree of local competition.

With regard to below-cost sales, the CC found that the main foods sold below cost are alcohol, tinned and packet goods, non-alcoholic beverages and confectionery. It concluded, though, that 'the below-cost selling engaged in by grocery retailers could not be characterized as a broad-based predatory strategy aimed at operators of independent non-affiliated or symbol group convenience stores or specialist grocery retailers' and that 'we find that below-cost selling is unlikely to mislead consumers in relation to the overall cost of shopping at a particular grocery store'.

The CC also looked at store accessibility as part of an analysis of market concentration. Its research indicated that 95 per cent of the UK population living in urban areas have access to at least one grocery store larger than 1,400 square metres within a 10-minute drive-time, with 20 per cent having a choice of at least four such stores within a 10-minute drive. In rural areas, 71 per cent of the population have access to at least one grocery store larger than 1,400 square metres within a 15-minute drive and 13 per cent of the population have a choice of at least four such stores. This, though, does not address the concerns of the 'food desert' literature cited in Section 2.1, which focuses on those without access to private transport.

[8] http://www.competition-commission.org.uk/inquiries/ref2006/grocery/pdf/emerging_thinking_pricing_practices.pdf.

CHAPTER 3
Theoretical Background: Why Could Prices Vary?

Some simple theoretical models of competition predict that the price charged by all sellers for an identical product will be the same. Models of Bertrand competition (in which sellers compete on prices) with identical products and perfect information predict that even when there are only two sellers, both will charge a price equal to the marginal cost of production. As the product is the same, consumers will purchase from whichever seller charges the lower price. The equilibrium outcome of the price-setting game is that each seller continues to just undercut the other until both price at marginal cost.

The crucial theoretical question, then, is why variation in price for identical products is actually observed. The answer lies in the ability of retailers to group consumers in some way according to their willingness to pay for the product (assuming consumers cannot profitably resell to one another and that willingness to pay varies) and to charge the groups different prices accordingly. This is known as 'price discrimination'. A simple example will illustrate. Suppose there are two towns between which travel costs are very large; one is populated with rich, price-insensitive consumers and one with poor households that are very price sensitive. A retailer with a shop in each town will find it optimal to charge a higher price in the rich town and a lower price in the poor town.

This is an example of geographical price discrimination. Other forms of direct discrimination can include selling the same good at different prices even within the same establishment (for example, selling cheaper cinema tickets to pensioners). Within grocery retail, it may be harder to detect these forms of direct discrimination. As discussed in Chapter 2, national retailers now largely operate national pricing policies. However, there may well be other ways in which firms can price-discriminate, such as the use of individually targeted vouchers and loyalty schemes, or operating multiple-fascia formats (such as Tesco Express or Sainsbury Local formats in inner-city areas which can charge higher prices). Discrimination can also be indirect, such as charging different unit prices according to the package size. We explore this issue in Chapters 4 and 5.

Another form of discrimination surrounds information and search costs. Strictly speaking, price variation arising in this strand of the literature is referred to as 'price dispersion' – in which different firms charge different prices for the same product – as opposed to 'price discrimination', which refers to one company charging different prices to different consumers for the same product. Nonetheless, the ability to distinguish between consumer types (direct discrimination) or to offer price schedules that induce consumers to self-select (indirect discrimination) is crucial to both. The difference is that price discrimination requires the seller to have some monopoly power whereas some models of price dispersion result in variation in final prices even in an otherwise competitive environment; search costs confer some price-setting ability to sellers.

The simplest Bertrand models assume that all buyers and sellers are fully and costlessly informed about all the available prices. The seminal contribution of Stigler (1961) was to point out that acquiring information – 'searching' for low prices – is costly. For grocery shopping, these costs might include travel costs to different stores and time costs involved in comparing prices. The marginal cost of visiting each additional store may be increasing – consumers will look at stores near their homes first, or stores with which they are already more familiar. The marginal benefits – the saving you expect to make from an additional search – are almost certainly decreasing, since the probability of finding a better price declines if you have already visited most of the stores in your area. Consumers will continue to search up to the point where the marginal cost is equal to the marginal benefit. Since the costs and benefits of search will vary across households (in terms of the number of stores in the local area, the availability of transport, their opportunity cost of time and so on), different consumers will search to different extents and so pay different prices.

The theoretical literature is somewhat technical and concerned with the precise mechanisms of information acquisition and competition among retailers, but the basic idea is that retailers face a trade-off between charging a high price to non-searchers and a low enough price to avoid driving searchers away. Given that price dispersion is supported, a general intuitive outcome is that, conditional on the benefits, consumers with lower search costs search more and pay lower prices.

Thus, to the extent that prices paid depend on search, we would expect to see relationships emerge in the data between prices and household characteristics correlated with search. Households with cars can feasibly search across a wider geographical area and may have lower search costs (subject to the costs of running a car), so we would perhaps expect people who shop by foot to pay more. The costs of search also include the opportunity cost of time spent searching, which is likely to be positively correlated with income and wages, number of children, hours worked, higher for lone parents and so on.[9] The benefits of search presumably depend on the total quantity of groceries purchased, so a household's total grocery expenditure would be negatively correlated with price, all else equal.

The empirical relationship we would expect between price and household income is ambiguous. High-income households may have higher opportunity costs of time, but also spend more on groceries and so stand to gain more from search. The relationship may also be complicated by unobserved household characteristics. For example, high-income households may have better cognitive skills, which improves the efficiency of each search as they can more easily process the information obtained from it. Given this theoretical ambiguity, an empirical investigation is extremely important.

In addition to spatial price dispersion, there is the issue of temporal price variation – temporary price reductions of goods, or sales. Several rationales for sales have been advanced in the literature. In some, such as that in Varian (1980), temporal price variation is the result of a randomised pricing strategy by firms in a search costs model to prevent

[9] See Burchardt (2008) for a discussion of time and income as constraints on economic well-being.

consumers learning which stores charge high and low prices. Salop and Stiglitz (1982) introduce storability. Consumers visit one store – if they find a high price, they buy only for immediate consumption, but if they are lucky and find a low price, they stock up. Again, sales are a randomised pricing strategy by firms which leaves them indifferent between selling fewer units at a higher price for immediate consumption and selling more at a lower price to consumers who also buy for future consumption. In Sobel (1984), temporary price reductions are a means of intertemporal price discrimination between consumers with a high reservation price and those with a low reservation price who are prepared to wait to buy until there is a sale. A more recent strand of literature[10] looks at promotions in a dynamic setting where consumers can stockpile the good. Some of this work suggests that sales of storable goods are a means to discriminate between consumers with high and low storage costs; Hendel and Nevo (2006b), for example, analyse detergent purchases and find that households in suburban areas with larger houses and lower storage costs hold larger volumes of inventory. They also conclude that these types of households can benefit from non-linear pricing schedules in which larger package sizes are offered at a per-unit discount.

The ability and willingness of households to exploit these bulk discounts (for example, a 24-pack of cola rather than a 6-pack) and store them in their homes for consumption, or take advantage of special offers that temporarily give additional volume for the same price (for example, buy-one-get-one-free offers), depend on the costs faced in transporting and storing higher additional volume relative to the price saving. Again, *ex ante*, it is not clear whether we expect poorer or richer households to use bulk discounts and special offers more: high consumption and low storage and transport costs mean searching for a low price is more worthwhile, but on the other hand those with large houses (i.e. room for storage) and high consumption are likely to be richer households with higher opportunity costs of time. Poorer households might also find making large, up-front payments difficult. Using the same data source we use in this study, Griffith et al. (2009) find that middle-income households tend to spend more on sale items than the poorest or the richest, but that buying in bulk is negatively related to income.

Within grocery retail, there are other reasons besides search intensity and use of sales to expect that households will pay different prices for identical products. Different stores may sell the same good at different prices across locations to reflect local cost conditions. Stores in urban areas may face higher rental prices and labour costs, which would lead them to charge more, though they might also face more intense local competition, leading them to charge less. Independent 'corner' shops are unable to exploit economies of scale in wholesale purchasing or other costs and so charge higher prices than large, national retailers for the same product. If poor households are concentrated in areas with high retail costs, then this may explain any finding that the poor pay more (as was suggested by some of the studies considering the 'food deserts' issue we examined in Chapter 2). Using our data, we can control for local area characteristics and see whether there is any residual role for household income in explaining variation in prices paid.

[10] For example, Boizot, Robin and Visser (2001), Erdem, Imai and Keane (2003) and Hendel and Nevo (2006a and 2006b).

Similarly, different stores may charge different prices to reflect the 'retail experience' of shopping there. Consumers can trade off the utility of a higher-quality shopping environment with a higher price. We are able to use information in our data on the stores at which people purchase to assess the extent to which different stores charge systematically higher prices. The product differentiation literature suggests, though, that retailers may also attempt to differentiate themselves from one another to ameliorate competition.[11] Thus any systematic difference we see between shops may only partly reflect service costs, with other differences reflecting market power induced by product differentiation. One specific dimension along which supermarkets can attempt to differentiate themselves is price format. In the marketing literature, two common price formats are EDLP (Everyday Low Pricing) and HiLo (promotion oriented). It has been suggested – for example, by Lal and Rao (1997) and Bell and Lattin (1998) – that these price strategies help stores attract different consumer types. For example, time-constrained consumers will tend to visit EDLP stores whereas bargain hunters will prefer the HiLo format.

[11] See, for example, Hotelling (1929), Salop (1979), Shaked and Sutton (1982) and Economides (1989).

CHAPTER 4
Data and Methodology

4.1 Data

This report uses data from Kantar Worldpanel.[12] The data contain the grocery purchases made by a nationally representative panel of British households (the data exclude Northern Ireland) using bar-code scanners installed in the home. Panel members scan the bar codes of items that are brought into the home. Data on the products are collected from the bar code and sent electronically to Kantar, and information on the prices paid is collected from till receipts which are mailed in periodically and matched to the purchase record. The data contain a record of each product purchased on each trip by each household, and products are distinguished by bar code such that different brands and package sizes, for example, are all uniquely distinguished. Households can remain part of the survey for as long as they wish and are rewarded for participation by high-street vouchers.[13] In principle, all household members should take part and the data should be a complete record of a household's purchases, though in practice (see below) there are known difficulties in recording purchases by household members other than the designated 'main shopper'.

We use data on households' purchases of 'fast-moving consumer goods' (FMCG) – essentially, items purchased largely in supermarkets including some non-foodstuffs such as personal care products and cleaning materials. We limit our analysis to purchases of food and drink items, however, including alcohol.[14] Data on non-bar-coded items, such as loose fruit and vegetables, are collected by means of a booklet of bar codes which the panel member scans; the member then enters some details about the product, such as type, weight and country of origin.[15] In principle, the data should provide complete records of all food and drink purchases brought into the home, from all stores (not only supermarkets but also corner stores, speciality shops, online food shopping and so on). The data we have currently span a six-year period from November 2001 to November 2007, though our primary focus is on results covering calendar year 2006 (in particular because income data are only available relating to this year).

The data include information on the characteristics of the products purchased (price, flavour, whether it is organic, packaging, size and so on) and demographic information on

[12] Formerly TNS Worldpanel. See http://www.kantarworldpanel.com/#/Home.

[13] The vouchers do not cover stores that are included in Worldpanel so should not affect purchase records.

[14] This is largely because the data we have on food and drink have been much more meticulously cleaned, improved and made consistent relative to the raw data received from Kantar than the data on non-food items. As far as we are able to tell, there is relatively little difference in the pattern of results if our data include non-food or exclude alcohol.

[15] From 2006, the sample size was increased from around 15,000 to 25,000 households. Members of the additional sample were issued with a new type of scanner unit which, Kantar argues, made scanning quicker and simpler and increased compliance among panellists in recording their bar-coded purchases at the expense of no longer requiring them to scan any non-bar-coded items. Again, we do not think that affects our results.

panel members such as age, sex, family composition, employment status, housing tenure, shopping habits (including the usual mode of transport taken) and durable ownership. Demographic data are in principle updated approximately every nine months or so. We also know the postcode district (the first 'word' of the postcode) in which the household lives and can use this information to map in data on the local area from government surveys such as the Census and Neighbourhood Statistics. Income information was collected from 2006 for a subsample of households.

As well as information on the products and households, we know the store from which the purchase was made. For chain stores, the data record precisely which store was used (to the postcode), whereas for corner shops and small stores, only the type of shop is known.

There is a vast amount of information. Over the whole six years, around half a million unique products (bar codes) are recorded, and a typical week of data records around 600,000–1,000,000 purchases worth some £1 million. An analysis such as we carry out here simply could not be carried out using other, more familiar data sets on household spending used in social science research. The UK Living Costs and Food Survey (LCFS),[16] for example, has information on only around 7,000 households in any year and records details of their spending on a large but still quite aggregate set of products, which means that there is considerable within-category variation in precisely what is purchased. In addition, this sort of consumer scanner data is unique in recording both prices (p) and quantities (q) purchased at the product level. Other spending data sets group many products together within a spending category and either report only total category-level expenditures or report both expenditure and total volume from which a 'unit price' can be inferred as the ratio of the two. These unit prices still suffer from aggregation, however, meaning that they are affected by quality variation of the goods within the group.

Recent studies of the quality of these data for use in empirical social science research[17] suggest that in terms of expenditures and demographics, the data match well to other spending information. There are some known issues: levels of spending in the Kantar data appear lower than those reported in the Expenditure and Food Survey, for example, though the *patterns* of expenditure appear similar in the two surveys, suggesting that the extent of this 'under-reporting' is similar across broad product groups. Discussions with Kantar suggest there are particular issues in collecting information from small trips, such as top-up trips to corner stores, and any purchases made by household members who are not the 'main shopper'. To some extent, this may affect our results, since prices in corner stores tend to be higher than in supermarkets, though it is likely that the vast majority of a household's typical food purchases occur as part of main shopping trips, which appear to be reported well.[18] These studies have also suggested that demographic transitions are not reported particularly well in the data (for example, too few households appear to retire or

[16] The LCFS replaces the Expenditure and Food Survey (EFS) from 2008. The EFS was itself an amalgamation of the old Family Expenditure Survey (FES) and National Food Survey (NFS) that began in 2001–02.

[17] For example, Leicester and Oldfield (2009a and 2009b) and Griffith and O'Connell (2009).

[18] A study of similar data in the US by Einav, Leibtag and Nevo (2008) compared scanner data with records of trips from supermarket checkouts and reported that most trips in the scanner data were reported accurately and fully.

change employment status) and discussions with Kantar have backed up this claim.[19] This does not pose any particular issues for our analysis since we make relatively limited use of the panel aspect of the data. The data also deliberately oversample some groups who are hard to recruit and retain, to ensure a reasonable sample of compliant households can be obtained for Kantar's own analyses – these include multiple-adult households and households with young children. We can control for these observable characteristics in our analysis.

Since much of our analysis will focus on household incomes, it is worth spending a little time describing the income variable in the data. Demographic information on the households is collected at the time of sign-up to the survey via a telephone interview with the person designated the 'main shopper' in the household. Income information was not collected in the Kantar data before 2006. In 2006, a one-off question was asked to a subsample of the active panel members that, as we understand it, asked the main shopper to report the household's gross income from all sources, grouped into bands of £10,000 (up to £70,000 or more). Around 73 per cent of households in the entire data set that we observe as active participants during 2006 have a non-missing response to the income question, though the earlier a household signed up to the data (our sample includes households that began reporting as far back as 1989), the less likely they are to have a valid income response. Although the income question was not repeated in subsequent years, it is our understanding that Kantar plan to include an income question on a regular basis going forward.

In their comparative analysis of the demographic information, Leicester and Oldfield (2009b, figure 8) find that, on average, incomes were lower in the 2006 Kantar data than in the 2006 Expenditure and Food Survey. They find that 69 per cent of Kantar households reported incomes below £30,000 per year compared with 56 per cent of EFS households, whilst just over 5 per cent of Kantar households reported income over £70,000 compared with 13 per cent of EFS households. However, similar fractions of households in the two surveys reported very low incomes, of less than £10,000, and what might be termed 'middle incomes', of £30,000 to £60,000.

This suggests two obvious possibilities. First, along the income dimension, the sample of households in the Kantar data may not be particularly 'representative' of the wider British population. Second, incomes may be measured with non-random error and more often under-reported than over-reported. This may be because the main shopper does not know the true income of others in the household (perhaps from non-labour sources) or misremembers it, whereas in the EFS data incomes are derived from a long series of detailed questions asked to each household member individually.

If the first possibility is true, it is not particularly worrying for our analysis. We are not looking to document the distribution of prices paid across the whole population, but

[19] We infer age from information on the date of birth of household members, so age does change over time. It is likely that the largest problems occur with variables such as employment status and housing tenure, which are not often used as part of market research analysis, which puts relatively little commercial pressure on Kantar to maintain these variables over time. By contrast, information on household composition, including number of children, appears to be better maintained, though fewer births are observed in the Kantar data than in other panel data such as the British Household Panel Survey (BHPS). See Leicester and Oldfield (2009b).

rather seeking to assess how prices vary along various observable dimensions that we can control for. Leicester and Oldfield (2009b) provide some evidence that the Kantar data are more likely than the EFS data to contain both older households and part-time and unemployed households, which would probably have lower income on average. If the second case is true, then it may be that some high-income households appear to have lower incomes in our data, which may muddy our results somewhat. However, given the relatively wide bands of income that we have, it seems fairly reasonable to assume that the ordering of reported incomes across households is a meaningful reflection of their true rankings. It is not very likely, for example, that a household with a true gross total income in excess of £70,000 would make such a large reporting error that it appears to have a very low income.

4.2 Methodology

4.2.1 Calculating price indices

Given information on households' purchases of thousands of products, we need to find some way to compare the overall prices paid by different households. This requires the computation of a price index. There are many ways to do this, and one of our key contributions is to compare a range of such methods to assess whether they tell similar stories about which sorts of households pay higher prices. Here, we describe briefly the different indices. Full technical details of their definition and calculation can be found in Appendix A.

One way to define a price index is to take some 'fixed basket' of items and compare the price paid for that basket by different households. This means that variation in the price index reflects differences in prices rather than differences in what is bought, an attractive property of such indices. However, not all households buy every item in a given basket. This means prices will be missing in some cases, so we need to find some way to impute the price that would have been paid had the household bought the item.

Our approach is to group households together according to observable characteristics, such as income, age and family composition, and to use the average price paid for a good within that group. In effect, we impute the missing price for a household as the average of the price paid by households in the group to which it belongs. This significantly reduces the problem of missing prices because now we can examine fixed baskets containing goods purchased at least once by a household in each group, rather than baskets containing goods purchased at least once by every household.[20]

Drawing on the literature on multilateral indices for making international price comparisons (Balk, 1996; Hill, 1997; Diewert, 1999), we choose three established indices with reasonable properties:

[20] In fact, there are essentially no goods (defined as bar codes) that are purchased by all the households in our data.

- the **Laspeyres** index;
- the **EKS** (Elteto and Koves, 1964; Szulc, 1964) index;
- the **CPD** (Country–Product Dummy) index.

The Laspeyres index takes an 'average' basket of goods and compares its cost at group average prices to its cost at national average prices. The EKS index takes an average of all bilateral comparisons between different groups. The CPD is a regression approach where the (log) price of each good in each group is regressed on a set of group and good dummies. The estimated coefficients on the dummies for each household are then the logs of the price index numbers.

There are two main drawbacks to these fixed-basket indices. First, the assumption that the group average price is a good estimate of the missing price for a household may be poor. Second, indices based on fixed baskets can only be calculated at the group level rather than at the household level.

One index that is not subject to the problem of missing prices is the **Paasche** (or 'own-basket') index, which compares the actual price of each household's own basket and its hypothetical cost at national average prices. It basically asks 'How much do households gain or lose buying their own basket at their own price compared with the average price?'. One benefit of this index is that we obtain one price index for each household which can be related to household-specific characteristics. This approach has been used in previous literature.[21] One potential drawback of this approach is that indices vary by what is bought as well as by prices paid. To take an extreme example, imagine that all households purchased unique products that no other household buys. Then every household will have a price index of 1 since the 'average' price paid by all households is just the same as the household price. More realistically, there may be products that tend to be bought only by certain types of households. Suppose that rich and poor households buy completely different subsets of products. Then the Paasche index will compare the price paid by a particular rich household to the prices paid by other rich households, and the price paid by a particular poor household to the prices paid by other poor households. This may somewhat limit the variation in prices paid and the usefulness of the index. However, if our supposition – that rich and poor households buy completely different subsets of products – were true, then the fixed-basket indices would fare no better since there would be no overlap between the products purchased by rich and poor and so no way to price a single product for both groups. The act of calculating a fixed-basket index does, though, force us to address the issue of how far products bought by different household groups overlap, which might otherwise escape notice.

By comparing the results from the different indices, we can assess the extent to which the choice of price index methodology drives the pattern of which households pay more or less. Ultimately, we would like to use the own-basket index to examine the factors associated with paying higher or lower prices, but if the results for that index (averaged to the household-group level) look substantially different from the results for other indices,

[21] For example, Carlson and Gieseke (1983) and Aguiar and Hurst (2007).

we need to be wary. Thus our first set of results in Chapter 5 compares price indices across groups.

4.2.2 Defining a 'product'

One of the key benefits of using the household scanner data for this analysis is that we observe precisely, at the bar-code level, what has been purchased. This means that when we compare the price paid by different households for the same product in the data, we know that we are comparing like-with-like and so do not need to worry about potential unobserved quality variation that may drive price differences. For example, if we were to try our analysis on data such as the Living Costs and Food Survey, where purchases are aggregated into a large number of different groups such as 'eggs', then it may be that rich households would buy expensive, free-range eggs and poorer households would buy cheaper, battery eggs. The rich would then appear to pay more for eggs, but this is partly driven by differences in product quality rather than different prices for the same product.

However, whilst the extremely disaggregate nature of the data is a virtue in this respect, it can also mask some ways in which households may be able to economise on their grocery purchases, which could be important. First, many goods are offered in a range of pack sizes and, typically, larger pack sizes are sold at a lower per-unit price. Because the data record different pack sizes of the same item as different products, our price indices will not capture the impact of bulk discounts on the unit price paid. As discussed in Chapter 3, economic theory is ambiguous about whether poor households are more likely to take advantage of these 'bulk discounts' than rich households.

Second, some households choose to buy own-brand or 'generic' versions of groceries rather than branded versions. Most national stores offer own-brand varieties of almost all grocery items, and in recent years there has been an increasing differentiation even within own-brand goods, with retailers offering 'economy', 'standard' and 'premium' versions of own brands targeting different consumers. Typically, own-brand products are substantially cheaper but of lower perceived or actual quality than branded variants of goods. Using Kantar data, Griffith et al. (2009) find own brands to be around 25 per cent cheaper than branded products and economy-range own brands 39 per cent cheaper. They also show that own brands are very popular: standard own brands account for around 40 per cent of expenditure and economy own brands about 4 per cent, with poorer households being much more likely to buy economy own brands.

To the extent that own brands are of different quality, we would not want to compare the price paid for, say, a generic pack of cornflakes with that paid for a branded pack, but we might be more willing to assume that the generic cornflakes offered by different stores are ostensibly the same product and of the same quality. Our data, however, treat them as different goods. This may limit the overlap of products purchased by different groups if some household types shop mainly in one chain and some types shop mainly in another.

We use the detailed information on product characteristics in the data to look at the impact of bulk discounting and own-brand purchases by aggregating products together and re-running our analysis. We define two different product identifiers to use alongside the bar code for our analysis: *packsize* and *ownbrand*. Products in the data are assigned to one of almost 200 different categories, each with a different set of observed product

characteristics. To generate the *packsize* variable, we group together products that have the same observable characteristics other than the package size; to generate the *ownbrand* variable, we group together generic products that have the same observable characteristics (including size), treating economy and premium versions of own brands as different products from standard versions. For the *ownbrand* definition, branded items are left as unique products rather than grouped together.

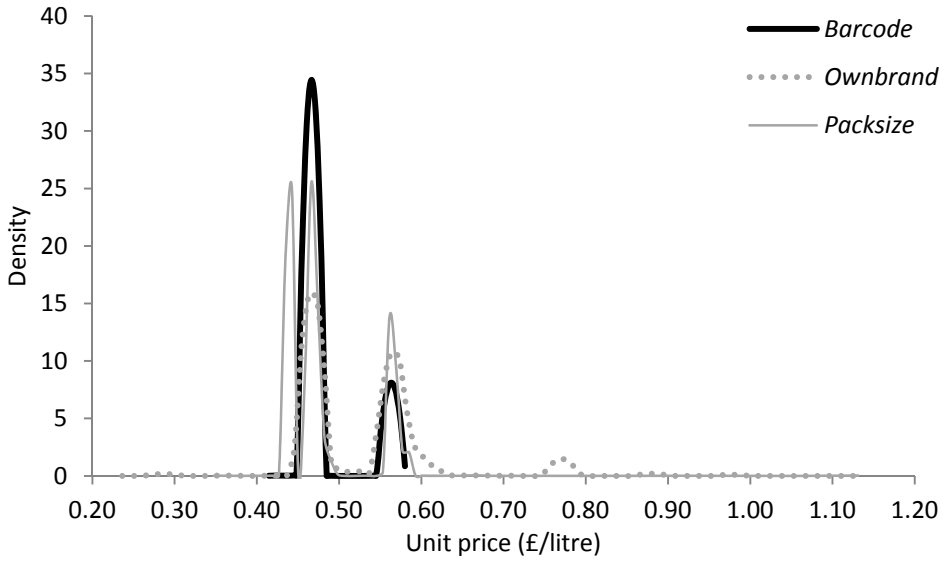
To illustrate both our data and these different definitions of 'product', we present below some descriptive information regarding prices and products in a single month, June 2006. In that month, we observe 2,006,158 separate purchases by 18,750 different households, buying 64,347 unique products (bar codes). Grouping products together into the *packsize* and *ownbrand* variables reduces the number of 'products' fairly substantially, to 50,296 on the *packsize* definition and 48,709 on the *ownbrand* definition, falls of 21.8 per cent and 24.3 per cent respectively.

Perhaps unsurprisingly, there are significant differences across different product categories in the extent to which redefining bar codes into pack-size and own-brand groups reduces the number of products we observe. For items such as fresh meats, milk, fruit and vegetables, cold drinks and canned goods, grouping together items of different sizes typically reduces the number of products by around one-third to one-half. These are items often sold in large numbers of different pack sizes. Some of these goods – for example, meat and fresh fruit and vegetables – also see large falls in the number of products once we group together different stores' own brands. Some items – for example, bread – do not appear to have much variation in pack size (the number of products falls by only around 14 per cent when we group sizes together) but do see very similar competing own-brand versions (the number of products falls by 42 per cent on the *ownbrand* definition). Others such as mineral water, by contrast, see much more variation in pack size (the number of products falls by 40 per cent when we group by size) than store own brands of a given size (falling by 7 per cent).

Figure 4.1 shows an example of how price variation changes when we redefine products. The black line shows a kernel density plot of the unit price (£/litre) of a single bar code in June 2006: a two-pint carton of a particular national supermarket's semi-skimmed milk. There were 1,569 purchases of this product in the period, sold at two prices: 47p/litre and 56p/litre. The dotted grey line shows the unit price of all the products in the same own-brand group: in effect, the two-pint cartons of semi-skimmed milk sold in different supermarkets.[22] Although the modal points of the price distribution are similar, there is now more variation: unit prices range between around 25p/litre and £1.13/litre for ostensibly the same item purchased in different stores. The thinner grey line shows the unit price for the products in the same pack-size group, effectively showing bulk discounting for semi-skimmed milk purchases from one supermarket. We can see clearly different peaks for different sizes of milk. Modal unit prices are 56p/litre for one-pint cartons, 47p for two-pint cartons and 44p for four- or six-pint cartons.

[22] Other characteristics, such as the type of container and whether it is organic, are held constant.

Figure 4.1. Price variation for a single 'product' – semi-skimmed milk



Source: Calculated from Kantar Worldpanel data, June 2006.

This graph gives us an idea of how price variation changes with product definition for a single item. To look more generally at how much price dispersion exists, we calculate the coefficient of variation of the unit price for each product and each product definition in June 2006. The coefficient of variation is the ratio of the standard deviation to the mean. The standard deviation is a measure of dispersion – in a normal distribution, for example, 68 per cent of observations lie within one standard deviation of the mean – and dividing by the mean gives a unit-free measure of dispersion.[23] For each product definition, Table 4.1 shows the median value of the coefficient of variation (expressed as a percentage), first for all products and then limiting the sample to products purchased multiple times during the month.

Table 4.1. Median coefficient of variation by product type, June 2006

	Barcode		***Packsize***		***Ownbrand***	
	Median	N	Median	N	Median	N
All products	0.37%	64,347	2.14%	50,296	1.49%	48,709
>1 purchase	2.89%	54,874	4.95%	43,888	4.55%	41,543
>10 purchases	6.22%	28,953	8.80%	23,991	8.69%	22,354
>100 purchases	11.02%	3,967	15.09%	3,915	14.48%	3,793

Source: Calculated from Kantar Worldpanel data, June 2006.

[23] A coefficient of variation of 0.1, for example, means that the standard deviation of the observed purchase prices is 10 per cent of the average purchase price for that product; for a normal distribution with mean 200, say, this would imply that 68 per cent of observations lie between 180 and 220.

There is considerable variability of product prices. Even when products are defined at the bar-code level, the average standard deviation is more than 6 per cent of the mean price for products purchased more than 10 times during a month. Unsurprisingly, more frequently purchased products have more variation in the price paid. Changing the definition of a product has a large positive impact on the variability of prices.

4.2.3 Household selection

Our results in Chapters 5 and 6 will focus on two main samples of households in the data. In Chapter 5, we look first at different price indices and focus on households observed in June 2006. There are 18,750 households in that month. We do not exclude any households from these results on the basis of their total observed spending; we carried out some sensitivity analysis which suggested this had very little impact on our findings.

Later in Chapter 5 and throughout Chapter 6, we focus on how the own-basket price index varies with observable household characteristics, using an average of a household's monthly price indices over calendar year 2006. In order to strip out households that report infrequently in the data, we use only those households that record at least £10 of (equivalised) expenditure in six or more separate months during 2006. This gives a sample size of 17,990 households, of which 12,885 have non-missing income values. Some of our analysis is restricted to households in England, of which there are 15,479 observations of which 11,066 have non-missing incomes.

CHAPTER 5
Results

5.1 Comparing price indices

We begin by comparing the different price indices described in Section 4.2.1 to assess whether they tell a consistent story as to which groups pay higher prices. We focus on results across different income groups, though results for other demographic groups, including by number of children, age and family composition, are summarised in Section 5.1.2 and detailed in Appendix B.

5.1.1 Prices paid by different income groups

Table 5.1 shows our estimated price indices according to the equivalised income group of the household.[24] We use data from a single month, June 2006. As discussed in Chapter 4, the Laspeyres, CPD and EKS indices are calculated at the income-group level whereas the own-basket index is calculated at the household level and then averaged across the groups for comparison.[25] We show the own-basket index in two ways: first, using all of the products in the data and, second, using only those products that are bought by all the income groups and so form part of the 'overlap' of products that are used to construct the Laspeyres and EKS indices.[26]

There are several interesting features of the results:

- Prices rise with income group: richer households pay more on average than poorer households.
- There is relatively little variation in the price index on average across income groups, despite the variation in the underlying product prices. The range between highest and lowest index is typically only 1–2 percentage points.
- The results for different indices are broadly comparable across different income groups. All the indices suggest the richest households pay the highest prices and the lowest prices are paid by households with equivalised incomes between £10,000 and £20,000. The group with incomes below £10,000 appear to pay slightly more but have the second-lowest price index in all cases.

[24] Income data are available in bands of £10,000 of household gross income. To equivalise, we take the mid-point of each band and adjust according to the modified OECD equivalence scale, which assigns a value of 0.67 to the first adult, 0.33 to all other adults and children aged 14+ and 0.2 to all younger children. Thus childless couples have a value of 1. The highest band in the data is £70,000 or more, for which we take £75,000 as the mid-point. Note that because relatively few households have very high incomes, we report equivalised incomes for bands of £10,000 width up to £60,000 or more.

[25] The own-basket index is normalised so that the average over all households is 100. The CPD index is set to 100 for an arbitrary base group (in this case, those with incomes over £60,000) and other indices are expressed relative to this base group.

[26] The CPD index can only be calculated for products that are bought by two or more of the groups.

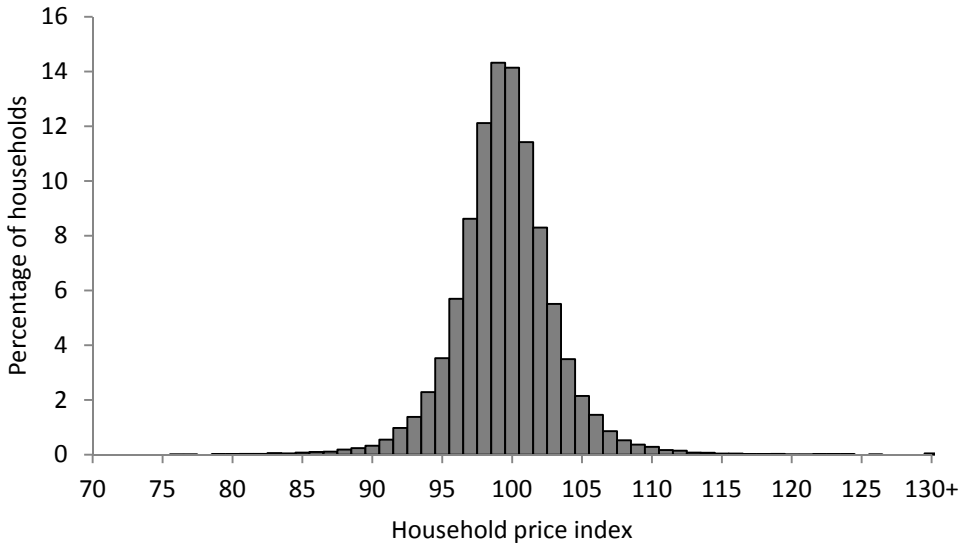
- Those households for which no income data were collected appear to have indices that are similar to those in the middle of the income distribution. This suggests that this group of households is not systematically different in income terms from those whose income data were collected.

- The overlap of products bought by all the income groups accounts for around half of total spending. Calculating the own-basket index just using these products does not change the results much.

Table 5.1. Price indices by equivalised income group, June 2006

Income group	**Laspeyres**	**CPD**	**EKS**	**Own-basket**		**N**
				All goods	Overlap	
Not collected	100.11	98.93	99.72	100.08	99.98	5,807
Under £10k	99.97	98.73	99.52	99.86	99.72	2,296
£10k–£20k	99.76	98.68	99.36	99.79	99.56	3,596
£20k–£30k	100.06	98.97	99.67	100.06	99.87	3,629
£30k–£40k	100.09	99.00	99.66	100.08	99.81	1,827
£40k–£50k	101.19	99.64	100.58	100.54	100.55	670
£50k–£60k	101.34	99.89	100.70	100.79	100.66	541
£60k+	101.53	100.00	100.82	100.85	100.70	384
Range	*1.77*	*1.32*	*1.46*	*1.06*	*1.14*	
Overlap	50.5%					

Figure 5.1. Household-level price index distribution, June 2006



The fact that the results are similar across different price indices gives us confidence in using our own-basket index to explore price variation at the household level. Looking at the household-level indices reveals considerably more variation than the group-level averages. Figure 5.1 shows the distribution of the own-basket index in June 2006 (restricted to households with an equivalised total spend of £10 or more in the month). Each bar has a width of 1 such that households in the '100' category have a price index

between 100 and 100.99, households in the '101' category have an index between 101 and 101.99 and so on.

The vast majority of households – over 98 per cent – have a price index in the range 90–110, but this still represents quite a lot of variation relative to the small average differences across income groups. Table 5.2 shows within-income-group quantiles of the household price index. These show similar patterns across groups to the averages shown in Table 5.1: within income group, poorer households (i.e. those at a lower quantile point) tend to have a lower price index than richer households but the variation across income groups is small.

Table 5.2. Within-income-group percentile points of household-level index, June 2006

	Percentile point				
Income group	**10th**	**25th**	**Median**	**75th**	**90th**
Under £10k	95.55	97.82	99.85	101.87	104.14
£10k–£20k	96.04	97.96	99.65	101.47	103.40
£20k–£30k	96.06	98.04	99.98	101.85	103.94
£30k–£40k	96.09	98.20	100.06	101.98	104.04
£40k–£50k	96.73	98.57	100.45	102.16	104.00
£50k–£60k	97.06	98.84	100.52	102.43	104.69
£60k+	96.48	98.72	100.28	102.35	104.48
All households	96.00	98.06	99.96	101.86	103.97

Note: Restricted to households with an equivalised total spend of £10 or more in the month.

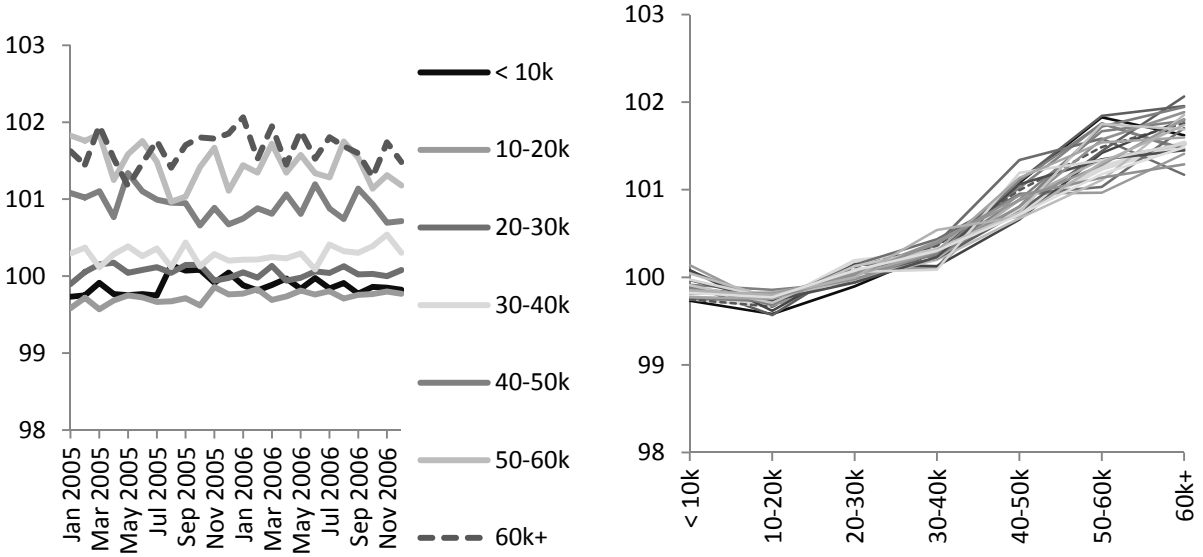
Taken together, these results suggest that although there is variation across households in the average prices paid for identical products (households at the 90th percentile of the distribution of prices have an index around 8 per cent higher than those at the 10th percentile), it does not appear to be much driven by income.

A key question is whether these results look stable across time. Repeating this analysis for 24 months from calendar years 2005 and 2006 suggests that the pattern of which household income groups pay more is not driven by the choice of month. Figures 5.2a, b and c show the Laspeyres, EKS and own-basket indices across time by income group (left-hand panel in each) and the indices across income group over time (right-hand panel).[27] The left-hand charts in each case show that there is no particular trend for any income group in their price index over time and the right-hand charts show clearly that the extent of month-on-month variation in the group-level indices is quite small. Typically, there is a little more variation in the index for higher income groups than lower income groups, which may partly reflect smaller sample sizes. The Laspeyres index shows the greatest range across income groups whilst the own-basket index shows the smallest range. Most important, though, is that the indices look stable across time.

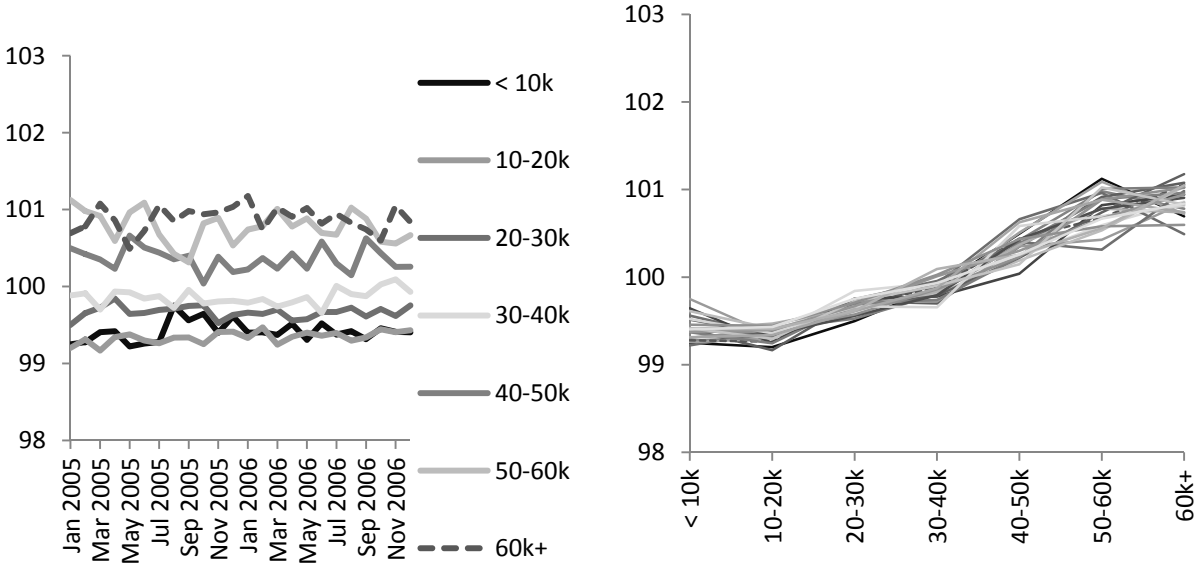
[27] The CPD is omitted for space reasons, but is available on request, as are similar figures for other demographic groups.

Figure 5.2. Indices across time, January 2005 – December 2006

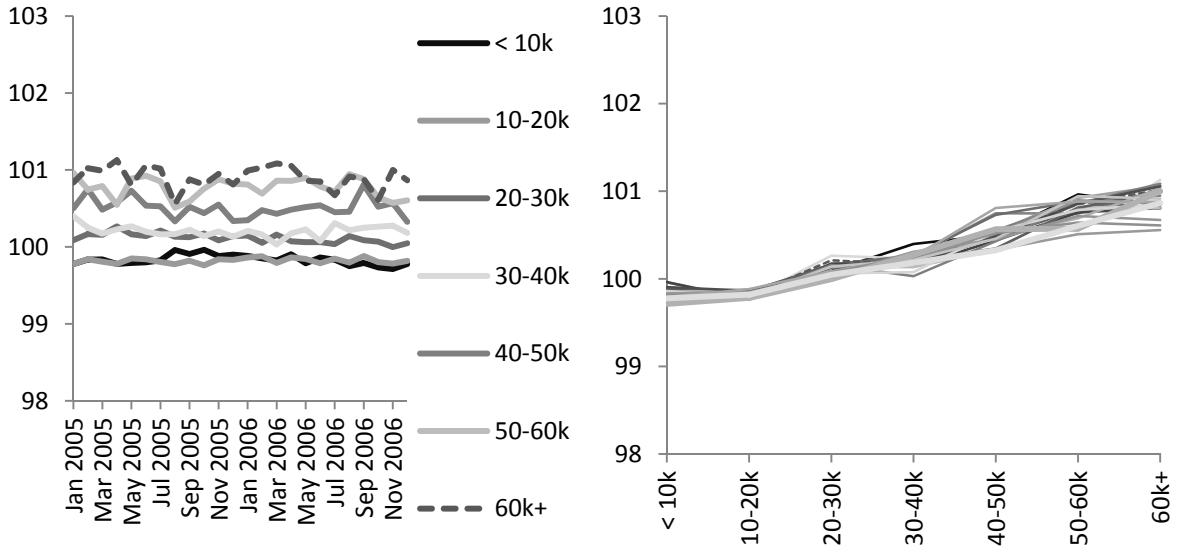
(a) Laspeyres index



(b) EKS index



(c) Own-basket index



Another way to assess whether the choice of a particular month affects the results is to consider the ranking of the different income groups' price indices in each month and whether it changes. The left-hand panels of Figure 5.2 show that there are occasional changes in the ranking of which groups have higher or lower indices from month to month, but broadly the rankings are quite stable. In all months and for all indices, the income group with the lowest index is always either the under £10,000 group or the £10,000–£20,000 group. The highest index is always that of the £50,000–£60,000 group or the £60,000+ group.

The group average price indices therefore do not appear very sensitive to either the choice of index or the choice of a particular month of data. One other robustness check is to see whether the period of data over which we calculate the index matters. Re-running the price indices using data from the entire calendar year 2006, we find very similar results (available on request).

As can be seen from Table 5.2, the variation of the price index within an income group is much larger than the variation of the average index across groups. Another interesting question is how much variation there is in a household's own price index in different months and how that variation compares with the average variation across households within an income group. Table 5.3 shows, by income group and across all households, the within-household and between-household standard deviation of the household-level price index. The 'within' standard deviation shows the typical variation of the price index of a particular household over time, whilst the 'between' standard deviation shows the variation of a household's average price index across different households (in that income group).

Table 5.3. Within- and between-household price index variation, 2006

Income group	Mean index	St. dev. (overall)	St. dev. (between)	St. dev. (within)	No. of households	No. of household-months
Under £10k	99.76	3.891	2.702	2.786	2,196	24,466
£10k–£20k	99.74	3.310	2.287	2.374	3,582	38,989
£20k–£30k	99.97	3.529	2.375	2.583	3,615	39,399
£30k–£40k	100.08	3.380	2.226	2.535	1,830	19,875
£40k–£50k	100.38	3.061	1.937	2.349	711	7,491
£50k–£60k	100.69	3.238	2.047	2.521	551	5,960
£60k+	100.70	3.671	2.300	2.827	400	4,258
All households	99.97	3.549	2.408	2.591	17,990	193,722

Note: 'All households' includes households with unobserved incomes, which explains why the total number of observations is not just the sum of the individual income groups' observations.

The results suggest that the variation in price index within a household is as large as, if not larger than, the variation in price index across households. Within-household variation in price indices seems to be similar across different income groups. Interestingly, between-household variation appears to be slightly larger for poorer groups than for richer ones. Overall, it appears that poorer households pay, on average, slightly lower but more variable prices than richer households.

In summary, the average price index increases modestly across income groups and this result is stable over time. However, there is considerable variation around that mean across households within an income group and over months for a particular household.

Given the variation in price index over time within households, we can compare the month-by-month results from Table 5.1 and Figure 5.2 with the results obtained from taking an average of a household's own-basket price indices over the months in which they are observed.[28] Table 5.4 shows the results of a simple regression of the natural logarithm of this household-specific average index on dummy variables for the equivalised income groups, excluding those with missing income data.[29] The coefficients, reported in the β column, represent estimates of the percentage difference in the price index of each income group compared with a reference group, which in this case is households with equivalised incomes below £10,000. The dependent variable (log average price index) has been multiplied by 100 such that a coefficient of 1 means the group pays 1 per cent higher prices than the reference group. We also show the estimated standard error (se) around these estimates and the *t*-value (values above +2 or below –2 suggest the estimated coefficient is statistically significantly different from zero at (at least) the 5 per cent level.

Table 5.4. Variation in average price index by equivalised income group

Income group	β	se	t	N
< £10k (excluded)	—	—	—	2,196
£10k–£20k	0.000	0.063	0.00	3,582
£20k–£30k	0.213	0.063	3.38	3,615
£30k–£40k	0.320	0.073	4.36	1,830
£40k–£50k	0.630	0.100	6.29	711
£50k–£60k	0.951	0.111	8.60	551
£60k+	0.911	0.126	7.22	400
R^2	0.013			

Notes: Dependent variable is log of household average price index multiplied by 100. Coefficients can be interpreted as approximate percentage difference compared with the excluded group.

The results confirm the broad findings so far: the price index increases with income but the variation across income groups is small. Households with an equivalised income above £50,000 pay prices that are just less than 1 per cent more than those with incomes below £10,000. The index rises across income groups, though those with incomes between £10,000 and £20,000 do not appear to pay any more or less than those with incomes below £10,000. The R^2 is very low, suggesting that income alone explains very little of the variation in price indices across households.

[28] Precisely, we use the household's own *plutocratic* average price index. That is, we weight the monthly household index by the household's total spending in that month to calculate the overall average. This just gives more weight to the months in which a household spent a lot than to the months in which it spent a little as a measure of its typical price index. Results using simple average price indices are not very different.

[29] If we regress the household-level average price index on a dummy variable for households without a reported income, the coefficient is insignificant. This suggests that households with a missing income value do not, on average, pay different prices from those with a recorded income value.

5.1.2 Prices paid by other household groups

The analysis so far can be replicated for other household groups, not just income. Appendix B shows sample sizes and regression results similar to Table 5.4 for households by number of children, age group and family composition. In each case, we examine whether there is evidence that 'poor' households pay different prices, where 'poor' households are those with an equivalised income below £20,000 (we choose this definition of 'poor' as Table 5.4 found no evidence that households with incomes between £10,000 and £20,000 paid more than those with incomes below £10,000). We do this by including, in a second specification, a dummy variable indicating that the household is poor and also interacting this dummy with the demographic group. This interaction term tells us whether the impact of being poor varies across our groups of interest.[30] Also in Appendix B, we look at households according to a measure of local area deprivation. We now provide a brief summary of the main findings.

Households with children

We define children as all individuals under 18 in our data. About two-thirds of households in the sample are childless, a similar proportion to that in the 2006 Expenditure and Food Survey. Households with more than one child are substantially more likely to have equivalised incomes below £20,000, which at least partly reflects the higher equivalence scales for such households.

Since few households (only 167) have more than three children, our regression analysis groups together all households with three or more children into a single group. The results show that as households have more children, the prices they pay fall slightly on average. Households with three or more children pay around 0.7 per cent lower prices than those without any children. The effects are small but statistically significant.

The interaction results show that having children does not significantly affect the impact of low income on prices paid. Being poor is associated with significantly lower prices but, for example, poor households with three or more children do not pay even lower prices than poor households with no children.

Age group of household head

Younger households (with the exception of those with a head aged under 25) are significantly less likely to have low income than older households. Around 60 per cent or so of households with a head aged 65+ (above the male state pension age) have equivalised incomes below £20,000 compared with around 40 per cent of households with a head aged 25–59.

Our regression estimates suggest that prices do not vary much with age, except for older households. Those aged 70 or over pay significantly higher prices than those aged 25–29, with those aged 80+ paying around 1 per cent more. Looking at the interactions of age and income, we find again that, except at older ages, there is little evidence that age affects the

[30] The interaction term does *not* say whether, for example, a household with one child that is poor pays more or less than a household with one child that is not poor. This comparison requires us to test the statistical significance of the joint coefficient on poor and the interaction between poor and one child.

impact of low income on prices. However, those aged 65 or over who are poor do seem to pay higher prices than younger poor households (except those poor aged 80+, though the effect is insignificant). Taken together, our estimates suggest that poor households with a head aged 65–79 in fact pay no less than non-poor households with a head of the same age. For all other age groups, other than those under 25, there is evidence that poor households pay less than non-poor households in the same age group.

Family composition

The extent to which different types of households have low equivalised incomes differs substantially in the data. One- or two-person pensioner households are more than twice as likely to have low income as one- or two-person non-pensioner households. Single people are less likely to have low income than couples or multiple-adult households, and households made up of two adults with children are much less likely to have low income than households containing a single adult with children.

Compared with single non-pensioners, our results show that single pensioners pay slightly higher prices (consistent with the evidence on age), couples with children pay slightly lower prices (consistent with the evidence on children) and lone parents pay even less. The lowest prices are paid by perhaps slightly unusual households – those made up of three or more adults with children, which suggests that household size is negatively correlated with prices paid. The interaction results suggest that poor childless couples and poor couples with children pay less than poor single people, but otherwise there is no significant difference in prices paid by different types of households conditional on being poor. Comparing poor and non-poor households of a given type, our estimates show no evidence that poor lone parents, single adults or households made up of multiple adults with children pay lower prices than equivalent non-poor households, but in all other cases poorer households pay less.

Local area deprivation

One of the findings from the literature that has explored how prices vary has been that local area is a more important determinant of price paid than household income. To explore this issue, we examine the importance of local area deprivation in explaining the variation in prices paid and whether there is any additional role for household income.

Our measure of deprivation is taken from the Department for Communities and Local Government Indices of Multiple Deprivation (IMD) for England.[31] The data assign a deprivation index to each 'Super Output Area' (a very fine level of geography; there are 32,482 SOAs in England) based on various domains: income, employment, health and disability, education, housing, crime and living environment. We map SOA-level deprivation to the postcode district in which our households reside using postcode data from the Office for National Statistics (ONS) and we define deprivation 'tertiles', splitting postcode districts into three equal groups from least to most deprived.

Two patterns emerge. First, within deprivation tertile, higher-income households pay higher prices. This means local area deprivation alone cannot fully account for price

[31] For an explanation of the indices and the raw data, see http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/.

differences. Second, for households in a given income group, those living in less deprived tertiles typically pay more. The relationship is less clear for higher-income households, though sample sizes for high-income households are small once we split by deprivation tertile. Overall, the richest households in the least deprived areas pay prices that are around 1.5 per cent higher than those paid by the poorest households in the most deprived areas.

5.2 Comparing product definitions

In Section 4.2.2, we described how we could use the information about the items in our data to redefine what is meant by comparing prices for 'identical products'. This section examines how the price indices change when we do this.

Table 5.5. '*Packsize*' price indices by income group, June 2006

Income group	Laspeyres	CPD	EKS	Own-basket	
				All goods	Overlap
Not collected	100.08	97.79	99.11	100.00	99.72
Under £10k	100.34	97.69	99.14	99.77	99.61
£10k–£20k	99.58	97.39	98.57	99.56	99.08
£20k–£30k	100.30	97.94	99.34	100.08	99.82
£30k–£40k	100.69	98.05	99.55	100.06	99.78
£40k–£50k	102.42	99.06	100.82	100.76	100.73
£50k–£60k	102.82	99.42	101.10	100.96	100.73
£60k+	105.63	100.00	102.48	101.19	101.08
Range	*6.05*	*2.61*	*3.91*	*1.63*	*2.00*
Overlap	59.7%				

Table 5.6. '*Ownbrand*' price indices by income group, June 2006

Income group	Laspeyres	CPD	EKS	Own-basket	
				All goods	Overlap
Not collected	100.06	97.73	99.36	100.15	100.09
Under £10k	99.61	97.27	98.84	99.69	99.48
£10k–£20k	99.46	97.27	98.76	99.67	99.42
£20k–£30k	100.07	97.78	99.36	100.15	100.03
£30k–£40k	100.33	97.99	99.60	100.34	100.23
£40k–£50k	101.82	98.96	100.90	101.14	101.35
£50k–£60k	102.35	99.45	101.36	101.56	101.83
£60k+	102.93	100.00	101.92	102.21	102.46
Range	*3.47*	*2.73*	*3.16*	*2.54*	*3.04*
Overlap	60.3%				

First, we group together items that differ only according to the packet size (the '*packsize*' definition) to examine which groups benefit from bulk discounts. Table 5.5 presents results for June 2006 comparing price indices for the different definitions across income groups. Second, we group together equivalent own-brand products from different stores (the '*ownbrand*' definition) to examine which households appear to buy cheaper own-brand varieties. Results are reported in Table 5.6.

Comparing these indices with those for the '*barcode*' product definition in Table 5.1 reveals several interesting results:

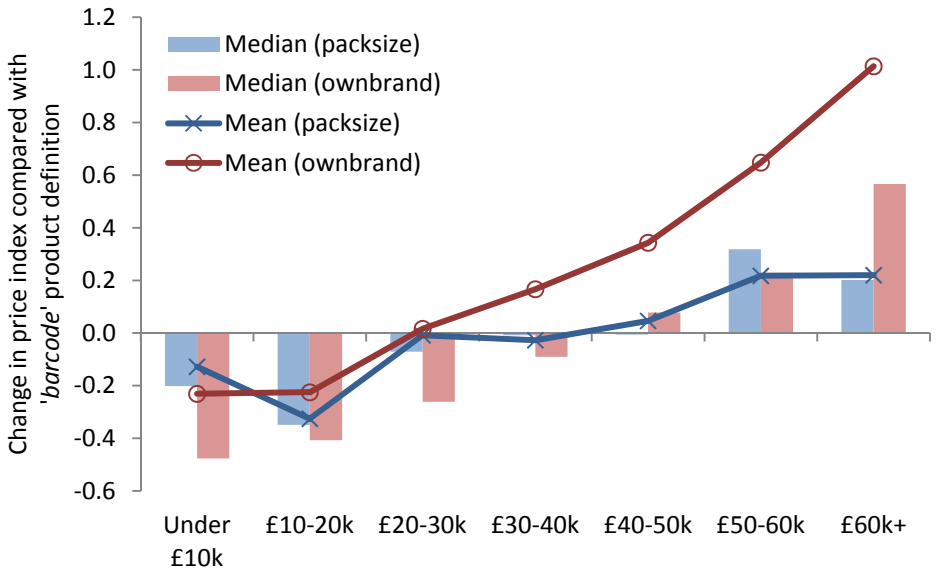
- On almost all price index definitions, it is poorer households that benefit from bulk discounts (in the sense that their index falls when we use the *packsize* definition rather than the *barcode* definition). The largest gains appear to be for those with incomes between £10,000 and £20,000, who often had the lowest indices on the *barcode* definition. The biggest increases in prices come for the richest group. Thus bulk discounting appears to benefit low-income households and accentuates the differences in prices paid across income groups quite substantially.

- Perhaps less surprisingly, it is also poorer households that buy cheaper own-brand products. The largest falls in the price index when we move to the *ownbrand* product definition come for those with incomes below £10,000, and for all indices there is a monotonically negative relationship between income group and the gains made from buying cheaper own brands. Again, the range of prices is accentuated by changing the product definition.

- The overlap of products purchased by all groups is raised substantially by changing the product definition, from about 50 per cent of spending in the *barcode* case to about 60 per cent of spending on both the *packsize* and *ownbrand* definitions.

Figure 5.3 illustrates the differences between the *barcode*, *ownbrand* and *packsize* product definitions across income groups graphically for the own-basket index, showing both the mean and median change in the household-level price index when moving from the *barcode* definition to the other definitions. Negative figures suggest the income group gains on average from bulk discounts or from cheaper own-brand products. The pattern that poorer households benefit and richer households lose out is clear; the median household in the £10,000–£20,000 income band sees its price index fall by about 0.4 percentage points when the product definition is changed, whereas those in the middle of the income distribution see very small average changes and rich households see their prices increase. Buying relatively more expensive own brand products (that is, relative to cheaper own brands as opposed to branded goods) appears to be a more significant factor than pack size in raising the prices paid by the richer households (the fact that the mean increase in the index is much larger than the median increase suggests there are a small number of households that lose out substantially from buying relatively more expensive own brands).

Figure 5.3. Average change in own-basket price index relative to '*barcode*' product definition, by income group



Note: Negative figure indicates group gains on average from switch in product definition (price index decreases).

Similar results based on number of children, age of the head and household composition are presented in Appendix B. Here is a summary of the findings:

- **Children:** Households with more children benefit substantially both from bulk discounting and from cheaper own brands. The median fall in the price index for households with three or more children on the *packsize* product definition is over 0.6 percentage points, whereas those without children see no change in their index. The median fall in the price index for those with three or more children on the *ownbrand* product definition is over 0.5 percentage points, compared with a fall of just over 0.1 percentage points for those without children.
- **Age group:** Younger households benefit substantially from buying cheaper own brands whereas older households lose out. Those under 25 see a median fall in their price index of around 0.5 percentage points on the *ownbrand* definition; those over 80 see an increase of around 0.5 percentage points. However, for bulk discounts, it is middle-aged households that see the largest fall in their price index; the youngest households see no change. The oldest households again lose out, seeing their price index rise by around 0.4 percentage points on average.
- **Household composition:** Single adults see the largest increase in their price index once bulk discounting is taken into account. For non-pensioners, the median increase is 0.2 percentage points; for pensioners, the increase is 0.4 percentage points. Households made up of three or more adults with children see the largest fall, of over 0.5 percentage points. The median households in all composition groups see their index fall when own brands are grouped together, but again the largest gain is for households of three or more adults with children.

CHAPTER 6
Explaining the Variation in Prices

This chapter attempts to provide some evidence on why prices vary. The aim is to use predictions and intuition derived from the various competing hypotheses regarding search and shopping behaviours discussed in Chapter 3 to examine the factors associated with higher or lower prices. We begin by looking at evidence that household search – both spatially across stores and temporally in terms of sales and bulk discounts – affects prices. Then we look at whether prices vary systematically over different types of stores.

6.1 Search

6.1.1 Spatial search

In Chapter 3, we discussed the possibility that prices vary because households differ in the extent to which they search for low prices. Households that search more should pay less, all else equal. It is not unambiguously clear whether richer or poorer households would be expected to search more, so it is important to investigate these issues empirically using the data.

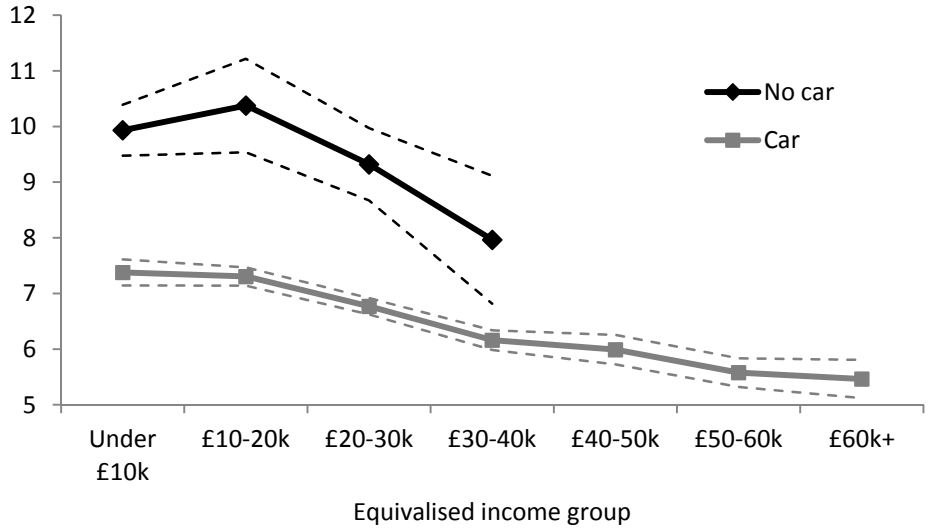
We begin by looking at shopping behaviours that reflect search. One possibility is to use the number of shopping trips that a household makes (where a trip is defined as a visit to a particular store on a particular day) – households that search more would make more trips. However, one concern might be that we then interpret households that make a regular – say, weekly – trip and a number of small, top-up trips as 'searching' more heavily than a household that just makes the weekly trip without additional top-ups. It seems unlikely that top-up trips are really about searching for low prices. We try to exclude such trips from our calculations by defining 'main' shopping trips as trips in which the household spends at least 20 per cent of its monthly average per-trip spend. The results when we use all trips are, however, much the same, though some of the relationships are marginally less significant.

Figure 6.1 shows how the number of main shopping trips per month varies with income group and whether or not the household has access to a car for shopping, which is likely to have an important effect on the ease with which households can search.[32] Not having access to a car makes search more difficult – households may be aware that a distant large supermarket offers lower prices but are simply unable to travel there. This might facilitate some forms of price discrimination. In addition, not having a car makes it more difficult to transport large quantities, which means households may have to shop more often (all else

[32] The survey asks households both whether they have a car and the mode of transport they usually use to go shopping. In our main sample, around 86 per cent own a car; of those, about 11 per cent say they do not mainly use the car to shop. Of the 14 per cent who do not own a car, around 18 per cent say they mainly shop by car, presumably relying on taxis, friends or relatives to do so. To the extent that both those who shop by car and those who own a car potentially have lower search costs, we define the 'no car' group as only those that neither have a car nor shop by car. In our full sample, 88 per cent have access to a car.

equal). In our data, those who have access to a car shop less frequently than those who do not – an average of 6.6 times a month compared with 9.4 – and the difference is strongly statistically significant. Table 6.1 shows that those without access to a car for shopping come predominantly from the lowest income groups: around 30 per cent of those in the lowest income group do not have access, compared with 10 per cent or so of those with incomes between £10,000 and £30,000 and 5 per cent or so of those with higher incomes. Therefore, when we look at how the number of trips varies with income, we want to make sure we are not mixing 'opportunity cost of time' effects with 'access to car' effects and so we split the sample in this way.

Figure 6.1. Average monthly number of main shopping trips, by income group and access to car, 2006



Notes: Dotted lines indicate 95 per cent confidence intervals. Figures are averages by income group and car use of household-level averages calculated from monthly estimates. High-income no-car groups are omitted because of small sample sizes.

Table 6.1. Price index by income group and access to car, 2006

Income group	N	% no car	Price index	
			No car	Car
< £10k	2,196	29.7%	99.9	99.7
£10k–£20k	3,582	9.6%	100.1	99.7
£20k–£30k	3,615	10.1%	100.2	100.0
£30k–£40k	1,830	4.3%	100.5	100.1
£40k–£50k	711	1.8%	(100.4)	100.4
£50k–£60k	551	6.2%	(101.5)	100.7
£60k+	400	4.3%	(101.3)	100.7
All	12,885	11.7%	100.1	100.0

Note: Figures in parentheses have cell sizes of less than 50.

There is a steady, and statistically significant, decline in the average number of shopping trips made per month as equivalised income increases for both the car and no-car groups. From the search theory perspective, this suggests that, as income increases, the additional

opportunity costs of time outweigh the greater absolute savings from higher spending and so total search decreases.

For a given income group, the no-car households always make a higher average number of trips and it seems likely that this is partly driven by buying smaller amounts on each trip. We would not therefore expect making more trips to translate into lower prices for the no-car households in exactly the same way as for households with access to a car. Table 6.1 shows that this is the case. Conditional on car access, the price index increases with income group; but, for a given income, those in the no-car group pay slightly higher prices even though they make more trips on average. Again, these average differences are very small. The difference between the price index for the car and no-car groups for each income group is statistically significant for the first three income groups (after this, the sample size for the no-car group becomes very small). This confirms that, in a search model, we do not want to simply mix the car and no-car groups together since households in the no-car group pay higher prices and make more trips on average.

Table 6.2 presents the results of regression analyses of our most successful specification for the effect of the number of shopping trips made on the household price index. In its simplest form, the model we have in mind is as follows. Suppose the price a household pays, p_h, is a function of a base price level, $\bar{p}$, and of how hard it searches, taking the form of a percentage reduction, α, for each additional search. Then

$$p_h = \bar{p}(1-\alpha)^n$$

where n is the number of searches. Taking natural logarithms of each side, we get

$$\ln p_h = \ln\bar{p} + n\ln(1-\alpha)$$

$$\approx \ln\bar{p} - \alpha n$$

for small values of α. In our actual specification, we allow the percentage price reduction to vary with the number of trips by including a 'trips squared' term, which turns out to be statistically significant.

Table 6.2. Effect of search: overall and by population density

Dependent variable: log average price index	(1)	(2)	(3)	(4)	(5)	(6)
No car	0.287*	0.280*	0.691*	0.200	0.226*	0.401*
No. of main trips	–0.076*	–0.078*	0.053	–0.082*	–0.079*	–0.090*
No. of main trips squared	0.001*	0.001*	–0.004*	0.001	0.001*	0.001*
Population density	—	—	Lowest quartile	2nd quartile	3rd quartile	Highest quartile
N	17,990	15,479	1,640	3,660	5,334	4,845
R^2	0.010	0.010	0.007	0.011	0.009	0.015

* = significant at 5 per cent level; † = significant at 10 per cent level.
Note: Results in columns 2–6 include English households only.

Note that here and in all regressions in this chapter where the price index is the dependent variable, we multiply the log average price index by 100 so that the coefficients can be interpreted as (approximately) percentage effects – a coefficient of 1 represents a 1 per cent increase in prices.

Column 1 of Table 6.2 shows that additional search has a statistically significant negative effect on the price index, but that the returns to search decrease as the number of trips increases – as shown by the positive sign on the number of trips squared term. We looked at different specifications of the regressions and found that the percentage reduction associated with additional trips did not appear to vary by access to a car,[33] but the intercept of the regression did. So it seems that the proportional return to search is the same for both groups but that those without access to a car face a marginally higher general, or base, price level – around 0.3 per cent higher than for those with access to a car. The effect of one additional trip is small: at 10 trips per month, one additional trip is associated with a reduction in prices of around 0.05 per cent.

Population density

When interpreting the results in column 1 of Table 6.2 in the context of search, we might be worried that the association between trips and price in the data is partly to do with the following circumstances: in sparsely populated areas, the market cannot support many shops, stores are further apart and people make fewer trips, but because there is less competition there is also a higher general price level in these areas; this could at least partly drive the observed correlation between trips and prices.

To examine this, we look at whether the relationship between number of shopping trips and the price index holds if we divide our sample by population density and perform the analysis separately within each subgroup. We use data from the 2001 Census in England taken from Casweb,[34] which records for each English 'Super Output Area' the population density measured as number of people per hectare. As with the deprivation measure discussed in Chapter 5, we map this variable to the postcode district using ONS postcode data to give a density measure for the households in our data. We then define quartiles of population density; that is, we split the postcode sectors in our data into four groups based on their population density from least to most dense.[35] Sample sizes by density quartile and car access are shown in Table 6.3. Note that for the lower two density quartiles, the number of households without access to a car for shopping is very small (57 and 241 respectively) and so any results for these groups should be interpreted with caution.

[33] That is, if we interacted main trips and main trips squared with a dummy for households without access to a car, the interaction terms were not significant.

[34] http://casweb.mimas.ac.uk/.

[35] These quartiles are based on postcode districts, not households, so there will not be equal numbers of households in each quartile. The boundary points for the 2nd, 3rd and highest density quartiles are around 15.7, 29.0 and 42.5 people per hectare respectively.

Table 6.3. Use of car by population density, 2006

Density	Car	No car	Total
Lowest quartile	1,583 (97%)	57 (3%)	1,640 (100%)
2nd quartile	3,419 (93%)	241 (7%)	3,660 (100%)
3rd quartile	4,809 (90%)	525 (10%)	5,334 (100%)
Highest quartile	3,956 (82%)	889 (18%)	4,845 (100%)
All	12,267 (79%)	1,712 (11%)	15,479 (100%)

Note: Figures include English households only.

In Table 6.4, it can be seen that for areas of similar population density, those who do not have access to a car make (statistically significantly) more trips than those who do. There appears to be an increasing relationship between density and number of trips made, but the only statistically significant difference is between those with cars in the least dense quartile and those with cars in the other quartiles.

Table 6.4. Shopping trips by use of car and population density, 2006

Density	Car	No car	All
Lowest quartile	6.2	8.6	6.3
2nd quartile	6.4	9.1	6.6
3rd quartile	6.6	9.6	6.9
Highest quartile	6.5	9.5	7.1
All	6.8	9.4	6.8

Table 6.5 shows the average price index by access to a car and population density. The price index tends to decrease slightly with density. For those with cars, households in the least dense quartile pay statistically significantly higher prices than those in the other quartiles, and those in the second quartile pay statistically higher prices than those in the top two quartiles. For those without cars, the only statistically significant difference is that those in the lowest density quartile pay higher prices than those in the third quartile. Within density quartile, those without cars have higher price indices, though the difference is only significant for the densest quartile.

Table 6.5. Average price index by density quartile and access to car, 2006

Density	Car	No car	All
Lowest quartile	100.2	100.9	100.3
2nd quartile	100.1	100.2	100.1
3rd quartile	99.9	100.0	99.9
Highest quartile	99.9	100.2	100.0
All	99.9	100.1	100.0

Again, we can analyse the effect of population density using regression analysis. Column 2 in Table 6.2 above replicates column 1 for England, and the results are very similar to those for the British data. Next, we run a separate analysis by density quartile, reported in

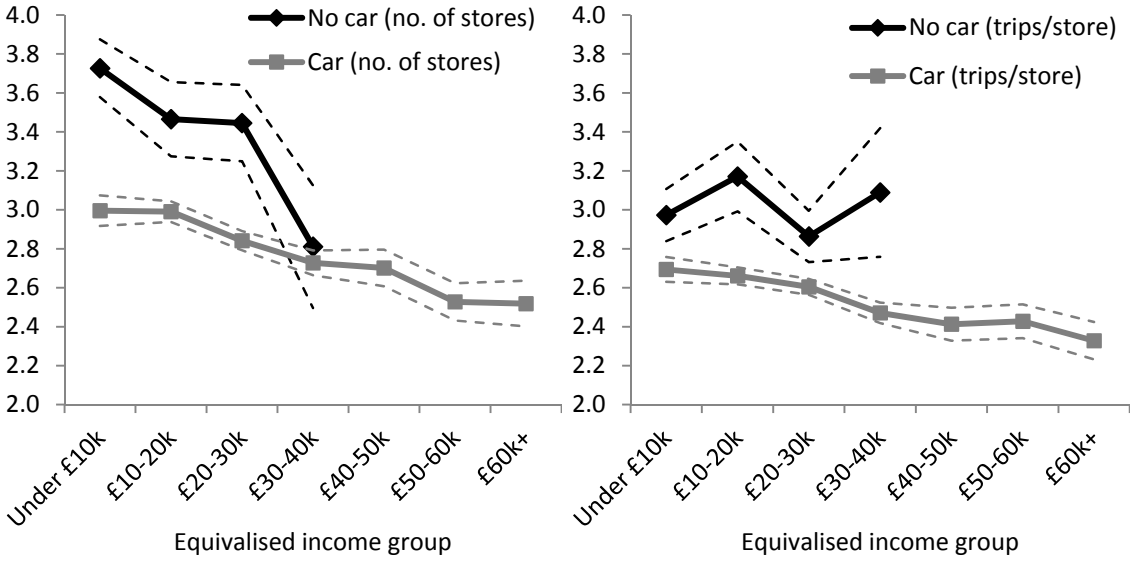
columns 3–6 from least to most dense quartile respectively. The same general pattern holds in each density quartile as for the whole sample. It does seem that the no-car households in the lowest population density quartile face the highest general price level and have the lowest returns to search, but the differences are still very small and the sample of no-car households is also very small, particularly in the least dense quartile.

It does not therefore appear that our finding that more trips are associated with lower prices is driven entirely by population density and local competition. For all areas, we still find statistically significant evidence that additional trips reduce prices paid.

6.1.2 Temporal search

Our previous results demonstrated that making more trips is associated with a slightly lower price index. This is consistent with the model of search. As discussed in Chapter 3, there are two ways in which prices can vary: spatially and temporally. Some shops may have generally lower prices than others, and shops will sometimes have temporary reductions in prices, or sales. Thus households can search for low prices across stores when they go shopping and over time both within and between stores. To look at this more closely, we decompose the total number of shopping trips a household makes each month into the number of different stores they visit and the average number of trips per store they make.[36]

Figure 6.2. Average monthly number of stores and trips per store, by income group and access to car, 2006



Notes: Dotted lines indicate 95 per cent confidence intervals. Figures are averages by income group and car use of household-level averages calculated from monthly estimates. High-income no-car groups are omitted because of small sample sizes.

[36] Again, we focus on 'main' shopping trips only for these definitions.

Table 6.6. Temporal search by population density

Dependent variable: log average price index	(1)	(2)	(3)	(4)	(5)	(6)
No car	0.287*	0.283*	0.640†	0.164	0.219*	0.407*
No. of main stores	–0.093*	–0.094*	0.142	–0.106	–0.104	–0.130†
No. of main stores squared	–0.019*	–0.020*	–0.045*	–0.022†	–0.017*	–0.017*
No. of trips/store	0.074*	0.074*	0.135*	0.063*	0.085*	0.056*
Population density	—	—	Lowest quartile	2nd quartile	3rd quartile	Highest quartile
N	17,990	15,479	1,640	3,660	5,334	4,845
R^2	0.028	0.029	0.021	0.030	0.026	0.034

* = significant at 5 per cent level; † = significant at 10 per cent level.
Note: Results in columns 2–6 include English households only.

Figure 6.2 shows that, for those with access to a car, both measures decrease with income group. However, the correlation between number of stores and trips per store is negative (both within income and car-access group and as a whole), suggesting that people who visit a higher number of stores than average also make fewer than average trips per store. We also see that those without access to a car tend to visit more shops and make more trips per store than the car group. Within the no-car group, the number of stores visited declines with income but the pattern for trips per store is less clear.

Table 6.6 shows estimates from various regressions of the log of the household's average price index on the number of stores and trips per store. We allow the marginal effect of the number of stores to vary by including a squared term, but could find no evidence that the relationship between the price index and trips per store was non-linear. Within the group totals shown in Figure 6.2, the variation in the number of stores visited and trips per store across population density quartile is very small, with the exception of the number of stores visited by the no-car group. Although the sample sizes are small in this group, the difference between the top and bottom density quartiles (3.6 stores and 3.1 stores respectively) is statistically significant.

Column 1 of Table 6.6 shows that for a given number of trips to each store, increasing the number of stores visited reduces the price index. The coefficient on the number of stores squared is also negative. This means that, unlike the decreasing returns to the total number of trips seen in Table 6.2, the marginal proportional reduction in the price index from visiting more stores increases with the number of stores visited. This would be consistent with the return to search being more like a fixed absolute reduction per search rather than a fixed proportional reduction. However, this must be weighed against the fact that the costs of visiting more stores are also likely to rise with the number visited (visiting stores further away, for example). In our data, only around 10 per cent of households visit an average of five or more stores in a given month for main shopping trips. Households that visit one store have a marginal saving of 0.15 per cent from an additional store and those that visit five stores have a marginal saving of 0.31 per cent, so the magnitude of the differences is small across the vast bulk of the households in our data.

Interestingly, there is a positive relationship between the number of trips made per store (keeping the number of stores visited constant) and the price index. The effect is again quite small – making an additional two trips per store increases prices by around 0.15 per cent – but is statistically significant. This says that, given a certain number of stores visited, households that shop more frequently tend to pay a higher price. This specification controls for access to a car (as in Table 6.2, there is no improvement in fit by allowing the effect of stores or trips per store to vary by access to car) and again suggests that those without cars have base prices around 0.3 per cent higher than those with cars.

Column 2 of Table 6.6 replicates the results for England only, and columns 3–6 show that the main relationships continue to hold across density quartile: the marginal proportional reduction in the price index increases with the number of stores visited; those without cars have higher base prices; and making additional trips per store increases prices paid. If anything, those in the least densely populated areas appear to lose out more from additional trips per store. There was no evidence of a non-linear relationship between prices and trips per store in any density quartile.

The positive relationship between price and trips per store is of note, and could be consistent with a model of search where consumers with low storage costs stockpile non-perishable goods when they find them on sale. More specifically, imagine a group of consumers who have plenty of space to store goods and another group who have limited storage space and/or cannot transport a large shop. The first group has a larger incentive to search hard for a low price and then stock up. Therefore, compared with the second group, we might find that members of the first group search harder when they go shopping, but leave a greater interval between trips because of their stockpiles. If we now think of the monthly number of stores as being a measure of how intensely the household searches, and the number of trips per store as being their shopping frequency, this could explain the negative correlation between the number of stores and trips per store in the data and why more trips per store (i.e. less stockpiling) is associated with higher prices given the number of stores visited.

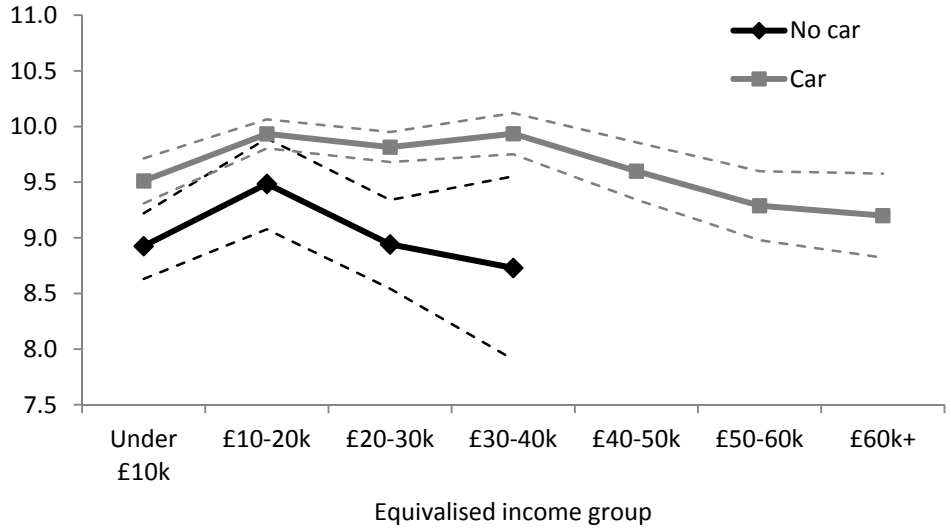
It could also explain why stores visited and trips per store both decline with income. We might expect that house size would increase, and hence storage costs decline, as income increases, enabling households to stockpile and make fewer trips per store. We might then expect richer households to visit more stores to search for low prices when building their stocks. However, as income rises, the opportunity cost of search also increases, acting to reduce search effort along both margins.

This is also consistent with the use of sales and promotions across income groups. Figure 6.3 shows an estimate of how much households in different income groups save on average from buying food products when they go on sale, given their access to a car. Using the data, which record the type of sale used, we estimate how much the sale items would have cost in the absence of the sale[37] and thus how much the household would have spent

[37] In the case of a price discount, this is simple, as we just add the price discount to the sale price. For quantity discounts, we assume the unit price would remain unchanged in the absence of the sale (for example, a 'buy one, get one free' would cost twice as much without the sale). For other types of sales, such as unrelated multibuys or 'free gift' style promotions, we cannot estimate a saving. Note that this measure assumes that the household would have bought what it did in the absence of the sale, which may well not be the case; this estimate should probably therefore be interpreted as an upper bound of the savings from sales.

without sales. This saving is expressed as a percentage of the 'non-sale' expenditure. Across income groups, savings from sales are hump-shaped – first increasing with income, then peaking at middle incomes and decreasing again at high incomes. This could be because at first, as storage costs decrease and spending increases, the benefit of searching for sale prices increases, but then at higher incomes the opportunity cost of time starts to outweigh the benefits of search and people reduce their search effort for sales again. The figure also shows that those without access to a car make smaller savings from sales than those with a car, which is consistent with them being unable to transport large volumes of sale items for stockpiling.

Figure 6.3. Average monthly percentage saving from sales, by income group and access to car, 2006



Notes: Dotted lines indicate 95 per cent confidence intervals. Figures are averages by income group and car use of household-level averages calculated from monthly estimates. High-income no-car groups are omitted because of small sample sizes.

Table 6.7. Temporal search by population density including sales

Dependent variable: log average price index	**All households (England)**	**Least dense**	**2nd quartile**	**3rd quartile**	**Most dense**
No car	0.047	0.468†	–0.077	–0.015	0.177*
No. of main stores	–0.056	–0.049	–0.051	–0.028	–0.088
No. of main stores squared	–0.023*	–0.020	–0.028*	–0.023*	–0.020*
No. of trips/store	–0.025†	–0.002	–0.028	–0.028	–0.025
Saving from sales	–0.322*	–0.356*	–0.327*	–0.317*	–0.311*
N	15,479	1,640	3,660	5,334	4,845
R^2	0.296	0.333	0.321	0.287	0.280

* = significant at 5 per cent level; † = significant at 10 per cent level.
Note: Results include English households only.

Table 6.7 replicates the model of Table 6.6 but includes the average share of expenditure saved through sales, which tries to control for the hypothesised relationship between shopping frequency and use of sales. We no longer find evidence that more trips per store

increases prices paid. Sales have a large negative impact on prices paid: increasing the savings from sales by 1 per cent reduces the price index by around 0.3 percentage points. This result suggests that conditional on the number of stores visited and the use of sales, households that shop more frequently do not pay higher prices. However, including sales in the model does not eliminate the negative impact on price of the number of stores, suggesting that search is not simply related to the ability to find and exploit promotions but is also to do with finding lower everyday non-sale prices. There is still some evidence that even controlling for use of sales, those without cars pay slightly higher base prices, particularly in the most and least densely populated regions.

In Table 6.8, we show the association between the number of trips per store and the overall savings the household makes by buying goods on sale. The dependent variable is the proportion of the household budget saved by buying goods on sale. There is a strong, significantly negative relationship between trips per store and share saved from sales, though the marginal impact of additional trips declines, as shown by the positive coefficient on trips per store squared. People who visit more stores tend to save slightly more from sales, though the effect is not consistent across different density quartiles. Those without access to a car tend to save less from sales: conditional on the number of stores visited and trips made to each store, not having access to a car reduces the savings from sales by around 0.7 per cent.

Table 6.8. Effect of search behaviours on savings from sales

Dependent variable: proportion of budget saved from sales	**All households (England)**	**Least dense**	**2nd quartile**	**3rd quartile**	**Most dense**
No car	–0.735*	–0.508	–0.737*	–0.724*	–0.741*
No. of main stores	0.055*	–0.016	0.035	0.085*	0.059†
No. of trips/store	–0.455*	–0.846*	–0.416*	–0.501*	–0.420*
No. of trips/store squared	0.016*	0.063*	0.015	0.014*	0.016†
N	15,479	1,640	3,660	5,334	4,845
R^2	0.016	0.015	0.013	0.021	0.016

* = significant at 5 per cent level; † = significant at 10 per cent level.
Note: Results include English households only.

If search and stocking up are important drivers of price, we might expect the impact of the number of stores visited and the number of trips made to vary across *goods* according to how perishable they are. To investigate this, we calculated good-specific household-level price indices using a breakdown of products supplied by Kantar into market-specific categories such as bread, alcohol and milk. In total, there are 198 of these categories for food and alcohol. We calculated a household-level price index for each product category using data from the whole calendar year 2006.[38] In our data, we do not observe a household visiting a shop and buying nothing; similarly, we do not observe the prices of

[38] We chose to calculate the index using the full year of spending information rather than taking an average index over each month because a household's spending on a given product group in a single month was often very low, which led to large variations in the price index from month to month. The index was calculated in precisely the same way as the own-basket index, comparing how much a household paid for the products it bought in each category over the year with how much it would have paid had it purchased those products at national average prices over the year.

goods that were considered but ultimately not purchased on any given shopping trip. While we might reasonably expect households not to make many grocery trips where they buy nothing, the same can probably not be said for purchases of individual goods. Thus using observed stores visited and trips per store is probably a good measure of general search for the whole basket but not for an individual good. For simplicity, we look at the annual frequency of purchase for individual goods – if the good is bought on sale and stockpiled, we would expect the household to make fewer annual purchases and pay a lower price.

Limiting our sample for each category to households that we observed purchasing something (from any category) in each month during 2006 and that made at least four trips in which they purchased from that category over the year (thus eliminating non-regular consumers of the good who try it once, perhaps because they see it on offer), we regress the log of this price index on the annual purchase frequency. We also regress the share purchased on sale of the good by each household on annual purchase frequency.

Our results lend some support to the idea that more storable goods are those that see the largest positive effects of purchase frequency on the price paid. In the regression of the price index on purchase frequency, the only category to have a negative relationship is fresh fruit and vegetables (the coefficient is small but significant), which are not storable for long periods. Goods with small positive effects of purchase frequency on price include milk and bread, which are also perishable, whilst those with the largest positive coefficients include highly storable categories such as tinned and ambient puddings, vinegar, syrup and treacle, and couscous. When we regress the share bought on sale on purchase frequency, there is a negative relationship for all categories, as would be expected. Goods with small coefficients are again non-storable items such as fresh fruit and vegetables, bread and milk, whilst those with the largest (absolute) coefficients are syrup and treacle, salt, mustard, tinned and ambient puddings, vinegar and couscous.

6.1.3 Bulk discounting

Households could make savings not just from buying goods on sale, but also from making bulk purchases with lower per-unit prices. We might expect that those households that make savings from stocking up during temporary price reductions are also those that make greatest use of bulk discounts if at least part of the reason for buying in bulk is to stockpile.

There seems to be some evidence that this is the case. Table 6.9 shows the relationship between the change in the household-level price index when we redefine products using our '*packsize*' definition (see Section 4.2.2) and household shopping behaviours. The results are consistent with the evidence in Table 6.6 above that households who make more trips per store pay higher prices, which we suggested may be due to less use of stockpiling when goods were on sale. Here, we find that households making more trips per store lose out from bulk discounting in the sense that their price index increases when we group together similar products of different pack sizes and recompute the index. Those who make more trips per store appear to buy smaller packages and so are not able to exploit the gains from bulk discounts. Adding in the saving from sales suggests that

households that make more use of sales also benefit more from bulk discounts,[39] but that there is still a strong effect of additional trips per store. Table 6.7 showed that, conditional on sales, we could not find strong evidence that those who shopped more frequently paid higher prices, but in Table 6.9 we find evidence that they still lose out from buying smaller packages. Interestingly, there does not appear to be any relationship between the number of stores visited and the gains from bulk discounting. Households do not appear to 'search' for bulk discounts.

Table 6.9. Effect of shopping behaviours on gains from bulk discounting

Dependent variable: change in index (moving from barcode *to* packsize *product definition)*	**(1)**	**(2)**
No. of main stores	–0.001	0.004
No. of trips/store	0.089*	0.059*
Saving from sales	—	–0.088*
N	17,990	17,990
R^2	0.004	0.005

* = significant at 5 per cent level; † = significant at 10 per cent level. Positive values mean that the effect is to increase the price index once we group products using the *packsize* definition rather than the *barcode* definition.

6.2 Store choice

In Chapter 3, we suggested that one reason why prices may vary is that different households shop in different types of store, and some households may be willing to trade off store quality and product prices. Furthermore, if household preferences for different retailers are correlated with their price responsiveness, then variation in prices across stores may not only reflect different 'service' costs but also the ability of different stores to price-discriminate.

Is there evidence in our data that different types of store charge very different prices for identical products? To examine this, we constructed our group-level price indices (Laspeyres, CPD and EKS) for different stores using the June 2006 spending data and the '*ownbrand*' product definition. Because different stores sell their own brand of many products, if we had restricted attention to products defined as bar codes bought in all types of stores we would have found very little overlap across stores (and any overlap would have covered branded goods only).[40] We group together different stores into a number of relatively broad categories given in Table 6.10. 'Discounters' include chains such as Aldi, Lidl and Netto as well as freezer centres. 'Large' stores are the large-format versions such as Tesco Extra as well as cash-and-carry supermarkets. 'Quality' stores are those chains known to charge higher prices but that promise a better 'shopping experience'. 'Stores 1, 2 and 3' are standard formats of three large, national supermarket chains – of these, store 2 generally charges the highest prices and store 3 usually the

[39] There may well be overlap between sales and bulk discounts since promotions are often of the 'extra volume for the same price' variety.

[40] In particular, some store types in our data (particularly some of the quality retailers as well as the discounter supermarkets) sell essentially only their own-brand products so we would have expected virtually zero overlap of products across store at the bar-code level.

lowest (though on the CPD price index, store 1 charges the same prices as store 3). 'Local' stores are defined as corner shops, high-street retailers, speciality stores (such as butcher and baker shops) and so on. 'All others' includes other national supermarkets, as well as 'metro' formats of national retailers, home delivery and online shopping.

Table 6.10. '*Ownbrand*' price indices by store type, June 2006

	Laspeyres	**CPD**	**EKS**
Local	110.74	100.00	102.44
Discounter	103.14	93.23	95.26
Large	100.75	92.25	97.17
Quality	119.94	113.66	115.77
Store 1	101.05	92.49	97.09
Store 2	105.11	97.28	100.64
Store 3	100.86	92.54	96.41
All others	102.20	94.27	98.01
Range	*19.19*	*21.41*	*20.51*
Overlap	17.8%		

The results show considerable variation in price indices across store types but largely match intuition about relative prices in different types of stores. Prices tend to be highest in local shops and quality supermarkets and lowest in large-format stores, discounters and supermarkets. Note that even having relaxed our notion of a product by grouping together different stores' own brands, the overlap of purchases across the different store types is still quite small, at less than a fifth of total expenditures, which might account for the relatively large variation in price indices.[41]

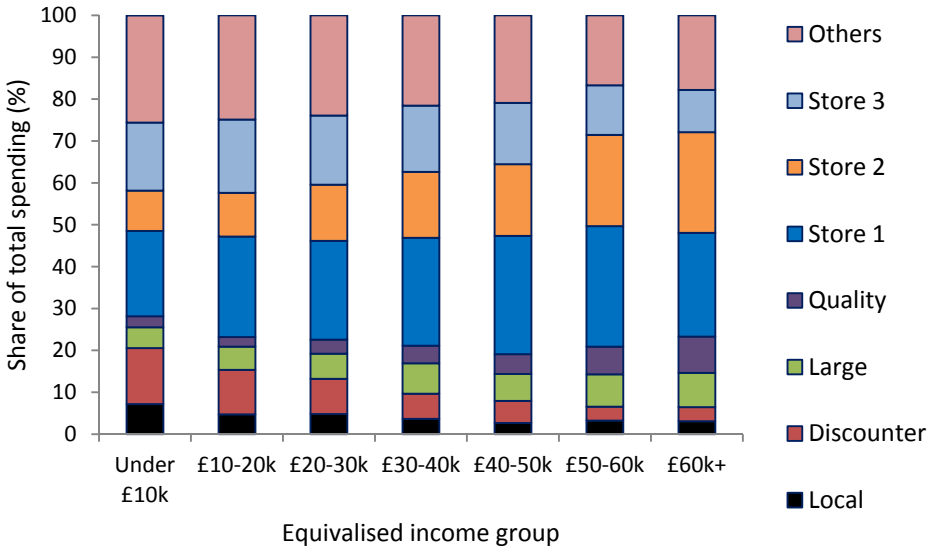
Our finding that higher-income households pay slightly higher prices on average could partly be driven by them shopping in more expensive stores. Figure 6.4 breaks down the average share of spending by store type for different equivalised income groups. The results lend some support to the idea that some of the variation in prices is driven by store choice: richer households on average pay higher prices and spend less in discounters and the cheapest of the national supermarkets, and more in the quality supermarkets and more expensive national supermarket. However, not all the relationships between income and store choice point in this direction: poorer households spend more in (expensive) local stores and less in (cheaper) large-format stores.

Indeed, store choice alone does *not* eliminate the relationship between income and prices. If we run a regression of the household's average log price index on income and store choice, then the variation in prices across income groups is reduced but not eliminated. Households in the richest income groups (over £50,000) are still estimated to have a price

[41] Note that all the Laspeyres indices are above 100. This is perfectly consistent with the definition of the Laspeyres index, which is the ratio of the 'average' basket at stores' own prices to the 'average' basket at national average prices. Since stores will typically sell a larger proportion of goods that they sell relatively cheaply, it is possible for all stores to sell the national average basket at above national average prices.

index around 0.5 percentage points higher than those in the poorest income group (below £10,000) even conditional on store choice. Moreover, income and store choice together still only explain about 8 per cent of the price variation we see across households.[42]

Figure 6.4. Average share of total spending by store type and income group, 2006



Note: Figures are averages by income group of household-level averages calculated from monthly estimates.

Table 6.11. Effect of store choice on search and frequency

	No. of stores	No. of trips/store
Local %	0.058*	–0.001
Discounter %	0.023*	–0.005*
Large %	–0.003*	–0.002*
Quality %	0.013*	0.003*
Store 1 %	–0.004*	0.001†
Store 2 %	–0.002*	0.001*
Store 3 %	–0.005*	0.000
N	17,990	17,990
R^2	0.267	0.009

* = significant at 5 per cent level; † = significant at 10 per cent level.
Note: Independent variables are shares of total household spending in different store types (spending in 'other' stores excluded from the regression); dependent variable is given in column header.

As discussed in Chapter 3, different stores may wish to use different pricing strategies (for example, HiLo or EDLP) to attract different types of consumer who like to search more or less. To this extent, we would expect there to be a considerable correlation between search variables (for example, numbers of stores and trips) and store choice. Table 6.11 shows this to be the case, particularly for search as measured by the number of stores visited. We regress both the number of stores and number of trips per store on the percentage of households' spending in the different store types defined above. Store

[42] Full results of the regression are available from the authors on request.

choice alone can explain more than a quarter of the variation in the number of stores visited. Households using national supermarkets or large-format stores visit fewer stores, whilst those visiting discounters, local shops and quality retailers visit more. This is consistent with the idea that households that mainly use supermarkets largely engage in one-stop shopping whilst those that use other types of store more may be splitting their shopping across a range of stores which might be indicative of search. Store choice can explain far less of how many trips households make per store, which we interpreted before as shopping frequency.

CHAPTER 7
Conclusions

Using detailed data on the grocery purchases of thousands of households, this research has assessed whether there is evidence that low-income and other potentially vulnerable household groups pay higher prices for identical food products. It has also empirically investigated various reasons as to why prices paid differ across households. Prices paid for individual goods vary substantially across both time and space in our data. However, once we construct a household's price index to measure overall price dispersion, we find that, on average, the differences in prices paid across groups are very small. In common with almost all other broad-based studies that have looked at this issue (which have mostly used US data), we find no evidence that 'the poor pay more' for food. If anything, our results suggest that poor households pay slightly lower prices than richer households. Using data from 2006, for example, we estimate that households with an equivalised income below £10,000 paid around 0.9 per cent less on average for food in the home than those with incomes over £60,000, a small but strongly statistically significant difference.

Strikingly, we find that the variation in monthly household-level price indices is, if anything, slightly larger over time for a given household than it is across households of the same income group. We find that variation across households is larger for the poorest households. Overall, it seems that poorer households pay, on average, slightly lower prices than richer households but with a more dispersed distribution.

Other household demographics that appear to be correlated with paying lower prices are the following:

- the presence of children in the household: households with one, two or more children pay, on average, 0.15 per cent, 0.41 per cent and 0.75 per cent less than those with no children respectively;

- not being over state pension age: compared with households headed by someone aged 25–29, those aged 70–74 pay 0.25 per cent more on average, those aged 75–79 pay 0.47 per cent more and those aged 80+ pay 0.98 per cent more; differences for all other age groups are statistically indistinguishable from zero;

- being a lone parent or having three or more adults in the household: compared with single adults, these groups pay 0.37 per cent less and 0.23 per cent less on average. Households of three or more adults with children pay even less, with an average price index 0.72 percentage points lower than that for single adults.

Unlike previous work, we demonstrate that our conclusions as to which groups pay higher or lower prices are robust to the precise way in which prices are compared across different demographic groups and that they do not seem to vary over time.

Another unique contribution of our work is to explore whether our results change when we re-define what is meant by 'identical products'. A focus on products at the bar-code level (the same brand, package size and so on) could miss important ways in which

households are able to economise, including bulk discounting and buying cheaper store own-brand products. Once we allow for these behaviours, our results show that there is more variation in average prices paid across groups and that it is poorer households that benefit most. The median reduction in price index for households with incomes below £10,000 when bulk discounting is taken into account is around 0.2 percentage points, whilst those with incomes over £60,000 see a median increase of around 0.2 percentage points. Grouping together stores' own brands yields even bigger effects, with a median fall for the poorest group of 0.5 percentage points and a median increase for the richest group of a similar magnitude. The impact of bulk discounts and buying cheaper own brands on prices paid by other demographic groups also tends to reinforce rather than mitigate the differences seen when products are defined as unique bar codes.

Using a household-level price index that is amenable to regression analysis, we investigate different reasons derived from economic theory as to why prices could vary. Some of our findings are as follows:

- Households that search more are able to obtain slightly lower prices. Measuring search by the average number of main shopping trips made per month and conditioning on whether or not the household uses a car, we find that making an additional trip reduces the average price index by around 0.05 percentage points (given a baseline of 10 trips). This does not appear to be driven entirely by local population density.

- Search can be carried out both over time and across stores. Given the number of stores visited, households that shop more frequently (make more trips per store) pay slightly higher prices: an additional two trips per store increases the price index on average by around 0.15 percentage points. This is consistent with high-frequency households being less willing or able to stock up when prices are low. Once we condition on how much households save from buying on sale, this effect disappears, which also lends weight to the stockpiling result. However, even controlling for sales, households that visit more stores still pay lower prices.

- The use of sales is, unsurprisingly, strongly related to paying lower prices. Households that save an additional 1 per cent of their shopping budget from sales pay around 0.3 per cent lower prices on average. Households without cars are less able to exploit sales and, on average, pay slightly higher prices than those with cars.

- There is considerable price variation across different stores (though relatively little overlap in the goods that are common to them, even when we try to account for different stores' own brands). Lower prices are available in large-format and discounter stores, whilst local stores and 'quality' retailers charge higher prices. The relationship between income, store choice and prices is not clear-cut. High-income households spend a larger share of their total grocery budget in quality retailers, but low-income households spend more on average in both expensive local stores and cheap discounter stores.

- Once we control for store choice, we still find that those in the richest income groups (over £50,000) pay 0.5 per cent more than those in the poorest income group (below £10,000) on average.

- In general, although we find (sometimes strongly) statistically significant effects on price of search, store choice and so on, they do not explain much of the overall variation in prices. Adding in the use of sales improves the explanatory power of our results quite substantially.

The fact that we find, on average, very little variation in food prices across groups does not mean that there are no households that pay substantially more for food. Looking at monthly household-level price indices averaged over a calendar year, households at the 95th percentile of the price distribution pay prices around 7.5 per cent higher than those at the 5th percentile of the distribution. This suggests that some households could make sizeable savings were they able to buy at average prices. What our results show is that individual demographic variables such as income are not able to explain much of this variation in prices. It may be that if we had sufficient data to identify precisely which these households were, we would find some consistent patterns. For example, in our research, we study the effects of income, car ownership, population density and local area deprivation. It may be that there are groups of low-income or elderly households in rural, deprived areas without access to a car that do indeed pay significantly more, but we do not have sufficient observations (even in our very large data set) of such households to be sure. As best we can tell from the data, though, when we do cut our sample in many ways like this we still cannot find much variation in average prices.

Although we find some evidence of variation in food prices across households, the fact that, on average, the variation is quite small does offer some reassurance for policymakers. Typically, for example, when calculating national poverty lines or comparing the living standards of different groups in different periods, a common price index across households is assumed. This is problematic when rich and poor households have different expenditure shares for different spending subcategories and/or when they pay different prices for identical goods. We address the latter concern in this research[43] and find, for food at least (which is still a very large component of total spending for poor households), that this assumption may not be too problematic. Of course, this does not rule out the issue for other important components of spending, such as housing and transport, and it would be interesting to examine empirically, using existing or new data, how much variation in prices exists for these spending groups as well.

[43] Leicester, O'Dea and Oldfield (2008) discuss variation in inflation due to household budget share differences.

APPENDIX A
Calculating Price Indices

One way in which we can define a price index is to take some 'fixed basket' of items and compare the price paid for that basket by different households. The advantage of this approach is that variation in the resulting price index reflects only variation in prices paid. The main problem is that not all households will buy every item in the basket, so the price for some items will be missing and we need to find some way to deal with this. It is also not clear whether taking some essentially arbitrary fixed basket of items is appropriate since no one household is likely to buy precisely that basket.

Suppose we have two households, *h* and *i*, that purchase a basket of goods $\boldsymbol{q}$ at prices $\boldsymbol{p}$, and both prices and quantities vary across households. Define P^{hi} as the price index comparing the prices paid by household *h* with those paid by household *i*. The two most straightforward indices to calculate (aside from the missing prices problem) are the *Laspeyres* index,

$$P_L^{hi} = \frac{\boldsymbol{p}^h \cdot \boldsymbol{q}^i}{\boldsymbol{p}^i \cdot \boldsymbol{q}^i},$$

which asks how much household *h* pays for household *i*'s basket compared with how much household *i* paid for it, and the *Paasche* index,

$$P_P^{hi} = \frac{\boldsymbol{p}^h \cdot \boldsymbol{q}^h}{\boldsymbol{p}^i \cdot \boldsymbol{q}^h},$$

which asks how much household *h* paid for its own basket compared with how much household *i* paid for it.

Both of these indices suffer from a *substitution bias*. If the price of an item in household *i*'s basket increases, the household will consume less of it. This means $\boldsymbol{q}^i$ will be low for items that *i* pays more for than *h*, biasing P_L^{hi} downwards and P_P^{hi} upwards compared with a true 'cost of living' index. The *Fisher* price index,

$$P_F^{hi} = \sqrt{P_L^{hi} P_P^{hi}},$$

is the geometric mean of the Laspeyres and Paasche indices, and is a 'superlative' index in that it does not suffer this substitution bias.

Our data contain many thousands of households, and so making bilateral comparisons like this does not tell us about which households pay more. We require a single index that summarises all the possible bilateral comparisons we could make. We make use of results from the multilateral price index literature on how to compare price levels across many different countries and apply them to comparisons across many households. The first

index we examine is the **EKS index** (Elteto and Koves, 1964; Szulc, 1964),[44] which writes the price index for household *h* as the geometric mean of all the possible bilateral Fisher price indices between that household and the other households in the data:

$$P_{EKS}^{h} = \prod_{i=1}^{H} \left(P_{F}^{hi}\right)^{\frac{1}{H}}.$$

An alternative approach is not to make direct bilateral comparisons between households *h* and *i* but instead to make an indirect comparison via some 'central' household, *X*. This gives rise to an alternative view of the Laspeyres index:

$$P_{L(X)}^{hi} = \frac{\boldsymbol{p}^{i} \cdot \boldsymbol{q}^{X}}{\boldsymbol{p}^{X} \cdot \boldsymbol{q}^{X}} \Bigg/ \frac{\boldsymbol{p}^{h} \cdot \boldsymbol{q}^{\boldsymbol{X}}}{\boldsymbol{p}^{X} \cdot \boldsymbol{q}^{\boldsymbol{X}}} = \frac{\boldsymbol{p}^{i} \cdot \boldsymbol{q}^{X}}{\boldsymbol{p}^{h} \cdot \boldsymbol{q}^{\boldsymbol{X}}}$$

where the prices paid by households *h* and *i* for household *X*'s basket are compared; and of the Paasche index:

$$P_{P(X)}^{hi} = \frac{\boldsymbol{p}^{i} \cdot \boldsymbol{q}^{\boldsymbol{i}}}{\boldsymbol{p}^{X} \cdot \boldsymbol{q}^{\boldsymbol{i}}} \Bigg/ \frac{\boldsymbol{p}^{h} \cdot \boldsymbol{q}^{\boldsymbol{h}}}{\boldsymbol{p}^{X} \cdot \boldsymbol{q}^{\boldsymbol{h}}}$$

where the prices paid by household *X* for the baskets purchased by *h* and *i* are compared. A form of substitution bias (known as the *Gerschenkron effect*) also occurs for these indices, but, unlike for the simple bilateral comparison, the direction of the bias is unclear and will depend on whether $\boldsymbol{q}^{X}$ is more similar to $\boldsymbol{q}^{h}$ or $\boldsymbol{q}^{i}$. The second index we examine is this **Laspeyres** index given by comparing the price paid by different households for some 'national average' basket *X* to the national average price paid for that basket. In other words, our 'central' household's basket is taken to be the national average amount of each item purchased:

$$P_{L(X)}^{h} = \frac{\boldsymbol{p}^{h} \cdot \boldsymbol{q}^{X}}{\boldsymbol{p}^{X} \cdot \boldsymbol{q}^{\boldsymbol{X}}}.$$

A third index we consider that derives from the cross-national price comparison literature is the **weighted Country–Product Dummy (CPD)** index. This method effectively derives price indices using statistical inference. The (log) of the price of each good in each household is regressed on a set of dummy variables for each product and each household using weighted least squares, where the weight is the average budget share of product *j* across all households. Diewert (2005) shows that this approach gives coefficients on the dummy variable for each household that are the logs of that household's price index, P_{CPD}^{h}, where the resulting indices are Tornqvist indices (another superlative price index).

[44] The EKS approach is the main one used by OECD–Eurostat for calculating purchasing power parity (PPP) price relativities between countries to make GDP comparisons. See http://www.oecd.org/dataoecd/59/10/37984252.pdf.

All of these cases suffer potential missing prices problems. The index can be derived, in the case of the EKS and Laspeyres indices, only for products that are purchased by every household and, in the case of the CPD, only for products that are purchased by at least two households. Essentially, there are no products in our data that are bought by every single household. Additionally, the CPD index suffers from an enormous number of dummy variables should we wish to calculate a price index for every household (there are around 15,000 households), which makes computation very difficult. To resolve these problems somewhat, instead of looking at individual households we assign households to a number of demographic groups (based on income, age, household composition and so on) and assume that all households within a group pay the same average price for each product. In effect, this is a form of imputation of missing prices: we assume that, for example, a lone parent who we did not observe buying product *j* in our data period would have paid the average price paid by other lone parents for that product had they done so. This significantly reduces the missing prices problem, as now the EKS and Laspeyres indices can be calculated for products purchased by every *group* rather than every household. It also makes computation of the CPD index feasible, since we have dummy variables for each household group (and the regression weights are the average of each group's budget share for each product).

This approach suffers two drawbacks: first, our assumption that the missing prices can be imputed as the average group price may be a poor one; and second, we obtain a price index for each household group rather than each household. This means that we cannot really explore how different observable characteristics affect price since we can only derive indices for one group at a time (income groups, age groups and so on). We could, in principle, define quite narrow groups (single parents with incomes below £10,000 living in the north-west who shop on foot, say), but the larger the number of groups, the smaller the subset of products that are purchased by all of them.

A final index that we consider does not suffer from this missing prices problem because it compares the price paid by each household for its own basket with the national average price paid for that basket. It is just the household-level version of the Paasche index above:

$$P^h_{P(X)} = \frac{\boldsymbol{p}^h \cdot \boldsymbol{q}^h}{\boldsymbol{p}^X \cdot \boldsymbol{q}^h}.$$

This index essentially asks 'How much do households gain or lose buying their own basket at their own price compared with the average price?'. The benefit of this is that we obtain one price index for each household which is then amenable to multivariate regression analysis. The cost is that indices vary along $\boldsymbol{q}$ as well as $\boldsymbol{p}$ now. To take an extreme example, imagine that all households purchased unique products that no other household buys. Then every household will have a price index of 1 since the 'average' price paid by all households is exactly the same as the household price.

APPENDIX B
Results for Other Groups

Households with children

Table B.1. Price indices by number of children in household, June 2006

No. of children	Laspeyres	CPD	EKS	Own-basket		N
				All goods	Overlap	
0	100.26	100.00	100.34	100.20	100.08	12,198
1	100.06	99.80	100.14	100.01	99.81	2,743
2	99.97	99.62	99.97	99.81	99.60	2,656
3+	99.65	99.19	99.55	99.34	99.11	1,153
Range	*0.61*	*0.81*	*0.79*	*0.86*	*0.97*	
Overlap	78.4%					

Table B.2. Sample sizes by number of children and income group, 2006

	Non-poor	Poor	Total	% poor
No children	4,963	3,605	8,568	42.1
1 child	1,090	756	1,846	41.0
2 children	816	922	1,738	53.0
3 or more children	238	495	733	67.5
All households	7,107	5,778	12,885	44.8

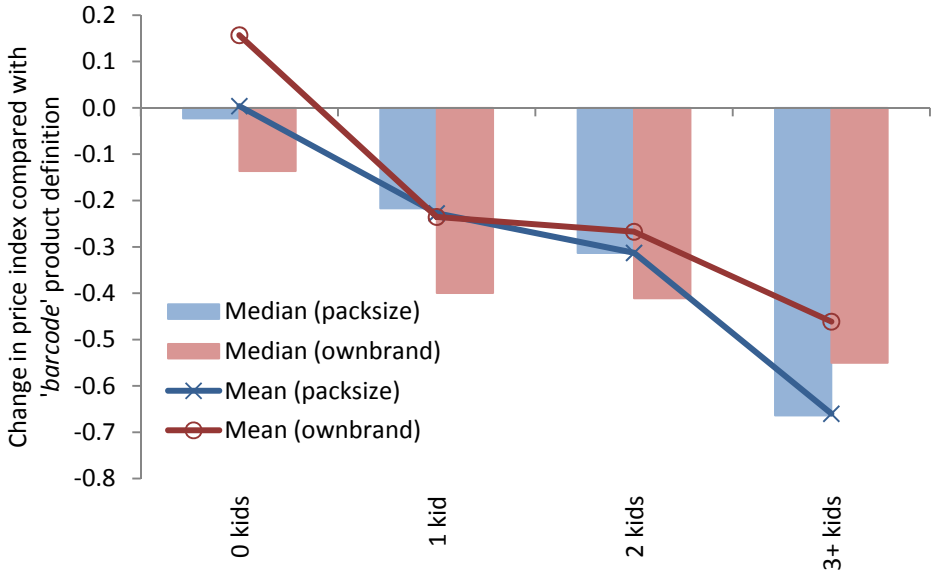
Note: Poor households are those with equivalised income below £20,000.

Table B.3. Variation in average price index by number of children

Excluded: 0 kids	β	se	t	*Excluded: 0 kids, non-poor*	β	se	t
1 kid	–0.146	0.060	–2.45	1 kid	–0.081	0.078	–1.05
2 kids	–0.413	0.061	–6.75	2 kids	–0.374	0.088	–4.27
3+ kids	–0.747	0.090	–8.35	3+ kids	–0.466	0.154	–3.03
				poor	–0.289	0.051	–5.69
				1 kid × poor	–0.166	0.121	–1.37
				2 kids × poor	–0.014	0.123	–0.11
				3+ kids × poor	–0.307	0.190	–1.62
N	12,885			N	12,885		
R^2	0.008			R^2	0.013		

Notes: Dependent variable is log of household average price index multiplied by 100. Coefficients can be interpreted as approximate percentage difference compared with the excluded group. Poor households are those with equivalised income below £20,000.

Figure B.1. Average change in price index relative to '*barcode*' product definition, by number of children



Note: Negative figure indicates group gains on average from switch in product definition (price index decreases).

Age group of household head

Table B.4. Price indices by age group of household head, June 2006

Age	Laspeyres	CPD	EKS	Own-basket		N
				All goods	Overlap	
Under 25	101.22	100.00	100.27	100.33	100.04	333
25–29	100.05	99.41	99.54	99.91	99.69	1,402
30–34	99.92	99.42	99.39	99.92	99.51	1,967
35–39	99.86	99.34	99.33	99.81	99.45	2,160
40–44	99.85	99.40	99.37	99.87	99.52	1,744
45–49	99.80	99.23	99.33	99.86	99.56	1,650
50–54	99.91	99.36	99.44	99.89	99.56	1,907
55–59	99.98	99.40	99.53	99.94	99.67	2,005
60–64	100.54	99.79	100.05	100.25	100.13	1,657
65–69	100.72	99.86	100.17	100.22	100.16	1,367
70–74	101.04	100.14	100.42	100.45	100.39	1,327
75–79	101.92	100.53	101.15	100.69	100.98	774
80+	103.02	101.51	102.09	101.57	101.82	457
Range	*3.22*	*2.28*	*2.76*	*1.76*	*2.37*	
Overlap	35.2%					

Table B.5. Sample sizes by age and income group, 2006

Age	Non-poor	Poor	Total	% poor
Under 25	73	71	144	49.3
25–29	575	349	924	37.8
30–34	823	471	1,294	36.4
35–39	879	598	1,477	40.5
40–44	729	473	1,202	39.4
45–49	693	453	1,146	39.5
50–54	800	532	1,332	39.9
55–59	803	596	1,399	42.6
60–64	624	554	1,178	47.0
65–69	411	594	1,005	59.1
70–74	389	582	971	59.9
75–79	205	325	530	61.3
80+	103	180	283	63.6
All households	7,107	5,778	12,885	44.8

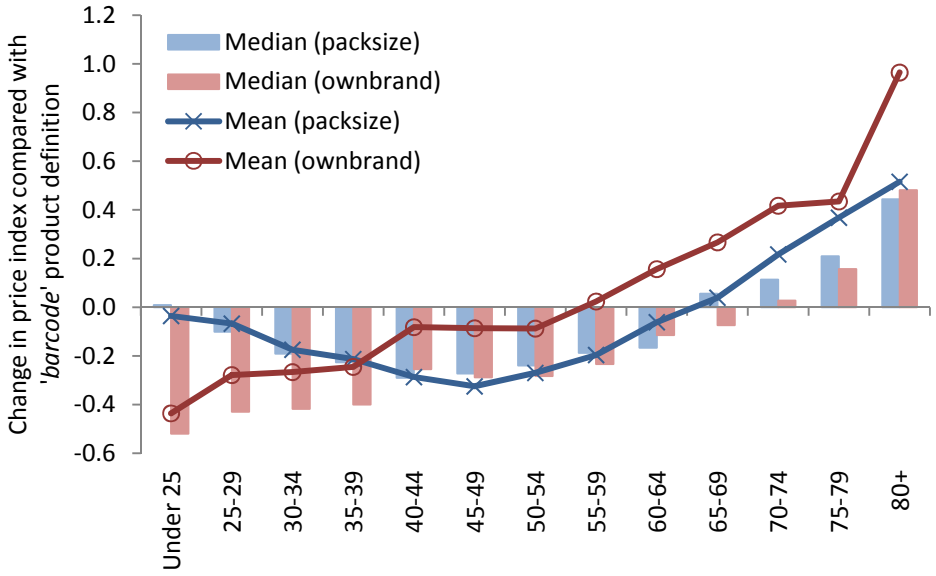
Note: Poor households are those with equivalised income below £20,000.

Table B.6. Variation in average price index by age group

Excluded: 25–29	β	se	t	*Excluded: 25–29, non-poor*	β	se	t
Under 25	–0.102	0.208	–0.49	Under 25	–0.222	0.288	–0.77
30–34	–0.047	0.100	–0.47	30–34	–0.082	0.126	–0.66
35–39	–0.156	0.098	–1.60	35–39	–0.160	0.124	–1.29
40–44	–0.150	0.102	–1.48	40–44	–0.205	0.129	–1.59
45–49	–0.139	0.103	–1.36	45–49	–0.179	0.131	–1.37
50–54	–0.118	0.100	–1.19	50–54	–0.171	0.127	–1.35
55–59	–0.084	0.099	–0.85	55–59	–0.179	0.127	–1.41
60–64	0.131	0.102	1.28	60–64	0.108	0.134	0.81
65–69	0.100	0.106	0.95	65–69	–0.053	0.150	–0.35
70–74	0.248	0.107	2.32	70–74	0.141	0.152	0.93
75–79	0.466	0.127	3.68	75–79	0.272	0.188	1.45
80+	0.983	0.158	6.22	80+	1.411	0.248	5.69
				poor	–0.648	0.157	–4.12
				< 25 × poor	0.394	0.417	0.95
				30–34 × poor	0.074	0.206	0.36
				35–39 × poor	0.053	0.199	0.27
				40–44 × poor	0.165	0.208	0.79
				45–49 × poor	0.129	0.210	0.61
				50–54 × poor	0.166	0.203	0.82
				55–59 × poor	0.297	0.201	1.48
				60–64 × poor	0.176	0.207	0.85
				65–69 × poor	0.493	0.216	2.28
				70–74 × poor	0.418	0.218	1.91
				75–79 × poor	0.565	0.260	2.18
				80+ × poor	–0.410	0.326	–1.26
N	12,885			N	12,885		
R^2	0.009			R^2	0.019		

Notes: Dependent variable is log of household average price index multiplied by 100. Coefficients can be interpreted as approximate percentage difference compared with the excluded group. Poor households are those with equivalised income below £20,000.

Figure B.2. Average change in price index relative to '*barcode*' product definition, by age group



Note: Negative figure indicates group gains on average from switch in product definition (price index decreases).

Household composition

Table B.7. Price indices by household composition, June 2006

	Laspeyres	**CPD**	**EKS**	**Own-basket**		**N**
				All goods	Overlap	
1 non-pensioner	100.41	100.00	99.99	100.02	99.87	1,743
1 pensioner	101.58	100.98	101.07	100.77	100.88	1,746
2 non-pensioners	100.18	100.02	99.94	100.09	99.91	3,324
2 pensioners	100.72	100.45	100.36	100.35	100.20	3,046
2 adults + children	99.80	99.82	99.53	99.83	99.49	5,253
1 adult + children	100.10	99.67	99.72	99.83	99.60	850
3+ adults	100.23	100.00	99.99	100.08	99.95	1,892
3+ adults + children	99.75	99.49	99.42	99.59	99.30	896
Range	*1.83*	*1.49*	*1.65*	*1.18*	*1.58*	

Table B.8. Sample sizes by household composition and income group, 2006

	Non-poor	Poor	Total	% poor
One non-pensioner	1,000	256	1,256	20.4
One pensioner	691	601	1,292	46.5
Two non-pensioners	1,788	584	2,372	24.6
Two pensioners	829	1,333	2,162	61.7
Two adults with children	1,940	1,556	3,496	44.5
One adult with children	126	437	563	77.6
3+ adults, no children	534	657	1,191	55.2
3+ adults with children	199	354	553	64.0
All households	7,107	5,778	12,885	44.8

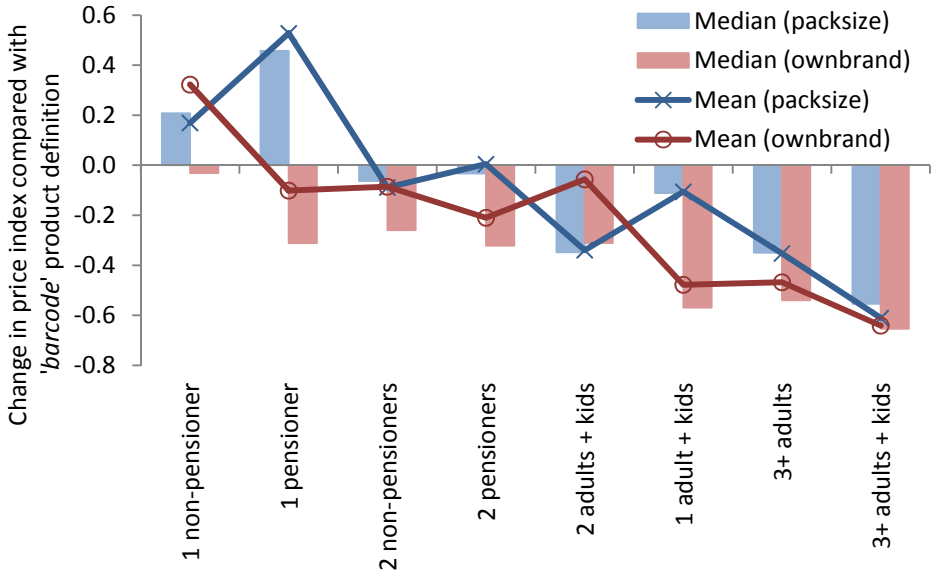
Note: Poor households are those with equivalised income below £20,000.

Table B.9. Variation in average price index by household composition

Excluded: 1 non-pensioner	β	se	t	*Excluded: 1 non-pensioner, non-poor*	β	se	t
1 pensioner	0.241	0.092	2.62	1 pensioner	0.350	0.115	3.05
2 non-pensioners	–0.083	0.081	–1.02	2 non-pensioners	0.039	0.091	0.42
2 pensioners	0.080	0.082	0.97	2 pensioners	0.258	0.109	2.37
2 adults with children	–0.342	0.076	–4.48	2 adults with children	–0.129	0.090	–1.43
1 adult with children	–0.370	0.118	–3.14	1 adult with children	–0.244	0.219	–1.11
3+ adults	–0.233	0.094	–2.48	3+ adults	–0.085	0.124	–0.69
3+ adults with children	–0.718	0.119	–6.05	3+ adults with children	–0.681	0.180	–3.78
				poor	–0.106	0.162	–0.65
				1 pensioner × poor	–0.175	0.207	–0.84
				2 non-pensioners × poor	–0.474	0.196	–2.42
				2 pensioners × poor	–0.218	0.192	–1.14
				2 adults with children × poor	–0.423	0.180	–2.34
				1 adult with children × poor	–0.085	0.285	–0.30
				3+ adults × poor	–0.201	0.211	–0.95
				3+ adults with children × poor	0.014	0.262	0.05
N	12,885			N	12,885		
R^2	0.009			R^2	0.016		

Notes: Dependent variable is log of household average price index multiplied by 100. Coefficients can be interpreted as approximate percentage difference compared with the excluded group. Poor households are those with equivalised income below £20,000.

Figure B.3. Average change in price index relative to '*barcode*' product definition, by household composition



Note: Negative figure indicates group gains on average from switch in product definition (price index decreases).

Local area deprivation

Table B.10 shows that, as would be expected, there are fewer high-income households in more deprived areas, and fewer low-income households in less deprived areas, relative to the national income distribution in our data. For example, 13 per cent of households in the least deprived areas have equivalised incomes below £10,000 compared with 17 per cent of all households in our sample, whilst 20 per cent of households in the most deprived areas have incomes over £30,000 compared with 26 per cent of all households. Because we only have deprivation measures for England, all figures exclude non-English households.

Table B.10. Sample sizes by income group and deprivation tertile, 2006

Income group	Least deprived	2nd tertile	Most deprived	All
< £10k	411 (13%)	580 (15%)	880 (22%)	1,871 (17%)
£10k–£20k	785 (25%)	1,070 (28%)	1,198 (29%)	3,053 (28%)
£20k–£30k	888 (28%)	1,086 (28%)	1,141 (28%)	3,115 (28%)
£30k–£40k	555 (18%)	580 (15%)	446 (11%)	1,581 (14%)
£40k–£50k	219 (7%)	209 (5%)	177 (4%)	605 (5%)
£50k–£60k	169 (5%)	178 (5%)	139 (3%)	486 (4%)
£60k+	144 (5%)	115 (3%)	96 (2%)	355 (3%)
Total	3,171 (100%)	3,818 (100%)	4,077 (100%)	11,066 (100%)

Note: Figures include English households only.

The price index results are shown in Table B.11. Column 1 is comparable to Table 5.4 and shows how prices vary with income; as before when looking at British results, richer households pay higher prices. Column 2 shows that postcode district deprivation (where a higher value indicates a more deprived area) is negatively correlated with prices paid,

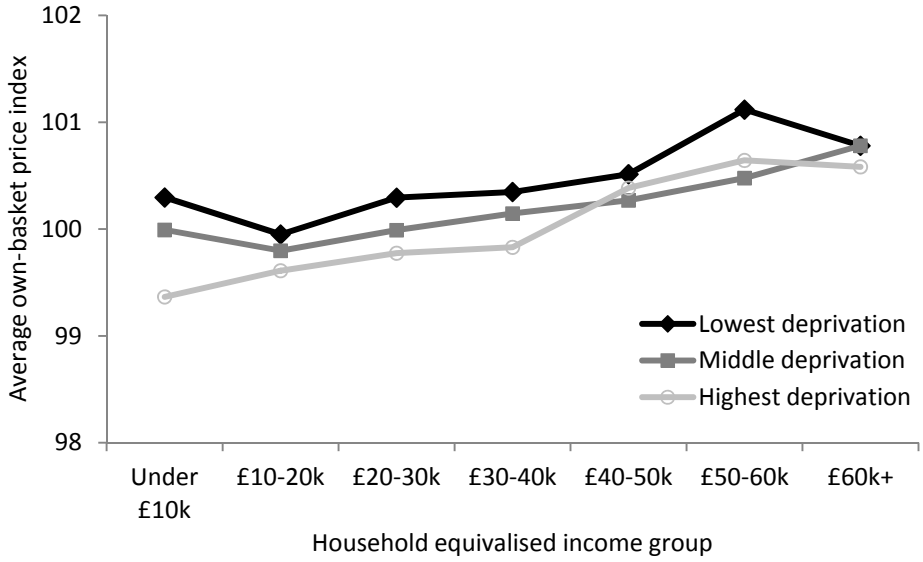
which tells the same story: households living in more deprived areas pay lower prices. Column 3 shows that even if we condition on the local area deprivation index, household income is still positively correlated with prices paid, though the size of the coefficients is slightly reduced.

Table B.11. Effect of local area deprivation and household income on prices

Dependent variable: log average price index	(1)	(2)	(3)	Least deprived	2nd tertile	Most deprived
< £10k	Excluded	—	Excluded	0.935*	0.638*	Excluded
£10k–£20k	0.010	—	–0.025	0.599*	0.450*	0.259*
£20k–£30k	0.243*	—	0.193*	0.938*	0.641*	0.424*
£30k–£40k	0.376*	—	0.288*	1.000*	0.802*	0.475*
£40k–£50k	0.644*	—	0.546*	1.167*	0.922*	1.044*
£50k–£60k	0.996*	—	0.908*	1.766*	1.132*	1.298*
£60k+	0.969*	—	0.864*	1.431*	1.421*	1.226*
IMD score	—	–0.022*	–0.019*	—	—	—
N	11,066	11,066	11,066		11,066	
R^2	0.014	0.011	0.023		0.024	

* = significant at 5 per cent level; † = significant at 10 per cent level.
Notes: Results include English households only. Dependent variable is scaled by 100; coefficients represent approximate percentage effects.

Figure B.4. Average own-basket index by income and local area deprivation tertile, 2006



Notes: Figures are averages by income and deprivation group of household-level averages calculated from monthly estimates. Figures include households in England only.

Figure B.4 shows the average price index for households in each income group by deprivation tertile. The final three columns of Table B.11 show results from a regression of the household's (log) price index on a fully interacted set of dummies for household income and deprivation tertile, which allows us to assess how significant these differences are. The coefficients compare the price index for each group with the poorest households

(incomes below £10,000) living in the most deprived areas.[45] These poorest households pay significantly lower prices than all other income groups in all deprivation tertiles. The richest households in the least deprived areas pay prices that are around 1.5 per cent higher than those paid by the poorest households in the most deprived areas. The patterns from Figure B.4 are also evident in these results. Within area, prices rise with income. Within the most deprived areas, all income groups pay significantly more than those with incomes below £10,000; in the other deprivation tertiles, those with incomes above £50,000 pay significantly more than those with incomes below £10,000. For a given income group, prices tend to fall as the local area becomes more deprived, at least for those with incomes below £40,000. There is, however, no significant difference in prices paid for those with incomes over £40,000 in areas with different deprivation levels.

[45] These coefficients are obtained by adding the relevant coefficients on the income groups, deprivation tertiles and their interactions and checking the joint significance.

References

Acheson, D. (1998), *Independent Inquiry into Inequalities in Health*, London: The Stationery Office (http://www.archive.official-documents.co.uk/document/doh/ih/ih.htm).

Aguiar, M. and E. Hurst (2007), 'Life-cycle prices and production', *American Economic Review*, 97(5), 1533–59.

Balk, B. (1996), 'A comparison of ten methods for multilateral international price and volume comparison', *Journal of Official Statistics*, 12(2), 199–222.

Beaumont, J., T. Lang, S. Leather and C. Mucklow (1995), *Report from the Policy Sub-Group to the Nutrition Task Force Low Income Project Team of the Department for Health*, Radlett: Institute of Grocery Distribution.

Bell, D. R. and J. M. Lattin (1998), 'Shopping behavior and consumer preference for store price format: why "large basket" shoppers prefer EDLP', *Marketing Science*, 17(1), 66–88.

Boizot, C., J-M. Robin and M. Visser (2001), 'The demand for food products: an analysis of interpurchase times and purchased quantities', *Economic Journal*, 111(470), 391–419.

Broda, C., E. Leibtag and D. E. Weinstein (2009), 'The role of prices in measuring the poor's living standards', *Journal of Economic Perspectives*, 23(2), 77–97.

Burchardt, T. (2008), *Time and Income Poverty*, CASEreport 57, London: Centre for Analysis of Social Exclusion (http://sticerd.lse.ac.uk/dps/case/cr/CASEreport57.pdf).

Carlson, J. A. and R. J. Gieseke (1983), 'Price search in a product market', *Journal of Consumer Research*, 9(4), 357–65.

Chung, C. and S. L. Myers, Jr. (1999), 'Do the poor pay more for food? An analysis of grocery store availability and food price disparities', *Journal of Consumer Affairs*, 33(2), 276–96.

Competition Commission (2000), *Supermarkets: A Report on the Supply of Groceries from Multiple Stores in the United Kingdom*, London: The Stationery Office (http://www.competition-commission.org.uk/rep_pub/reports/2000/446super.htm#full).

Competition Commission (2008), *The Supply of Groceries in the UK Market Investigation*, London: The Stationery Office (http://www.competition-commission.org.uk/rep_pub/reports/2008/fulltext/538.pdf).

Crawford, I. (2003), 'Variations in the price of foods and nutrients in the UK', Institute for Fiscal Studies (IFS), Working Paper 03/19 (http://www.ifs.org.uk/wps/wp0319.pdf).

Cummins, S. and S. Macintyre (2002), 'A systematic study of an urban foodscape: the price and availability of food in Greater Glasgow', *Urban Studies*, 39(11), 2115–30.

Diewert, W. E. (1999), 'Axiomatic and economic approaches to international comparison', in A. Heston and R. E. Lipsey (eds), *International and Interarea Comparisons of Income, Output, and Prices*, Chicago: University of Chicago Press.

Diewert, W. E. (2005), 'Weighted country product dummy variable regressions and index number formulae', *Review of Income and Wealth*, 51(4), 561–70.

Economides, N. (1989), 'Desirability of compatibility in the absence of network externalities', *American Economic Review*, 79(5), 1165–81.

Einav, L., E. Leibtag and A. Nevo (2008), *On the Accuracy of Nielsen Homescan Data*, Economic Research Report 69, Washington, DC: US Department of Agriculture (http://www.ers.usda.gov/Publications/ERR69/ERR69.pdf).

Elteto, O. and P. Koves (1964), 'On a problem of index number computation relating to international comparison', *Statisztika Szemle*, 42, 507–18.

Erdem, T., S. Imai and M. Keane (2003), 'Consumer price and promotion expectations: capturing brand and quantity choice dynamics under price uncertainty', *Quantitative Marketing and Economics*, 1(1), 5–64.

Griffith, R., E. Leibtag, A. Leicester and A. Nevo (2009), 'Consumer shopping behavior: how much do consumers save?', *Journal of Economic Perspectives*, 23(2), 99–120.

Griffith, R. and M. O'Connell (2009), 'The use of scanner data for research into nutrition', *Fiscal Studies*, 30(3–4), 339–65.

Hendel, I. and A. Nevo (2006a), 'Sales and consumer inventory', *RAND Journal of Economics*, 37(3), 543–61.

Hendel, I. and A. Nevo (2006b), 'Measuring the implications of sales and consumer inventory behavior', *Econometrica*, 74(6), 1637–73.

Hill, R. (1997), 'A taxonomy of multilateral methods for making international comparisons of prices and quantities', *Review of Income and Wealth*, 43(1), 49–69.

Hotelling, H. (1929), 'Stability in competition', *Economic Journal*, 39(153), 41–57.

Kaufman, P. R., J. M. MacDonald, S. M. Lutz and D. M. Smallwood (1997), *Do the Poor Pay More for Food? Item Selection and Price Differences Affect Low-Income Household Food Costs*, Agricultural Economic Report 759, Washington, DC: US Department of Agriculture (http://ageconsearch.umn.edu/bitstream/34065/1/ae970759.pdf).

Lal, R. and R. Rao (1997), 'Supermarket competition: the case of everyday low pricing', *Marketing Science*, 16(1), 60–80.

Leicester, A., C. O'Dea and Z. Oldfield (2008), *The Inflation Experience of Older Households*, Commentary 106, London: Institute for Fiscal Studies (http://www.ifs.org.uk/comms/comm106.pdf).

Leicester, A. and Z. Oldfield (2009a), 'Using scanner technology to collect expenditure data', *Fiscal Studies*, 30(3–4), 309–37.

Leicester, A. and Z. Oldfield (2009b), 'An analysis of consumer panel data', Institute for Fiscal Studies (IFS), Working Paper 09/09 (http://www.ifs.org.uk/wps/wp0909.pdf).

MacDonald, J. M. and P. E. Nelson, Jr. (1991), 'Do the poor still pay more? Food price variations in large metropolitan areas', *Journal of Urban Economics*, 30(3), 344–59.

Mooney, C. (1990), 'Cost and availability of healthy food choices in a London health district', *Journal of Human Nutrition and Dietetics*, 3(2), 111–20.

National Consumer Council (2004), *Why Do the Poor Pay More ... or Get Less?*, London: National Consumer Council.

Piachaud, D. and J. Webb (1996), *The Price of Food: Missing Out on Mass Consumption*, London: Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Salop, S. (1979), 'Monopolistic competition with outside goods', *Bell Journal of Economics*, 10(1), 141–56.

Salop, S. and J. E. Stiglitz (1982), 'The theory of sales: a simple model of equilibrium price dispersion with identical agents', *American Economic Review*, 72(5), 1121–30.

Shaked, A. and J. Sutton (1982), 'Relaxing price competition through product differentiation', *Review of Economic Studies*, 49(1), 3–13.

Smith, D. M., S. Cummins, M. Taylor, J. Dawson, D. Marshall, L. Sparks and A. S. Anderson (2010), 'Neighbourhood food environment and area deprivation: spatial accessibility to grocery stores selling fresh fruit and vegetables in urban and rural settings', *International Journal of Epidemiology*, 39(1), 277–84.

Sobel, J. (1984), 'The timing of sales', *Review of Economic Studies*, 51(3), 353–68.

Sooman, A., S. Macintyre and A. Anderson (1993), 'Scotland's health – a more difficult challenge for some? The price and availability of healthy foods in socially contrasting localities in the west of Scotland', *Health Bulletin (Edinburgh)*, 51(5), 276–84.

Stigler, G. (1961), 'The economics of information', *Journal of Political Economy*, 69(3), 213–25.

Szulc, B. J. (1964), 'Indices for multiregional comparisons', *Prezegald Statystyczny*, 3, 239–54.

US Department of Agriculture (2009), *Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences – Report to Congress*, Washington, DC: USDA (http://www.ers.usda.gov/Publications/AP/AP036/).

Varian, H. (1980), 'A model of sales', *American Economic Review*, 70(4), 651–9.