

Do Prices and Attributes Explain International Differences in Food Purchases?*

Pierre Dubois

Rachel Griffith

Aviv Nevo

Abstract

Food purchases differ substantially across countries. We use detailed household level data from the US, France and the UK to (i) document these differences; (ii) estimate a demand system for food and nutrients, and (iii) simulate counterfactual choices if households faced prices and nutritional characteristics from other countries. We find that differences in prices and characteristics are important and can explain some difference (e.g., US-France difference in caloric intake), but generally cannot explain many of the compositional patterns by themselves. Instead, it seems an interaction between the economic environment and differences in preferences is needed to explain cross country differences.

*We thank Martin Browning, Robin Lee, Arthur Lewbel, Peter Neary and participants in several seminars and conferences for comments. We gratefully acknowledge financial support from USDA, ERC-2009-AdG grant 249529, ANR-ESRC ORA grant number ES/I012222/1 and the Center for the Study of IO at Northwestern University. This research was funded in part by a cooperative agreement between the USDA/ERS and Northwestern University, but the views expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Department of Agriculture. Dubois is at the Toulouse School of Economics, Griffith is at the Institute for Fiscal Studies and the University of Manchester and Nevo is at Northwestern University and NBER. Contact information: pierre.dubois@tse-fr.eu, rgriffith@ifs.org.uk and nevo@northwestern.edu.

1 Introduction

Food purchases differ across countries, within countries over time and across demographic groups. These differences seem to be correlated with rates of excess weight gain and diet related illness. Economists tend to attribute the difference in food purchases across markets to differences in prices. An alternative explanation is that they are due to inherent differences in preferences and eating habits. In this paper we study the differences in food purchases and nutritional outcomes across countries, focusing on the US, UK and France.

We start by systematically documenting the differences in food purchases for consumption at home between the US, UK and France. We show that US households purchase different foods from French and UK households. For example, US households purchase more calories per person. A greater percentage of those calories come in the form of carbohydrates, and a lower share in the form of proteins. A higher share of expenditure is on drinks and prepared foods, and a lower share is on fruits and vegetables. We also document substantial differences in prices and nutritional characteristics of foods across the three countries.

This leads to the main contribution of the paper, which is to consider whether prices and nutritional characteristics can explain the observed differences in food purchases. To answer this question we develop and estimate a model of demand for food products and nutrients in each country. We estimate the model using household-level (home scanner) data that include detailed food purchases for an extended period for participating households in each of the three countries. The purchase data is merged with data on nutrient content at a disaggregated product level. We use the estimates to simulate the quantities and nutritional content of the food basket that US households would purchase if they faced the prices and food characteristics faced by households in France and the UK. This allows us to better understand how important differences in prices and nutritional attributes are in explaining differences in food purchases, as opposed to preferences or other factors. We consider this counterfactual scenario not necessarily because changing US prices and product attributes to those in France or the UK is a feasible policy, but because we think it is informative in helping us to understand why the nutritional balance of households' food baskets are so different across the three countries.¹ We conduct the simulations for the average US household, but

¹For discussion of the wide range of policies to address obesity that are under consideration see, *inter alia*, Acs et al. (2007), Gortmaker et al. (2011), Griffith and O'Connell (2010), and Philipson and Posner (2008).

we also separately simulate the effect on two other groups that are of particular policy interest: low income households and households that purchase more calories (per capita).

We find that, if faced with French prices and product attributes, the average US household would purchase substantially fewer calories: a similar level to the average French household when faced with the same environment. However, the composition of these calories would differ. The simulated change is mostly due to price differences. If we change only the nutrient characteristics this has little impact on the amount of calories the average US household obtains, though it does affect the form of those calories, shifting them away from carbohydrates and towards proteins and fats. In contrast, when we simulate the average US household's food basket with UK product attributes this has a substantial impact on reducing calories, whereas changing relative prices in fact increases calories. For some questions it can be misleading to focus only on total calories. The simulations suggest that, even when the total calories purchased is not affected, the composition of macronutrients and food groups can change substantially.

The results suggest that, while the economic environment, as reflected in prices and attributes, can have a large impact on food purchases and the nutritional composition of the food basket, prices and attributes do not fully explain the observed differences. Price differences mostly explain the large difference in caloric intake between the average French and US household. However, nutrient characteristics are important when comparing to the UK, and differences in preferences and eating habits are generally quite important, and in some cases can offset the influences of the economic environment. For example, we find that UK households have healthier purchasing patterns than US households despite the prices and product offering they face, not because of them.

There are several reasons to be interested in the differences in food purchases across the three countries. Of primary interest is the fact that the differences in nutritional characteristics are mirrored in a number of health outcomes. The National Research Council (2011) reports differences along a number of dimensions, most of which show that the US has poorer health than the UK and France. For example, 36.4% of men aged 65+ in the US report having heart disease compared to 28.8% in France and 32.2% in the UK; diabetes prevalence is 21.4% in the US compared to 13.0% in France and 11.2% in the UK (Table 2-1

of National Research Council, 2011). Obesity rates are also the highest in the US at 30.0%, compared to 14.5% in France and 23.6% in the UK.²

Differences in obesity rates across countries, and implications for health outcomes, are due to many factors, including exercise and general life style, but are likely to at least in part also be due to differences in food consumption patterns.³ More generally, nutrition is well understood to be an important determinant of health outcomes, leading to economic costs, including medical costs, lost productivity and a reduction in the quality of life. For example, in the UK poor diet is estimated to account for about one-third of all deaths from cancer and cardiovascular disease, and the US Center for Disease Control estimates that in 2008 medical costs related to obesity were as high as \$147 billion (CDC, 2011).

In order to address our main question, and to exploit the richness of our data, we develop a model of demand that nests models in product space and those in characteristics space. To understand the need for this model consider two commonly used alternatives. The first approach is to model demand at a disaggregate product level, for example demand for soft drinks, and assess the importance of prices and various characteristics. This approach will pin down preferences within narrowly defined product groups, but will not let us address questions of choice among product groups and differences in the overall food basket. Furthermore, narrowly defined products (i.e. brands) are very different across countries, creating problems with matching products across the countries. A second common approach would be to model demand for food at a much more aggregated level.⁴ However, this would not take advantage of the detailed information in our data, nor does it account for the differences across countries in the attributes of food offered; each food category would be assumed to be the same across all countries.

We instead propose a model in which a consumer chooses continuous quantities of each of a large number of products in order to maximize utility, which depends both on the characteristics of the products, as in Gorman (1956) and Lancaster (1966), and on the quantity consumed of each product. This nests commonly used models in characteristics space, such as the discrete choice model (McFadden, 1974, Berry, Levinsohn and Pakes, 1995), and the

²See also National Academy of Sciences (2013).

³See also, Abaluck (2011), Bleich et al (2007), Bawa (2005), Chou et al (2004), Duffey and Popkin (2011), French et al (2001), Drewnowski and Specter (2004), Finkelstein and Zuckerman (2008), Lakdawalla and Philipson (2002, 2009), Philipson et al (2004), Philipson and Posner (2011), Swinburn et al (2009).

⁴For example, following this approach Deaton (1997) studies demand for food in developing countries, and Seale et al (2003) provide a descriptive analysis of differences in food consumption patterns across countries.

hedonic price model (inter alia Court, 1939, Griliches, 1961, Rosen, 1974, Epple, 1987). Our model also generalizes classical demand models in product space, such as Cobb Douglas, Translog or the Almost Ideal Model, which typically rely on weak separability of preferences in order to make the analysis tractable. We relax the weak separability assumption by creating an interaction between products through the characteristics they supply.

Key to our analysis is the rich micro data we use. In each country we have a sample of several thousand households for whom we observe *all* purchases of food for consumption at home; that is we know the households' entire food basket. We know precisely what product was bought, the quantity that was purchased, how much was paid and crucially its nutritional content. We use data for 2005-2006. To facilitate cross-country comparison, and to make the estimation of demand tractable, we need to aggregate the data to similar categories of food items across countries. We specify a model of demand that explicitly aggregates from the individual product level and comes from a direct specification of the utility model. Our model yields a simple linear estimating equation, which relates expenditure on products to their nutritional content, and it allows us to control for rich patterns of heterogeneity using household level fixed effects. This tractability comes at a cost: for a given individual with a given draw of random parameters, the functional form of how income and prices affect quantity demands is very limited.

The variation over time and across households in the underlying available products (and their nutritional components) is key for our estimation. An endogeneity problem arises from the fact that quantities appear both in the dependent variable and as an explanatory variable in the quantity of nutrients purchased. To account for this endogeneity we use variation in the nutritional content of products available, which we assume is exogenous conditional on our controls. This idea is similar to using variation in product attributes to identify demand, which is popular in the IO literature (Bresnahan, 1981, Berry, Levinsohn and Pakes, 1995). To generate this variation we rely on the detailed nutrition information available in our data.

Our paper is related to a literature that tries to attribute the differences in obesity over time and across markets to differences in prices. For example, Cutler et al. (2003) suggest that the decrease in the price of calories has increased caloric intake and contributed to the increase in obesity. Philipson and Posner (2003) also suggest that a change in price is a key driver of the increase in obesity, but focus on the price of burning calories, which has gone

up over time.⁵ Neither of these papers provides direct evidence on the importance of the economic environment relative to other factors, such as the change in the nutritional content of food or difference in preferences. We are able to add to this literature by providing direct evidence on the economic determinants of the nutritional balance of households' food basket.

Another closely related paper is Atkin (2012), who documents regional consumption differences in India, and explores the origins of these differences, as well as their implications for nutrition. He introduces habit formation into a overlapping-generations general equilibrium model in order to study the causes and consequences of the regional taste differences. In his model, households develop tastes for locally-abundant foods that they were fed as children. Over generations regional differences in preference and consumption arise endogenously through habit formation. He tests the predictions of the model using cross-regional migration data. He concludes, as we do for our setting, that an interaction between preferences and the environment is needed to explain observed regional differences. The context and identification strategies in the two papers are different, yet we reach similar conclusions. We believe that the fact that we reach similar conclusions to Atkin, studying different settings and using different methodologies, adds to the credibility of the findings.

In the nutrition literature, Drewnowski (2004) and Drewnowski and Specter (2004) show that energy dense foods – foods with more calories per unit of weight – are negatively correlated with price per calorie. Similarly, Drewnowski et al (2007), and Maillot et al (2007), use French data to show that households who buy energy dense food baskets also tend to spend less on food. Based on these associations they conclude that relative price differences are a key cause to a poor diet. Our analysis differs in several significant ways. First, we have more detailed price and purchase data, along with nutritional information at a very disaggregated level (see Griffith and O'Connell, 2009, for the importance of detailed data). Second, we can account for several macro nutrients, not just calories or calories per unit weight. Third, we estimate the causal effect of prices and characteristics on consumer choice. Finally, we are able to go further than simply considering the direction of change, we can simulate and quantify the effect of a change in prices and nutrients.

The structure of the rest of the paper is as follows. The next section describes the data and provides an initial description of differences in the nutritional characteristics of shopping baskets and behavior across countries. In section 3 we develop a model of demand over

⁵In addition to the above papers, see also Goldman et al (2009) and Lu and Goldman (2010).

products and characteristics. In section 4 we estimate the model and discuss the implications of our estimates and simulations. A final section summarizes and concludes.

2 Comparison of Food Purchases

2.1 Data

We use detailed data collected by market research firms using similar methodology in the US, UK and France. In France and the UK the data come from the Kantar (formerly known as TNS) WorldPanel, while in the US the data were collected by Nielsen as part of the Homescan panel. These data include information on all food purchased and brought into the home by a large number of households over a two year period (2005-2006); the data are recorded by households using handheld scanners in the home. We have information on quantities, prices and characteristics of the products purchased at the level of the individual food product, as defined by the barcode or what is called the Universal Product Code (UPC) in the US. The characteristics include nutritional characteristics such as calories, proteins, fats and carbohydrates, as shown on nutritional labels.⁶

Table 1 provides some descriptive statistics on the demographics of the sample of households we use in each of the three countries. These are a sub-set of all households in the data. We drop households that are outliers (as described in the Appendix). In the US, Nielsen asks a sub sample to report purchases of random-weight products, which are products that are not pre-packaged and typically do not have a UPC. Random-weight items are common in fruit, vegetables, meat and deli items. Since collecting information on random-weight products is time consuming, Nielsen only asks a fraction of its panelists to collect these data. The US sample we use are those households that reported random-weight purchases and therefore the number of households in the US is smaller than in the UK and France, despite having a larger overall sample.

Family structure, as measured by household size, number of kids and age, is similar across the samples from the three countries. In what follows we use the household as the unit of

⁶The nutritional information is of the same form across the countries but it was collected somewhat differently in the three countries. In the UK the nutritional information was collected by Kantar from manufacturers, food labels and by direct measurement. In the US the data on purchases from Nielsen was matched with nutritional information from Gladson, and in France the nutritional information was collected directly from labels and public sources. The Data Appendix provides details on the construction of the data.

analysis. To control for difference in size and composition across households we use an adult equivalence scale based on caloric needs.⁷ We sum the daily caloric needs of each member of the household (based on age and gender) and divide by 2500, which is the caloric needs of an adult male (19-59). We provide details in Section A.4 in the Appendix. The sample average of this measure is also similar across the countries.

Table 1 : Demographics

	France	UK	US
# of households	11,677	12,698	8,484
Household size	2.7	2.6	2.4
# of kids	0.7	0.6	0.5
Adult equivalent	2.2	2.1	2.0

Notes: numbers are averages across households in the sample used in subsequent analysis. Adult equivalent is a caloric needs scale: we sum the daily caloric needs of each household member (based on age and gender) and divide by 2500.

A key advantage of the detailed data is that they allow direct measurement of prices and characteristics of a *substantial* part of households' food purchases. Precise information on prices and detailed attributes of the products allows us to estimate preferences. Estimates using more aggregated purchase data and coarse data on expenditures and prices provides much less accurate estimates of preference parameters. Panel data on households' purchases also allows us to control for individual heterogeneity in a rich way. Many standard sources of information on food purchases are cross-sectional, and therefore rely on cross household differences for identification of preference parameters. Instead we rely mainly on within household variation. Nonetheless, before proceeding with our analysis we should be up-front about several potential concerns with the data.

The data in all countries are collected by households themselves within the home, and as such might suffer from recording error. To document the extent of this problem Einav, Leibtag and Nevo (2010) compare Nielsen Homescan data to information from cash registers of a retailer and find that in some dimensions the US data are indeed prone to error, but the amount of noise seems equivalent to that found in many data sets commonly used. For example, Bound and Krueger (1991) find that the variance of the log of the ratio of earnings reported in the CPS with Social Security administrative is 0.114, while Einav, Leibtag and Nevo (2010) find the variance of the log of the ratio of Homescan and retailer price is 0.139.

⁷An alternative to the equivalence scale is the more general approach of Lewbel (1985) and Lewbel and Pendakur (2011), which takes into account observed and unobserved heterogeneity in equivalence scales.

In both studies the correlation between the reported and true variables is 0.88. Leicester and Oldfield (2009) compare the UK data to data from the Family Expenditure Survey and “suggest that problems of fatigue and attrition may not be so severe as may be expected.” We note that even if recording errors exist, as long as there is no systematic differences in reporting errors across the countries our findings should not be significantly impacted. In addition, the rich controls for heterogeneity we introduce in the econometric analysis will help to control for differences across households in recording.

Collecting the data is quite time consuming and therefore likely to generate a selection in who agrees to participate in the sample. Indeed, the demographics in Table 1 suggest that the households in the sample are from smaller households, have fewer children and are older than the respective national averages. However, the numbers in the table also suggest that the demographics of the sample participants are similar across the three countries.

The data we have does not include food purchases for consumption outside the home, for example in restaurants. It is also purchase data, not consumption data, so does not reflect food that is wasted. These considerations could matter for some of the descriptive analysis, but largely they make the differences larger - the best estimates that we can find suggest that US households eat out more and waste more food than either French or UK households. In the simulations our main assumption is that behavior along these two dimensions remains unchanged, however, we consider what implications these considerations might have for our analysis in Section 4.4.

2.2 Purchasing Patterns

We start by considering total food purchased.⁸ Table 2 describes total calories, nutrients and expenditure, with all figures reported per day per adult equivalent. The first row reports the average across households of total calories purchased per person per day. The second, third and fourth rows show the average amount of calories in terms of each of the three macronutrients (carbohydrates, proteins and fats), the fifth, sixth and seventh row show the average amount of each macronutrient in terms of grams, and the final row shows average household expenditure on food per person per day in US\$.

⁸We will use the terms "purchases" and "consumption" interchangeably. Our data records purchases, which we will assume equals consumption. In reality, some food might be thrown away without being consumed, or it might be consumed by someone who is not a member of the household.

There are some striking differences. US households purchase more calories per person. This is even more striking given the higher propensity to eat food out in the US. In terms of the balance of macronutrients the US and UK households are similar on aggregate, but comparing the US to the French the extra calories are in the form of carbohydrates. French households purchase both a larger amount and higher share of their calories in the form of proteins and fat. Average spending on food is higher in France than in the US or UK.

These broad patterns hold not just at the mean, but at other parts of the distribution as well. For example, in the counterfactual analysis we compare low income households, and also households that are at the top of the calorie consumption distribution.

Table 2 : Mean Consumption Across Countries

	FR	UK	US
calories	1776.6	1928.9	2102.7
<i>from carbohydrates</i>	667.4 (38%)	890.5 (47%)	1019.3 (49%)
<i>from protein</i>	287.9 (16%)	293.3 (16%)	264.9 (13%)
<i>from fats</i>	821.0 (46%)	694.5 (37%)	781.6 (37%)
carbohydrates (g)	178.0	237.5	271.8
proteins (g)	72.0	73.3	66.2
fats (g)	91.2	77.2	86.8
expenditure (\$)	5.03	4.71	4.59

Notes: Figures reported are the average per person per day using an adult equivalent scale over 2005-2006. Expenditure is in US\$ using an exchange rate of £1 = \$1.80 and €1 = \$1.25.

The recommended percentage of energy consumed in the form of each macronutrient that is consistent with good health are roughly similar across the three countries.⁹ The figures in Table 2 suggest that the French are the most out of line with the guidelines. However, we should be careful in making this inference since these figures do not include food consumed at restaurants.

To further study the cross-country differences we look at how these purchases are divided between nine broad food categories. The categories are commonly used by the USDA for descriptive analysis, and were chosen for their nutritional characteristics; foods within each category share a similar nutrient composition. The Appendix (Section A.4) details what products are included in each category.

⁹The French government agency coordinating nutrition information and policy, Programme National Nutrition Santé- Afssa (2002) recommends 50-55% carbohydrates, 11-15% protein and 30-35% fats. The UK Department of Health (1991) recommends 50%, 15% and 35%, while the US Guidelines, available at <http://www.health.gov/dietaryguidelines/dga2010/DietaryGuidelines2010.pdf>, recommend 45-65% carbohydrates, 10-30 % protein and 25-35% fat.

Table 3 reports average household expenditure, expenditure shares and quantities across the nine food categories. There are considerable differences in expenditure by food category across the countries. The UK and US expenditure patterns are more similar, while the French numbers are different. The average French household spends less on processed food, such as drinks and prepared foods, and more on basic ingredients such as meats, dairy, fruits and vegetables, both in dollar terms and as a fraction of overall expenditure. The average UK and US household spends less than French households on meats and the UK spend more on grains, while the average US household spends less on dairy and more on drinks and prepared foods.

Table 3: Expenditure and Quantity by Category

Category	Expenditure (\$ per qtr)			Exp Shares (%)			Quantity (kilo per qtr)			Calorie Share (%)		
	FR	UK	US	FR	UK	US	FR	UK	US	FR	UK	US
Fruits	29.65	42.34	34.54	6.6	9.3	8.1	14.6	14.0	17.2	4.5	4.5	5.3
Vegetables	44.22	46.13	34.43	9.7	10.4	7.9	18.2	20.2	14.0	5.3	6.0	3.0
Grains	25.33	34.31	30.97	6.0	8.4	7.8	6.7	13.4	8.8	14.3	19.8	14.3
Dairy	74.90	53.55	38.85	16.7	12.7	9.5	25.7	27.9	20.7	17.2	12.8	9.3
Meats	147.53	80.00	80.78	31.0	18.3	19.0	14.2	11.1	14.7	16.6	13.2	16.1
Oils	15.14	8.81	7.96	3.3	2.0	1.9	3.1	2.1	2.2	13.1	6.8	6.6
Sweeteners	5.85	4.30	5.54	1.4	1.1	1.4	2.4	2.4	2.6	5.1	4.9	4.4
Drinks	26.81	24.38	41.44	5.9	5.8	10.1	43.4	17.4	50.0	3.5	2.0	5.9
Prepared	96.35	138.98	151.48	21.2	32.7	36.1	16.4	26.2	30.0	22.8	31.2	38.0

Notes: Figures are the mean of the distribution across households and quarters, and are per person per quarter using an adult equivalent scale, conditional on strictly positive expenditure in that category in that quarter. Expenditure is in US\$ using an exchange rate of £1 = \$1.80 and €1 = \$1.25.

The next three columns present quantities, measured in kilograms.¹⁰ The numbers in these columns present a slightly different picture than the expenditure numbers. The US and UK consumption patterns are now quite different. For example, price differences across countries (discussed in the next section) explain why the French spend much more on meat but purchase a similar quantity to US households. Generally, the French tend to purchase less processed food, such as drinks and prepared foods, and more basic ingredients such as meats, dairy and vegetables. This is especially true compared to the US purchasing patterns. The UK and US purchasing patterns are more similar, but even here there are differences,

¹⁰Here, and elsewhere, we measure quantities in kilos, which should be innocuous when comparing similar products, but can lead to difficulties when aggregating across diverse products. Ideally, we could measure quantities in "servings" but our data does not allow us to do this.

with the average UK household consuming more vegetables, grains and dairy and the average US household consuming more meat and drinks.

In the final three columns we look at the share of calories from each food category. We see some of the same broad patterns as before. The French continue to be somewhat different, purchasing a larger fraction of their calories from vegetables, dairy and meat and less from prepared foods. This will serve as a key motivation for our analysis below: differences in the prices might explain the differences in the quantity of prepared food purchased, but to fully understand the health implications we need to account for the differences in the nutrient content of prepared food between the UK and US.

The numbers in Table 3 give us a first indication of the importance of differences across countries in both prices and the nutrient content of food. Consider the expenditure shares. They tell us something about differences in preferences across countries. Indeed, if we considered a simple Cobb-Douglas utility function these shares would be the preference parameters. However, by looking at the differences between quantities and expenditures, for example between the US and UK, we get a first indication of the importance of prices - similar expenditure shares can translate into quite different quantities purchased, and thus into quite different nutritional outcomes. Similarly, the differences in calorie shares suggest that it is important to control for differences in nutrient content across countries.

2.3 Prices and Product Attributes

In this section we document some of the differences in prices and nutrition attributes across countries that we saw indirectly in Table 3. Table 4 describes the average price per kilo in each product category. With a few exceptions the US prices are the lowest and the prices in France the highest. If we hold the quantities fixed at the levels of Table 3 the US expenditure would increase by 14.4% and 13.9% if paying the UK and French prices. On the other hand, French expenditure would decrease by 13.3% and 2.2% if the French paid US or UK prices. The UK expenditure would decrease by 4.3% if paying US prices but would increase by 12.7% if paying French prices. However, note that relative prices do not all go the same way, the relative prices differ across categories (for example, fruit is cheapest in France, while meat is cheapest in the US) and this will lead to important differences in our simulations.

The differences in relative prices are consistent with some of the differences we saw in Table 3, but it is clear that prices do not tell the whole story. For example, US households purchase more fruit and fewer vegetables than the UK, consistent with the lower relative price of fruit in the US and lower price of vegetables in the UK. On the other hand, the relative price of drinks in the US is somewhat higher than in France, even though the US households purchase slightly more (50 versus 43). We have to be a little careful in this comparison, since it could be driven by composition effects.

Table 4: Mean Prices by Category

	FR	UK	US
Fruits	2.09	3.21	2.12
Vegetables	2.53	2.32	2.64
Grain	3.89	2.63	3.73
Dairy	3.26	2.22	2.48
Meats	10.33	7.29	5.88
Oils	5.19	3.97	4.47
Sweeteners	2.79	2.38	4.61
Drinks	0.89	2.50	1.56
Prepared	6.04	5.43	5.13

Notes: units are US\$ per 1 kilogram using an exchange rate of £1 = \$1.80 and €1 = \$1.25.

The nutrient characteristics of foods on offer and purchased also vary. In Table 5 we show the mean nutrient content of the food products by category in each country. As before, the differences across countries could at least in part be driven by composition effects. For example, the lower carbohydrate content in French drinks is due to the fact that French households buy more water. One way around this is to zoom in on more disaggregated products. An alternative, which we think might be more meaningful here, is to focus on the difference in offerings. We do that by presenting in Table 5 an unweighted average of all the products we *ever* see purchased by households in our sample. In other words, we take the universe of all products ever purchased by any of the households in our sample, and keep one observation for each. This does not totally eliminate the impact of choice, since the product needs to be bought at least once, but it significantly reduces it.

We see large differences across countries. For example, the meat products that US households buy have on average much more fat and carbohydrate than the meat products that French households purchase, which are more protein intensive. Another example, we saw above that the higher fraction of calories from prepared foods in the US is consistent with

prepared foods in the US being more calorie dense relative to UK prepared foods. The difference in calories from prepared foods seems to come from the differences in carbohydrates and fats. Drinks are also much more carbohydrate intense in the US than in the UK, and even more than in France. The differences across countries remain even if we focus on more narrowly defined products.

Table 5: Calories from each Nutrient by Category

	carbohydrates			protein			fat		
	FR	UK	US	FR	UK	US	FR	UK	US
Fruits	57	68	70	3	5	2	8	7	1
Vegetables	39	38	49	20	22	13	76	85	7
Grain	211	129	227	34	22	38	96	20	36
Dairy	18	22	29	71	57	48	188	166	130
Meats	5	21	30	76	72	66	120	129	206
Oils	2	7	6	11	3	2	678	602	671
Sweeteners	305	307	345	3	4	0	0	1	0
Drinks	27	34	69	1	4	2	1	4	5
Prepared	126	95	194	24	23	22	127	88	117

Notes: Figures are means across all products purchased in our sample, with an equal weight to each food product (UPC). Units are calories from each nutrient (carbohydrates, proteins, fats) per 100 grams of food.

3 A Model of Demand

As we saw in the previous section there are cross-country differences in the choices households make and in the prices and product offerings they face. Our aim is to investigate the extent to which cross-country differences in purchases are attributable to differences in prices and the attributes of products (the economic environment). One could address this question by estimating disaggregate product level demand, say by using methods that are standard in the Industrial Organization literature (see, *inter alia*, Berry, Levinsohn and Pakes, 1995, and Nevo, 2000), and simulating the purchases if a household faced different prices and characteristics. This approach does not let us address questions of choice among product groups, does not take advantage of the coverage of our data and is problematic since narrowly defined products (*i.e.*, brands) are very different across countries and therefore hard to match.

Instead, we model demand for food at home more generally at an aggregated level, for example at the level of the nine categories we used in the previous section. We allow preferences to depend on characteristics, nutrients in our case, in order to account for the differences across countries in the attributes of each food category. The model we use builds on Gorman

(1956) and Lancaster (1966), where utility depends on the characteristics of the product. Special cases of the characteristics model are the discrete choice model (McFadden, 1974, Berry, Levinsohn and Pakes, 1995), and the hedonic price model (Court, 1939, Griliches, 1961, Rosen, 1974, Epple, 1987, as well as many others).

As noted by Gorman (1956), one constraint of the pure characteristics model is that it predicts that the number of goods purchased will not exceed the number of characteristics. In a discrete choice setting this is not a constraint, since the consumer chooses a single option. The same is true for the hedonic setting, where the consumer chooses a single option from a continuum of choices (or from a discrete choice set as in Bajari and Benkard, 2005). However, in many settings, of which the one we examine below is an example, consumers choose from a discrete menu of products but choose many products. The key is that the number of products chosen exceeds the number of observed characteristics. In the characteristics approach this can be explained by introducing product specific attributes. As we will see this essentially amounts to going back to a “standard” demand model where utility is defined in product space. The model we propose can nest standard models in characteristics space as well as demand models in product space.

3.1 Theory

A household¹¹ chooses from N products, where product n is characterized by C characteristics $\{a_{n1}, \dots, a_{nC}\}$. We primarily have in mind cases where C is smaller than N , in some cases much smaller. The utility of household i with demographics η_i is given by $U(x_i, \mathbf{z}_i, \mathbf{y}_i; \eta_i)$ where x_i is the numeraire, \mathbf{z}_i is a $C \times 1$ vector of characteristics of food and \mathbf{y}_i is a vector of the quantities purchased of all food products by household i . Define the $N \times C$ matrix $\mathbf{A} \equiv \{a_{nc}\}_{n=1, \dots, N, c=1, \dots, C}$. The household will maximize utility by choosing the quantity of the numeraire, x_i , and of food items, \mathbf{y}_i , subject to a budget constraint:

$$\begin{aligned} & \max_{x_i, \mathbf{y}_i} U(x_i, \mathbf{z}_i, \mathbf{y}_i; \eta_i) \\ s.t. \quad & \sum_{n=1}^N y_{in} p_n + p_0 x_i \leq I_i \quad ; \quad \mathbf{z}_i = \mathbf{A}' \mathbf{y}_i; \quad x_i, y_{in} \geq 0, \end{aligned}$$

¹¹As we saw in the previous section the data is at the household level. To match this we formulate the model as a choice by a household with a well defined utility function. We abstract from issues of intra household interactions.

where p_n is the price of one unit of y_{in} , I_i is the household's income, and p_0 is the price of the outside good x_i .

Following standard arguments (and dropping the i subscripts) this can be written as

$$\begin{aligned} \max_{\mathbf{y}} \quad & U\left(\frac{I - \mathbf{p}'\mathbf{y}}{p_0}, \mathbf{A}'\mathbf{y}, \mathbf{y}\right) \\ \text{s.t.} \quad & y_n \geq 0. \end{aligned}$$

Assuming that quantities $\{y_n\}_{n=1}^N$ are continuous, the first order conditions are

$$\sum_{c=1}^C a_{nc} \frac{\partial U}{\partial z_c} - \frac{\partial U}{\partial x} \frac{p_n}{p_0} + \frac{\partial U}{\partial y_n} = 0 \quad \text{if } y_n > 0.$$

The model we propose nests various models considered in the literature: discrete choice and hedonics on one hand and demand models in product space on the other. First, suppose the utility function is $U(x, \mathbf{z})$, which is the case in discrete choice models or in hedonic models. Because the transformation from products to characteristics is linear and in this case $\partial U / \partial y_n = 0$, at most C of the N products would be purchased. If we restrict $y_n \in \{0, 1\}$ and $\sum_{n=1}^N y_n \leq 1$, the model collapses to the standard discrete choice model. In general, the prediction that at most C products are purchased is a problem since we would like to consider cases where the number of products chosen is (much) greater than the number of observed characteristics.

Alternatively, if the utility function is $U(x, \mathbf{y})$ then we can generate standard demand systems in product space, such as Cobb-Douglas, CES, Translog and the Almost Ideal Demand System. Once we allow for a characteristic that is product specific then a model in characteristics space is equivalent to a model in product space, as long as the characteristics do not vary over time or markets. Note, that we need more than just different values on a small number of unobserved characteristics, but a totally different characteristic that can *only* be obtained from each product. A model with such a large number of characteristics would be intractable in many applications, where the number of products considered is large thus generating a serious dimensionality problem. In addition, for our purpose, a model in only product space would not allow us to incorporate differences in the characteristics and availability of products across countries.

To better understand the role of the characteristics in our model we can rewrite the first order conditions for n such that $y_n > 0$ as

$$\frac{\partial U/\partial y_n}{\partial U/\partial x} = \frac{p_n}{p_0} - \sum_{c=1}^C a_{nc} \frac{\partial U/\partial z_c}{\partial U/\partial x}.$$

Consider the case where characteristics do not enter the utility, i.e., $\partial U/\partial z_c = 0$. The first order conditions, in this case $\frac{\partial U/\partial y_n}{\partial U/\partial x} = \frac{p_n}{p_0}$, implicitly define the demand correspondence. Indeed, under invertibility conditions, we can write the (Marshallian) demand function as $Q(\mathbf{p}; \eta_i)$. A similar idea applies in our model. Demand depends on the *hedonic prices* of each good instead of prices. The hedonic prices, $\frac{p_n}{p_0} - \sum_{c=1}^C a_{nc} \frac{\partial U/\partial z_c}{\partial U/\partial x}$, depend on the marginal utility of the consumer from the characteristics. Indeed if $\frac{\partial U/\partial z_c}{\partial U/\partial x}$ are constants then one can simply plug the hedonic prices in the demand function in lieu of prices. If the marginal utility from a characteristic is positive then a consumer will adjust the price downward. In other words, if two products have the same price but one has more of a characteristic, with a positive marginal utility, then the effective price to the consumer will be lower for the product with the higher value of the characteristic.

We focus on a particular functional form for utility. We divide the large number of products into J food groups each with K_j products for each group $j \in \{1, \dots, J\}$. Each product n is now labelled kj if it is the k^{th} food item of food group j (where $k \in \{1, \dots, K_j\}$). We do this in order to keep the model tractable, and to work at a level of products that are readily comparable across countries, but we base the food group on conceptually sound aggregation from the individual items. The aggregation assumptions are thus transparent and we can assess which parameters can (and cannot) be identified. There is conceptually no problem with working at a lower level of aggregation, it is an economic question of what level is most meaningful for the particular application.

We assume that utility is given by:

$$U(x_i, \mathbf{z}_i, \mathbf{y}_i; \eta_i) = \prod_{j=1}^J \left(\sum_{k=1}^{K_j} f_{ikj}(y_{ikj}) \right)^{\mu_{ij}} \prod_{c=1}^C h_{ic}(z_{ic}) \exp(\gamma_i x_i) \quad (1)$$

where γ_i is the marginal utility of income, $z_{ic} = \sum_{j=1}^J \sum_{k=1}^{K_j} a_{kj,c} y_{ikj}$, $f_{ikj}(y_{ikj})$ and $h_{ic}(z_{ic})$ are individual specific utility functions that give the utility from products within a food group and the utility from nutrients respectively. If we assume that $h_{ic}(z_{ic}) = z_{ic}^{\beta_c}$ then the

utility from food groups and nutrients is Cobb-Douglas. The utility from products within a group can take different forms. One particular function that is easy to work with is the CES function $f_{ikj}(y_{ikj}) = \lambda_{ikj} y_{ikj}^{\theta_{ij}}$.

This specification of utility allows for a great deal of heterogeneity across goods and individuals. However, a serious limitation of the specific functional forms, relative to more flexible continuous demand system models, is that the way that income and prices affect quantity is limited. Expenditures on each food group will depend on individual parameters, but will not vary with an individual's income. In the estimation we can control for heterogeneity *across* consumers in income, but the functional form imposes restrictive income effects *within* a consumer as income changes. Similarly, the model limits the nonlinearities in price effects. As we will see below, while restrictive in limiting nonlinearities and interaction terms in prices and income, this functional form facilitates both aggregation across goods, and a tractable linear estimation equation that allows for the inclusion of substantial unobserved heterogeneity using a large number of fixed effects.

The way we define utility from products follows a long tradition in demand analysis of assuming weak separability across product groups when defining consumer preferences (Gorman, 1959, and follow up work). Denoting the vector of products $\mathbf{y}_i = (\mathbf{y}_i^1, \dots, \mathbf{y}_i^J)$ where $\mathbf{y}_i^j = (y_i^{j,1}, \dots, y_i^{j,K_j})$, we assume that the utility function satisfies $U(x_i, \mathbf{z}_i, \mathbf{y}_i; \eta_i) = U(x_i, \mathbf{z}_i, W_1(\mathbf{y}_i^1), \dots, W_J(\mathbf{y}_i^J); \eta_i)$, where $W_j(\cdot)$ are subutilities that are a function of the subvector \mathbf{y}_i^j . Without taste for overall nutrition characteristic \mathbf{z}_i , the utility function would be weakly separable across groups. However, entering characteristics into the utility function directly breaks this weak separability and generates more general preferences over products. We have a functional form which is weakly separable across groups conditional on indices $\mathbf{A}'\mathbf{y}$. We impose a sort of “characteristic contingent weak separability” across groups, because demand is weakly separable across bundles of goods \mathbf{y}^j for vectors $(\mathbf{y}^1, \dots, \mathbf{y}^J)$ in the sets $Y(\mathbf{z}_0) = \{\mathbf{y} | \mathbf{A}'\mathbf{y} = \mathbf{z}_0\}$ for any vector of values of characteristic \mathbf{z}_0 . Products from different food groups that have a non-zero amount of a characteristic will interact with each other through the utility from the characteristic, and not just through the group subutilities. This allows for a tractable way to relax the weak separability assumption.

The effect of relaxing weak separability is related to the concept of latent separability in Blundell and Robin (2000). However, our model is not nested within, nor does it nest, latent separability. In latent separability the subutilities, $W_j(\cdot)$, are defined over vectors $\tilde{\mathbf{y}}_j$

of size N , where $\sum_{j=1}^J \tilde{\mathbf{y}}_j = \mathbf{y}_j$. The subutilities can be thought of as utilities from various (latent) activities, each of which require a (non exclusive) subset of the products make. The total amount of each product consumed is a summation over the amount required for each activity. Weak separability is broken because products from different groups can interact through different subutilities. Like here, weak separability is generalized. However, the way weak separability is generalized in Blundell and Robin (2000) is different from our model.

Maximizing utility subject to budget constraint yields the following first order conditions:

$$\mu_{ij} \frac{f'_{ikj}(y_{ikj}) y_{ikj}}{\sum_l f_{ijl}(y_{ilj})} + \sum_c a_{kj,c} y_{ikj} \frac{h'_{ic}(z_{ic})}{h_{ic}(z_{ic})} = \gamma_i \frac{p_{kj}}{p_0} y_{ikj}.$$

Summing the first order conditions over k for a given j :

$$\mu_{ij} \frac{\sum_k f'_{ikj}(y_{ikj}) y_{ikj}}{\sum_k f_{ikj}(y_{ikj})} + \sum_c \frac{h'_{ic}(z_{ic})}{h_{ic}(z_{ic})} \sum_k a_{kj,c} y_{ikj} = \gamma_i \sum_k \frac{p_{kj}}{p_0} y_{ikj}.$$

Using $f_{ikj}(y_{ikj}) = \lambda_{ikj} y_{ikj}^{\theta_{ij}}$ and $h_{ic}(z_{ic}) = \exp(\beta_c z_{ic})$, this expression can be simplified to:

$$\sum_k p_{kj} y_{ikj} = p_0 \frac{\mu_{ij} \theta_{ij}}{\gamma_i} + \sum_c p_0 \frac{\beta_c}{\gamma_i} \sum_k a_{kj,c} y_{ikj}. \quad (2)$$

Moving to the empirical specification we introduce a time subscript t as we are using panel data. Quantities and prices vary over time, and as prices for a unique good may vary across markets we also introduce an individual subscript to price.

3.2 Estimation

Our estimating equation comes directly from equation (2). This allows us to define the error term from the theory and directly introduce unobserved heterogeneity of preferences. Following the recent literature in Industrial Organization, we assume one of the characteristics, indexed $c = 1$, is unobserved. Introducing a time subscript t , let $p_0 \frac{\mu_{ij} \theta_{ij}}{\gamma_i} + p_0 \frac{\beta_1}{\gamma_i} \sum_k a_{kj,1} \times y_{ikjt} = \delta_{ij} + \xi_{jt} + \varepsilon_{ijt}$. We normalize $p_0 = 1$ and $\gamma_i = 1$. The normalization of the price of the outside good p_0 will have to be taken into account when we consider counterfactual experiments that change real versus nominal prices. We return to this point below.

Our estimating equation is

$$w_{ijt} = \sum_c \beta_c z_{ijct} + \delta_{ij} + \xi_{jt} + \varepsilon_{ijt} \quad (3)$$

where $w_{ijt} = \sum_k p_{ikjt} y_{ikjt}$, is the expenditure on food group j by household i at period t , and $z_{ijct} = \sum_k a_{kj,ct} y_{ikjt}$ is the amount of nutrient c household i gets from group j at t .

The combined error term, $\delta_{ij} + \xi_{jt} + \varepsilon_{ijt}$, captures elements of preferences and the environment. One could imagine that preferences for food groups vary across households. For example, some households might derive more utility from vegetables than other households. The household-category effects, δ_{ij} , are meant to capture this. In addition, the products could have an unobserved attribute that varies over time. For example, fruit might be of higher quality during the summer months. The category-quarter effects, ξ_{jt} , will capture this. Finally, the term ε_{ijt} will capture interactions between these effects and could include preference shocks (if preferences are not fixed over time, or if these shocks are, for example, due to changes in unobserved physical activity) and variation over time in the unobserved characteristic.

If ε_{ijt} includes either (changes in the) unobserved characteristics of the goods or preference shocks it will likely impact the choice of quantities of products chosen. This raises a potential concern about endogeneity of the nutrient, z_{ijct} . Even if we allow for household-category, δ_{ij} , and category-time, ξ_{jt} , fixed effects, there remain shocks ε_{ijt} at the household-category-time level that might be correlated with quantity choice and hence with z_{ijct} .

To see the problem, assume that $K_j = 1$, i.e., there is a single product within each category, and for simplicity assume that there is a single nutrient, say carbohydrates. The estimating equation becomes

$$p_{ijt} y_{ijt} = \beta a_{j,ct} y_{ijt} + \delta_{ij} + \xi_{jt} + \varepsilon_{ijt}.$$

In words, we regress the expenditure of product j on the carbohydrates from product j . Consider the variation in the quantity of carbohydrates. This will in part be due to changes in y_{ijt} and in part due to changes in $a_{j,ct}$. As we discussed above, the error term ε_{ijt} consists of random preference shocks and of variation in the utility from unobserved attributes, which

will likely be correlated with y_{ijt} . Therefore, it is quite likely the quantity of carbohydrates from product j , z_{ijct} , will be endogenous.

To account for endogeneity of the z 's we exploit the variation of available products, and their prices and attributes, due to exogenous reasons. The variation in products and their attributes can be due to entry or exit of products or to changes in the market structure, say due to entry and exit of stores. We have a very rich set of controls in the model that account for heterogeneity in preferences, so when we say that available products are exogenous, we mean that they are exogenous conditional on the controls. For example, whether a particular product is offered, is likely correlated with the preferences of households in the market. However, we are able to control for these preferences and look at the effect of changes in product attributes.

The linearity of the estimating equation implies that we can use standard linear panel and instrumental variable (IV) methods. A key challenge for us is how to generate individual variation in the instruments. Ideally we would observe the actual availability of products in the stores where the household shops, and use this availability as an IV. Unfortunately, we only observe products that are actually purchased by some households in our data and not the complete set of available products. We therefore approximate the nutrients of products available to each household by computing the (unweighted) average nutrient content of products purchased, in that category and quarter, by households in a “reference” group.

In order to define the reference group we compute for each household the retailer where they shop for each category most frequently. We then define the reference group as other households in the local area who do their shopping in the same retail chain. Note that the reference group for each household will vary by category. We then compute an (unweighted) average nutrient content of products bought by members of the reference group. We think of this as the average nutritional content of the products in the household’s choice set. This variable will vary by household, category and time, and as we will see below is highly (conditionally) correlated with the endogenous variable. Our identifying assumption is that the variation in this average, conditional of the household-category and category-time fixed effects, is uncorrelated with the error term.

Formally, denote by $A_{h(i)jt}$ the choice set of products in category j for household type (the “reference” group) $h(i)$ in period t . We use the average nutritional content of the choice set, $\omega_{h(i)jct} = \frac{1}{\#A_{h(i)jt}} \sum_{k \in A_{h(i)jt}} a_{kj,c}$, as instrumental variables. Note, that these variables,

one for each nutrient, will depend on the definition of i 's reference group $h(i)$ and vary across periods, household type $h(i)$ and category because of the variations in the choice set $A_{h(i)jt}$. Our identifying assumption is that for $c = 1, \dots, C$

$$E(\varepsilon_{ijt} | \omega_{h(i)jct}, \delta_{ij}, \xi_{jt}) = 0.$$

It requires that, conditional on household-category and category-period fixed effects, (i) ε_{ijt} is not correlated with which products (UPCs) are bought by same type of households, and (ii) that the (changes) in the unobserved characteristic of category j , $a_{kj,1}$, is uncorrelated with other characteristics $a_{kj,c}$. As we will see below, these instrumental variables $\omega_{h(i)jct}$ are highly correlated with $z_{ijct} = \sum_{k \in A_{h(i)jt}} a_{kj,c} y_{ikjt}$, and thus are quite powerful instruments.

4 Empirical Results

4.1 Demand Estimates

In Table 6 we report the estimated coefficients for the demand equation described by equation (3), estimated separately for each country. An observation in the estimation is a household-category-quarter, where we define the nine categories as in Section 2 and describe them in the Appendix. The dependent variable in all the regressions is the expenditure in dollars per adult equivalent for a household in a quarter and category. We include the three macronutrients - carbohydrates, proteins and fats - all measured in kilograms. Data limitations do not allow us to introduce more detailed nutrients, such as different types of fats or proteins from different sources. To capture this heterogeneity in the nutrients we allow the effect of the protein and fat characteristics to vary with the food category. Proteins and fats in the dairy and meat category are largely from animal sources, prepared foods are a mix of animal and vegetal food products, and other categories, such as fruit, vegetables and grains, are primarily vegetal sources.

One might think that preferences are also over total calories, and not just macro nutrients. The caloric content of food is comprised of energy in the form of these three macro-nutrients; total calories is a weighted sum of grams of these macronutrients.¹² The estimates in Table

¹²Approximate weights are 4 calories for each gram of carbohydrates, 4 calories for each gram of proteins and 9 calories for each gram of fats.

6 allow us to check if households only care about calories. The coefficients on the nutrients are (mostly) statistically significant and different from these weights, they differ across countries, and we will see in the counterfactual analysis that varying them has an economically significant impact.

Table 6: Demand Estimates: preferences for nutrients

	OLS - Fixed Effects			IV - Fixed Effects		
	FR	UK	US	FR	UK	US
Carbohydrates	3.484*** (0.0240)	2.913*** (0.0236)	1.311*** (0.0145)	1.483*** (0.554)	1.751*** (0.219)	1.459*** (0.356)
Proteins						
Dairy and Meat	37.01*** (0.0972)	27.17*** (0.0856)	26.70*** (0.114)	23.62*** (1.660)	18.09*** (0.958)	19.79*** (2.327)
Prepared	46.93*** (0.358)	46.07*** (0.182)	59.11*** (0.230)	16.21*** (4.979)	18.36*** (3.015)	51.84*** (1.502)
Other	19.83*** (0.205)	18.10*** (0.189)	19.39*** (0.213)	2.948 (4.808)	2.660 (1.729)	-1.058 (2.306)
Fats						
Dairy and Meat	8.497*** (0.137)	6.424*** (0.104)	-1.810*** (0.105)	4.376*** (1.447)	1.315** (0.610)	1.184 (0.751)
Prepared	12.80*** (0.171)	8.793*** (0.113)	1.562*** (0.0941)	10.94*** (1.250)	10.56*** (0.592)	-2.224*** (0.715)
Other	4.509*** (0.0432)	5.835*** (0.0854)	3.359*** (0.0681)	1.748*** (0.606)	3.707*** (0.599)	1.661*** (0.517)
Validity of IV				340.72	1962.72	337.06
Weak IV				113.63	656.54	48.19
Observations	657,822	654,736	423,976	657,822	654,736	423,976
R-squared	0.700	0.610	0.532			

Note: The dependent variable is the expenditure in dollars per adult equivalent by a household in a category-quarter over 2005-2006. All regressions include household-category and category-quarter fixed effects. The IV results use the (unweighted) average nutrient content in the reference group by category and quarter, defined as household purchasing in the same store in the local area, as the instrumental variable. All standard errors are clustered at the household-food category levels. The validity of IV test is an LM test of whether the excluded instruments are correlated with the endogenous regressors; a rejection of the null indicates that the model is identified. The weak IV test is the Cragg-Donald statistic.

The first three columns present estimates from fixed effects OLS regressions; each regression uses data from the indicated country. All the regressions include household-category and category-time fixed effects. The former control for household specific tastes for particular products, while the latter control for category-specific seasonal effects. The coefficients are identified from within household-category variation, i.e. the correlation between nutrient content and expenditure within a category (and household) over time. Most of the coefficients are statistically significant and positive.

As we previously discussed, these results potentially suffer from endogeneity. Therefore, in the next three columns we examine the results from instrumental variable regressions. The instruments we use are the (unweighted) average nutrients of the products purchased in each quarter by households shopping in the same retail chain in the same area. They aim to capture the variation in the attributes of available products. The regression also includes category-time fixed effects, to control for category specific seasonal effects, and household-category effects, to control for heterogeneity in preferences. This wipes out a significant fraction of the variation in the instruments. However, as we can see from the standard errors, sufficient variation is left and as we see in the bottom of the table the first stage F-statistic of the excluded IVs is very high.

The coefficients are mostly significant, with some interesting differences across countries. The French have the highest relative preference for fats and proteins in dairy and meat. And the Americans have the highest preference for proteins in prepared food, and the lowest for fats in prepared foods. The ratio of the fats coefficients to the carbohydrates coefficient is the highest in France and the lowest in the US, while the ratio of proteins to carbohydrates tends to be higher in the US compared to France and the UK (this is mostly driven by the coefficient for the prepared category).

In Table 7 we report the average household-category and category-quarter fixed effects, averaged across households and quarters, within the nine categories. These fixed effects could be treated as preference parameters, or as unobserved country-specific attributes of the products. For each country these are,

$$\bar{\sigma}_j = \frac{1}{IT} \sum_{it} \hat{\sigma}_{ijt} = \frac{1}{IT} \sum_{it} \left(w_{ijt} - \sum_c \hat{\beta}_c z_{ijct} \right)$$

where w_{ijt} , $\hat{\beta}_c$ and z_{ijct} are the dependent variable, the estimated coefficients and the regressors from the regression defined in equation (3) for each country, and I and T are the total number of households and periods. These can be negative; a negative number suggests that households are purchasing this product for its nutrient characteristics. Indeed, when the value of the coefficients on nutrients is larger, as in the OLS fixed effects regressions, we see more negative numbers.

The numbers are consistent with the story we told above. US households have a higher preference for prepared foods and drinks, both high in carbohydrates, while the French have

a higher preference for dairy and meat, higher in protein. This is one factor that explains why the US households purchase more carbohydrates, while the French households purchase more protein.

Table 7: Demand Estimates: preferences for categories

	OLS - Fixed Effects			IV - Fixed Effects		
	FR	UK	US	FR	UK	US
Fruits	21.78	35.10	29.40	26.92	38.77	31.19
Vegetables	31.04	31.85	28.49	40.77	41.03	32.93
Grains	-3.59	-7.82	7.18	16.49	18.41	23.54
Dairy	-8.46	5.60	10.21	24.66	25.35	14.87
Meat	28.44	10.64	18.56	73.40	38.25	28.89
Oils	3.08	0.05	2.75	10.71	3.11	5.39
Sweeteners	-1.98	-2.30	2.15	2.53	0.50	1.77
Drinks	20.87	21.06	36.91	24.33	22.67	37.31
Prepared	15.48	29.23	51.01	54.75	73.81	71.45

Notes: The table reports the average of the household-category and category-quarter fixed effects across households and quarters, within the nine categories.

4.2 Counterfactual Analysis

To explore the role that differences in prices and product attributes play in the observed differences in the nutritional content of purchased food we simulate the behavior of households from one country if faced with prices and attributes from the other countries. In the discussion below we focus on US households, and ask what would US households purchase if faced with French and UK prices and product attributes. We start by examining the behavior of the average household, and then explore heterogeneity in the response by separately considering low income households and households that purchase a high calorie basket of goods. We focus on these two groups as they are of particular policy interest, and groups for which we might expect different behavior.

In all cases we view the simulation as leaving the US household in place and changing only their shopping environment, namely the prices they face and attributes of products offered to them, and asking what would they purchase. In particular, we envision the household's income as staying constant, which is important given the limitations we discussed above in how our functional form can capture income effects. We describe how expenditure and nutrient patterns compare in these counterfactuals to actual purchases made by the average household in the US, France and the UK.

In defining preferences we always use the estimated slope coefficients, $\widehat{\beta}_1, \dots, \widehat{\beta}_C$ from the home country, which in all but a few simulations below is the US. It is less clear how to think of the category fixed effects. In principle these could be treated as preference parameters, or we could treat them as unobserved country-specific attributes of the products. For example, if the average expenditure in France on meat is high, this could be because French households have a higher preference for meat or because meat is of higher quality. We explore the extreme cases where these fixed effects are either all attributed to preferences or all attributed to unobserved product characteristics.

We start with the “average” household and computing their purchases in the home country. This average household reproduces exactly the average consumption by category observed in the data. The implied category fixed effects, $\bar{\sigma}_j^H$ for country H , that imply this choice, given prices and attributes in each country, are given by:

$$\bar{\sigma}_j^H = \bar{y}_j \left[\bar{p}_j^H - \sum_c \widehat{\beta}_c^H \bar{a}_{jc}^H \right] \quad \text{for } H \in \{US, FR, UK\}$$

where $\bar{y}_j = \frac{1}{IT} \sum_{it} y_{ijt}^H$, $\bar{p}_j^H = \frac{1}{IT} \sum_{it} p_{ijt}^H$ and $\bar{a}_{jc}^H = \frac{1}{IT} \sum_{it} a_{ijct}^H$.

The counterfactual simulations we consider are defined by five variables, which can take the value estimated for any of the three countries. We label these V_1, \dots, V_5 : product attributes (V_1), food prices (V_2), category effects (V_3), nutrient preference parameters (V_4) and a conversion factor to reflect the price of the outside good (V_5).

In our model, simulated quantities purchased of each category j are given by¹³

$$\widehat{y}_j^{(V_1 \dots V_5)} = \frac{\bar{\sigma}_j^{V_3}}{\bar{p}_j^{V_2} / \hat{\tau}^{V_5} - \sum_c \widehat{\beta}_c^{V_4} \bar{a}_{jc}^{V_1}} = \frac{\hat{\tau}^{V_5} \bar{\sigma}_j^{V_3}}{\bar{p}_j^{V_2} - \hat{\tau}^{V_5} \sum_c \widehat{\beta}_c^{V_4} \bar{a}_{jc}^{V_1}} \quad V_i \in \{FR, UK, US\}, \quad i = 1, \dots, 5 \quad (4)$$

where $\hat{\tau}^{V_5} = p_0^{V_5} / p_0^{US}$ is a conversion factor that allows us to convert nominal prices into real prices. We use information from the Penn World Tables on the “price level of consumption” to proxy the price of the outside good. We find that the price of the outside good in France is 7.9% higher than in the US (implying $\hat{\tau}^{FR} = 1.079$), while it is 8.9% higher in the UK than in the US ($\hat{\tau}^{UK} = 1.089$).

¹³We take the max of $\widehat{y}_j^{(V_1 \dots V_5)}$ and 0, to deal with corner solutions. This constraint does not bind for our preferred specification.

We examine five scenarios that amount to assuming different values for V_1, \dots, V_5 . An important input into the simulation is the hedonic price $\bar{p}_j^{V_2} - \hat{\tau}^{V_5} \sum_c \hat{\beta}_c^{V_4} \bar{a}_{jc}^{V_1}$. In Tables A3 and A4 in the Appendix A.6 we show the hedonic prices under different scenarios. These prices will be informative when we consider the simulations below. The counterfactual scenarios we consider are:

Scenario A: hold the quantity that the average US household purchases fixed, but allow the products to have the average attributes from France or the UK. This scenario simulates the effect of changing the nutrients of food products while holding food choices constant, which allows us to separate the effect of a change in attributes on choice from the effect of a change in nutrients holding choice constant.

Scenario B: preferences and product attributes are those of the average US household but food prices are as in France or the UK. In this case quantities are given by equation (4), with $V = (US, FR, US, US, US)$ for France and equivalently for the UK. This scenario isolates the effect of prices. Choices are allowed to change according to the model, but the assumption is that the product attributes do not change (they remain as in the US).

Scenario C: preferences are those of the average US household but food prices and product attributes are as in France or the UK. Quantities are given by equation (4), with $V = (FR, FR, US, US, US)$ for France and equivalently for the UK. This scenario simulates the effect of the change of both prices and product attributes to those available in France and the UK.

Scenario D: preferences are those of the average US household but food prices and product attributes are as in France or the UK, as well as the price of the outside good. In other words, the prices are changed to the real prices in France or the UK, as opposed to nominal prices as in Scenario C. In this case quantities are given by equation (4), with $V = (FR, FR, US, US, FR)$ for France and equivalently for the UK.

Scenario E: preferences are those of the average US household but food prices, product attributes, and the price of the outside good are as in France or the UK, as well as the mean category effects $\bar{\sigma}_j^V$, which could reflect some characteristics of goods and aspects of the economic environment and not only preferences. In this case quantities are given by equation (4), with $V = (FR, FR, FR, US, FR)$ for France and equivalently for the UK.

In the simulations, prices and product attributes are the quantity weighted averages in country V for category j : we are not simulating the choices of the disaggregated quantities

y_{ijkt} , only the quantity at the category level. There are two ways to view our simulations. First, we can consider the category j as a homogenous or single good. In that case the simulation is directly linked to the theory. An alternative is to acknowledge that each category is an aggregate over heterogenous products, but to assume a two stage maximization problem. The simulated household takes the choice of the products within each of the categories as given, and then chooses how much to purchase of each category. This is not the same as the solution to the maximization problem we present in the theory section.

The reason we need to conduct the simulations at the aggregated level is twofold. First, to simulate quantities at a disaggregated level would require estimating many parameters. Given the number of products we have, this is not feasible to do at the level of narrowly defined products; the problem thus requires some aggregation. Second, even if we could estimate the parameters at a very disaggregated level, we could not use these estimates directly since very narrowly defined products are very different in the three countries. In order to use the parameters we would need to average prices and attributes to a level that is comparable across the different countries.

4.2.1 US Households Facing French Prices and Attributes

We start by considering the purchasing behavior of a household with preferences of the average US household facing French prices and product attributes. Table 8 shows the predicted change in calories and macronutrients purchased by the average US household under scenarios A to E described above. In column (1) we show the simulated purchases by a household with average US preferences facing US food prices and attributes. In column (7) we show the simulated purchases by a household with average French preferences facing French prices and attributes. Columns (2)-(6) show the simulated purchases for scenarios A-E.

Focusing first on the row showing total Calories, comparing column (2) to column (1) we see that holding quantities constant at the level of the average US household but using (average) French nutrient content has little impact on total calories. There is, however, an impact on the fraction of calories coming from the different nutrients, moving the US percentages closer to the French ones. Moving to column (3), where we use French food prices but keep nutrients as in the US, has a substantial impact, leading the average US household to substantially reduce the calories they purchase (to 85% of the level in column (1)). Indeed the total calories is very similar to those of the average French household (displayed in the

last column). However, this average US household would spend 12% less than the French for approximately the same total calories and the fraction of calories coming from each macro nutrient is very different and changes in the “wrong” way. Column (4) considers the average US household facing French food prices and product attributes, leading to a slight further reduction in calories to 84% of the amount in column (1).

Table 8: US Households Facing French Prices and Attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
scenario:		A	B	C	D	E	
attributes (V_1)	US	FR	US	FR	FR	FR	FR
prices (V_2)	US	US	FR	FR	FR	FR	FR
category effects (V_3)	US	US	US	US	US	FR	FR
nutrient pref (V_4)	US	US	US	US	US	US	FR
price adjustment (V_5)	1	1	1	1	1.079	1.079	1
Calories	2212.3	2158.3	1884.1	1839.9	2088.5	1946.6	1873.4
Expenditure	5.41	5.41	4.75	4.94	5.59	5.51	5.38
Carb (cal)	1092.6	903.0	1166.6	949.4	1070.5	766.6	709.6
	49.4	41.8	61.9	51.6	51.3	39.4	37.9
Prot (cal)	279.40	326.66	171.57	213.18	243.26	287.91	299.90
	12.6	15.1	9.1	11.6	11.6	14.8	16.0
Fat (cal)	840.3	928.7	545.9	677.3	774.8	892.1	863.9
	38.0	43.0	29.0	36.8	37.1	45.8	46.1
% calories							
Fruits	4.7	4.1	5.5	4.7	4.5	4.2	4.1
Vegetables	3.0	3.5	3.6	4.2	4.1	5.2	5.2
Grain	13.0	15.2	14.5	17.4	16.9	12.3	13.3
Dairy	10.2	12.5	7.0	9.9	10.2	13.9	17.8
Meats	16.0	14.5	6.5	6.4	6.4	14.4	16.1
Oils	5.7	6.8	5.6	6.8	6.6	12.8	12.5
Sweeteners	3.3	3.1	8.5	7.5	7.6	4.6	4.2
Drinks	8.0	4.6	17.8	9.6	9.2	4.7	4.5
Prepared	36.2	35.7	31.0	33.4	34.4	27.9	22.3

Notes: Figures are per adult equivalent per day. Expenditure is in US\$.

In column (5) we consider a change of the prices facing US households to the real French prices, by accounting for differences in the relative price of the outside good. Since the price of the outside good is higher in France this makes food more attractive, and leads to an increase in total calories. In column (6) we consider the scenario where we alter food prices, product attributes, the price of the outside good and change the category effects to the French values. Category effects capture household-category and category-time fixed effects.

As previously discussed, these probably combine preferences and aspects of the economic environment. Once we include these the level of calories purchased is similar to the average French household, around 4% higher, and around 12% lower than column (1). Interestingly, in this last scenario not only are total calories similar to French level, but the fraction of calories coming from different macro nutrients is more similar, as is the level of expenditure.

In comparing the total calories purchased by the average US and French household it appears that food prices can explain much of the difference (scenario B), but not all. However, this conclusion while true for calories, is overly simplistic. If we look at the macronutrients purchased we see that these remain different. The average US household purchases more carbohydrates and less proteins and fats than the average French household when faced with French prices and attributes (columns (2) - (5)). It is only when we change the category effects that the balance of macro nutrients looks more like the French diet.

Consideration of Tables 5 and A.3, in the Appendix, shows us in part where these results come from. Comparing columns (1) and (3) of Table A.3 we can see that the hedonic price of many of the large food categories for a household with US preferences but facing French food prices is higher than in the US, e.g. Dairy, Meats and Prepared Foods, and so the average US household reduces the quantity purchased of each of these. However, this is not true of all categories. Fruits, Vegetables, Sweeteners and Drinks are all cheaper, and so the average US household increases the quantity purchased in these categories. In some categories this change in quantity is offset by differences in product attributes (shown in Table 5). So, for example, the quantity of Fruit purchased by the average US household increases when faced with French food prices, but the difference in product attributes means that this represents a reduction in calories. This is also true for Sweeteners and Drinks. For Dairy products it goes in the other direction, the higher hedonic price means that the average US household purchases less Dairy when faced with French food prices, but the difference in product attributes means that this represents an increase in calories. This is also true for Oils. For the other categories the change in quantities goes in the same direction as the change in calories. The biggest change comes from Meats. The hedonic price is substantially higher for the average US household facing French food prices and product attributes, and this leads to a big reduction in Meat purchased, and the average attributes of Meat in France mean that it has fewer calories.

While the total share of macronutrients purchased by the average US household when faced with French food prices and attributes is similar to that when they face US prices and attributes, the food categories that these nutrients come from differs substantially. For example, a US household with average US preferences faced with US prices and attributes obtains 37% of proteins from Meats, 27.4% from Prepared foods, 14.5% from Dairy and 13.9% from Grains. When faced with French food prices and attributes they obtain only 14% of proteins from Meats (because Meat is much more expensive in France they purchase less), 36% from Prepared foods (because prepared foods in France contain more protein), 15.3% from Dairy and 20% from Grains.

4.2.2 US Households Facing UK Prices and Attributes

We now consider a household with the average US preferences facing UK food prices and attributes. Table 9 is laid out as Table 8, in column (1) we show the simulated purchases by a household with average US preferences, and in column (7) by a household with average UK preferences. Columns (2)-(6) show the simulated purchases for scenarios A-E.

These results are quite different to those for the average US household facing French prices and attributes. Comparing columns (1) and (2) we see that using US quantities and UK nutrient content leads to a substantial reduction in calories. Products in the UK tend to have less carbohydrates, about the same amount of fats, but more protein. Thus, in total holding quantities fixed leads to a decrease in calories, due to the decrease in fats and carbohydrates, but an increase in proteins.

Comparing columns (1) and (3) we see that the effect on the average US household of facing UK food prices is to purchase 6% more calories, and increase purchase of all the nutrients. This is the opposite of the impact of product attributes. In Section 2.3 we saw that a Laspeyres index, holding quantities fixed at US levels, finds the UK prices are higher than US prices. Here, however, we find that when faced with UK prices, US households purchase more calories not less. This is because some categories are cheaper in the UK, and have a lower hedonic price, leading to increase consumption in these categories and increase in total calories. Column (4) combines the two effects. This leads to a reduction in the total calories purchased to 97% of the level in column (1), though as we discuss below the products purchased differ substantially. In column (5) we consider also changing the price of the outside good to the UK level, because prices in the UK are higher than in the US this

makes food more attractive and total calories purchased increases substantially. In column (6) we also alter category effects. When faced with this “full” UK environment of food prices, attributes and the price of the outside good and category effects, the average US household would purchase around 7% more calories than when faced with the “full” US environment, and over 20% more than the average UK household. He would also spend considerably more than when faced with the US environment, and more than the average UK household faced with the same environment.

Table 9: US Households Facing UK Prices and Attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
scenario:		A	B	C	D	E	
attributes (V_1)	US	UK	US	UK	UK	UK	UK
prices (V_2)	US	US	UK	UK	UK	UK	UK
category effects (V_3)	US	US	US	US	US	UK	UK
nutrient pref (V_4)	US	US	US	US	US	US	UK
price adjustment (V_5)	1	1	1	1	1.089	1.089	1
Calories	2212.3	2015.3	2336.4	2157.3	2567.6	2372.5	1972.8
Expenditure	5.41	5.41	5.29	5.54	6.55	6.56	5.19
Carb (cal)	1092.6	936.1	1269.9	1095.9	1293.3	1066.5	926.1
	49.4	46.4	54.4	50.8	50.4	45.0	46.9
Prot (cal)	279.40	313.11	270.36	299.69	359.30	387.66	306.79
	12.6	15.5	11.6	13.9	14.0	16.3	15.6
Fat (cal)	840.3	766.1	796.1	761.8	915.0	918.4	739.9
	38.0	38.0	34.1	35.3	35.6	38.7	37.5
% calories							
Fruits	4.7	5.0	2.8	3.0	2.7	3.9	4.4
Vegetables	3.0	3.9	3.2	4.2	3.9	5.1	5.9
Grain	13.0	11.7	19.9	16.1	15.2	12.6	18.4
Dairy	10.2	10.3	12.6	10.3	10.4	15.2	14.3
Meats	16.0	17.2	9.5	11.7	12.0	14.9	13.4
Oils	5.7	6.0	6.4	6.4	6.1	3.4	6.7
Sweeteners	3.3	3.5	10.4	9.9	10.2	2.8	3.8
Drinks	8.0	8.7	4.6	4.9	4.5	2.7	3.1
Prepared	36.2	33.8	30.6	33.6	35.0	39.4	30.2

Notes: Figures are per adult equivalent per day. Expenditure is in US\$.

Again, Tables 5 and A.4 provide some intuition for why these results arise. Comparing columns (1) and (5) we see that the hedonic price of some of the large categories of food are much lower for a household with the average US preferences in the UK than they are in the US, or for a household with the average UK preferences in the UK. In addition,

US households value proteins more than UK households (see Table 6), and this leads to differences in their valuation of products.

As in the simulations for France, the composition of foods purchased differs substantially when a household with the average US preferences faces UK prices and attributes compared to when they are in the US, or compared to a household with the average UK preferences. The average US household facing UK prices and attributes purchases a higher share of their calories in Vegetables and Grains and a lower share in Fruit, Meat and Prepared Foods.

4.3 Distributional Implications

Up to now we have focused on the average household in the simulations. However, the impact of changing prices or product attributes might differ across households. We focus on two different groups of households – low income households and high calorie (per capita) households – that are of interest for somewhat different reasons. Low income households might be of particular interest, since obesity is negatively correlated with income, and by looking at low income households we can also check the sensitivity of the results to income variation. High calorie households are of interest, since they are more likely to be obese, and therefore might be of more direct policy interest.

4.3.1 Low Income Households

In this section we consider low income households. We define low income by computing for each household the adult equivalent income and then focus on the bottom quartile in each country. Households in the lowest income quartile spend less and purchase slightly fewer calories per adult equivalent than higher income households in all three countries (details available in an on-line appendix). They get fewer calories from fruits, vegetables and meats and more from grains. Prices paid by lower income households are lower than prices paid by households in higher income groups, and the attributes of the products they buy differ.

In the top panel of Table 10 we show a selection of the results from the simulated counterfactuals where we use the mean prices and attributes for households in the bottom income quartile in the US and France. Broadly we see a similar pattern to when we considered the average household. However, there are some differences. Simply facing the low income US household with French food prices, Scenario B in column (3), can no longer explain the bulk

of the differences between the low income US and French households. It is only in Scenario E, where we face the low income US household with the French category effects that the level of purchases by low income US households nears the level of the low income French household. In this case the low income US household will purchase around 23% fewer calories than when faced with the US environment, and around 3% higher than the low income French household faced with the same environment.

Table 10: Low Income US Households Facing FR/UK Prices and Attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
scenario:	US	A	B	C	D	E	FR/UK
FR							
Calories	2128.9	2105.5	1805.9	1775.4	2057.1	1639.0	1594.1
Expenditure	4.26	4.26	3.62	3.82	4.41	3.69	3.67
Carb (%)	50.5	42.7	64.9	53.2	52.8	41.3	39.3
Prot (%)	12.0	14.4	8.1	10.7	10.8	13.4	15.0
Fat (%)	37.5	42.9	27.0	36.1	36.4	45.3	45.7
UK							
Calories	2128.9	1891.4	2130.3	1954.8	2407.8	2304.4	1841.3
Expenditure	4.26	4.26	4.00	4.25	5.17	5.62	4.23
Carb (%)	50.5	46.7	57.8	53.2	53.1	45.1	47.5
Prot (%)	12.0	15.0	10.3	12.8	12.8	16.0	15.0
Fat (%)	37.5	38.3	32.0	34.0	34.1	38.9	37.5

Notes: Figures are per adult equivalent per day. Expenditure is in US\$.

In the bottom panel of Table 10 we show a selection of the results from the simulated counterfactuals where we use the mean price and attributes for households in the bottom income quartile in the US and the UK. Here as well we see a broadly similar pattern to when we consider the average household. Changing attributes alone, while holding quantities fixed (scenario A), brings the level of calories and the share from each of the macronutrients closest to the UK household. When we change prices and attribute and allow households to reoptimize, the US household always purchases a higher level of calories, with more carbohydrates and less proteins than the average low income household from the UK. As with the average household, when faced with the “full” UK environment of food prices, attributes and the price of the outside good and category effects (column (6)), the average low income US household would purchase around 8% more calories than when faced with the “full” US environment, and over 25% more than the average low income UK household. They would also spend considerably more than when faced with the US environment, and more than the average low income UK household faced with the same environment.

4.3.2 High Calorie Households

We now focus on a different group of households - those that purchase a high (per capita) level of calories. These households are of particular interest to any policy aimed at reducing obesity. We divide households in each country into quartiles based on adult equivalent calories purchased per person per day. Households in this group in the US purchase on average 57% more calories per adult equivalent than the average households, 54% more in France and 43% more in the UK (details available in an on-line appendix). The households in this group also spend more on food than the average household and buy slightly cheaper food. However, on other dimensions, such that the share of calories from macronutrients and from different food categories, households in this group are not noticeably different from the average household.

Table 11 shows a selection of the simulation results where we use the mean prices and attributes for households in the top quartile of adult equivalent calorie purchases. The patterns are broadly the same across the scenarios as they were for the average household, with a few notable exceptions. For example, in column (4) of the top panel we see a reduction of 28% in calories purchased, compared to a reduction of 17% for the average household in Table 8. Similarly, in column (4) of the bottom panel we see a reduction of 10% compared to column (1), while for the average household we saw a reduction of only 3% in Table 9.

Table 11: High Calorie US Households Facing FR/UK Prices and Attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
scenario:	US	A	B	C	D	E	FR/UK
FR							
Calories	3460.0	3239.2	2686.6	2475.2	2805.5	2995.2	2875.0
Expenditure	7.68	7.68	6.47	6.66	7.52	8.65	8.44
Carb (%)	49.5	41.0	62.9	51.3	51.0	38.1	36.6
Prot (%)	12.2	15.3	8.4	11.4	11.4	15.0	16.2
Fat (%)	38.4	43.7	28.6	37.3	37.6	46.9	47.2
UK							
Calories	3460.0	3105.1	3510.9	3119.9	3725.6	3372.6	2810.7
Expenditure	7.68	7.68	7.37	7.62	9.04	9.06	7.15
Carb (%)	49.5	46.4	55.1	51.8	51.4	45.3	47.2
Prot (%)	12.2	15.3	11.1	13.4	13.5	16.1	15.3
Fat (%)	38.4	38.3	33.8	34.8	35.1	38.6	37.5

Notes: Figures are per adult equivalent per day. Expenditure is in US\$.

4.4 Robustness

There are two potentially important limitations to our data, a lack of measurement of food eaten outside of the home, and lack of measurement of food waste.

Eating out. From other data sources we know that the fraction of calories from food outside the home differs across countries. For the US the National Health and Nutrition Examination Survey suggests that 35% of calories are consumed outside the home (USDA, 2010). In the UK, Griffith and O’Connell (2009) use the Expenditure and Food Survey (EFS) for 2005/2006 and find that 12% of calories are eaten outside of the home. In France, Afssa (2009) finds that 18% of total energy consumption intake is outside the home. This suggests that our descriptive analysis underestimates the cross country calorie differences.

The lack of food away from home has several implications for our conclusions. First, in an ideal world we could measure the calories consumed away from home and see how these change when we change prices of food at home. Second, in the simulations of food at home we would want to know the impact of changing the price of food away from home. Note, that due to the large number of fixed effects we would need the price (and availability) of food away from home to vary by household-quarter. Otherwise, it would be collinear with the household or time fixed effects.

We think the best way to interpret our results is as keeping the consumption of food away from home constant and therefore we miss the first effect above. Without more data we have no way to know how big is the substitution between food at home and food away. The impact of the second on the simulations is less obvious, especially once we view our simulations as trying to simulate the effect of changes in prices and/or attributes on food consumption in the US (and not relative to the average UK/France household). The different scenarios we offer, especially Scenario E, can help bound the size of the effect.

Finally, while we think that understanding patterns of food purchase and nutrients outside the home is also important, we believe that food purchased for consumption at home is of interest on its own.

Food waste. Our data measures food purchases, which we equate to consumption. In reality, a fraction of the food purchased is thrown away and never consumed. Estimates on how large this fraction is vary substantially depending on what is being measured. Many of the reported headline figures that you see measure total food waste (e.g. by looking at food

found in dumps); this includes waste from the catering trade and from retail establishments, and is not only food wasted by households. We could not find estimates that measure household food waste and that were comparable across countries. In principle, we could try to apply a fraction of waste, that possibly varies by category and country, in the simulation. However, in order to model this properly we would need to take a stand on whether this fraction is policy invariant or whether it would vary as we change prices. It is plausible that one of the reasons that food waste varies across countries is because of differences in price.

5 Summary and Conclusion

In this paper we document the differences in food purchases made by households in France, the UK and the US. US households purchase more calories than UK households, who purchase more than French households. Furthermore, the source of the calories differs in terms of which macro-nutrients they come from, across the countries. We estimate the determinants of demand for nutrients in each of the countries by extending the demand model of Gorman (1956), that nests classical demand models in product space, as well as models in characteristics space. Our model departs from the weakly separable case of utility functions across different food groups by allowing the marginal utility of each food category to be affected by the amount of nutrients provided by all other foods. It yields a simple linear estimating equation, which relates the expenditures on products to the nutritional content.

The estimates allow us to simulate counterfactual quantities purchased by households with preferences from one country but facing prices and product attributes from another country. We use the simulations to learn about the relative importance of preferences versus the economic environment. We find that, the average US household when faced with French prices and product attributes, will purchase substantially fewer calories, bringing the level close to that of the average French household when faced with the same environment. However, the composition of these calories would differ. The simulated change is mostly due to price differences. In contrast, when we simulate the average US household's food basket with UK product attributes this has a substantial impact on reducing calories, whereas changing relative prices in fact increases calories.

From these findings we conclude that the economic environment makes a substantial difference on the consumption basket. However, in general, it is the interaction of preference, prices and attributes that explains the cross country differences.

In terms of the “healthiness” of preferences and the environment we can attempt to rank countries. The French environment generally encourages healthier purchasing habits. The average US household generally improves her purchasing habits, (while we have not shown this in the above tables) the same is true for the average UK household. The UK environment, on the other hand, generates worse outcomes for the average US household (and the average French household as well). Indeed, the reason the UK households purchase less calories than US households is because of their preferences and *despite* their environment, not because of it.

While we have made significant progress in understanding the role preferences and the economic environment play in explaining cross country differences, significant work still remains. Our model and setup can be used to understand within country differences. For example, within the US there are significant regional differences, as well as differences across demographic groups. How much of these differences can be explained by differences in the economic environment versus preference differences?

Similarly, our model and methods can be used to evaluate various policies. For example, imposing a tax on sugar or on fat. Previous studies have examined these questions but usually in the context of a particular product, say soft drink. We can study the effect more generally, allowing for substitution across products.

Finally, as we discussed above the demand model we offered nests demand models in product space and those in characteristics space. We relied on a particular functional form, but the basic ideas can be extended and used more broadly to generate flexible demand models with more general income and price effects. Furthermore, the basic methods can be used to explore demand at a lower level of aggregation.

6 References

Abaluck J. (2011) “What Would We Eat if We Knew More: The Implications of a Large-Scale Change in Nutrition Labeling” mimeo

Acs Z., L. Henderson, D. Levy, A. Lyles and K. Stanton (2007) "The policy framework for confronting obesity" in Acs and Lyles (eds) *Obesity, Business and Public Policy*, Edward Elgar

Afssa (2002) "La Santé vient en mangeant : Document d'accompagnement du guide alimentaire pour tous destiné aux professionnels de santé", http://www.sante.gouv.fr/IMG/pdf/guide_alimentairepro.pdf

Afssa (2009) "Étude Individuelle Nationale des Consommations Alimentaires 2 (INCA 2) 2006-2007", <http://www.afssa.fr/Documents/PASER-Ra-INCA2.pdf>

Atkin D. (2012) "Trade, Tastes and Nutrition in India", forth. *American Economic Review*

Bajari P. and L. Benkard (2005) "Hedonic Price Indexes With Unobserved Product Characteristics, and Application to Personal Computers" *Journal of Business and Economic Statistics*, 23(1): 61-75.

Bawa S. (2005) "The role of the consumption of beverages in the obesity epidemic", *Perspectives in Public Health*, Vol. 125 no. 3, 124-128

Berry S., J. Levinsohn and A. Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4): 841-890 (July 1995).

Bleich S., D. Cutler, C. Murray and A. Adams (2007) "Why Is The Developed World Obese?" NBER Working Paper 12954

Blundell R. and J. M. Robin (2000) "Latent Separability: Grouping Goods Without Weak Separability" *Econometrica*, Vol. 68, No. 1, pp. 53-84

Bresnahan T. (1981) "Departures from marginal-cost pricing in the American automobile industry : Estimates for 1977-1978," *Journal of Econometrics*, Elsevier, Vol. 17(2), pages 201-227, November

Bound, J., and A. B. Krueger (1991) "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of Labor Economics*, 9(1), 1-24.

CDC (2011) "Obesity, At a Glance," Center for Disease Control and Prevention (available at http://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/2011/Obesity_AAG_WEB_508.pdf)

Chou S., M. Grossman and H. Saffer (2004) "An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System" *Journal of Health Economics*, 23, 565-587

Court A. T. (1939) "Hedonic Price Indexes with Automotive Examples". In *The Dynamics of Automotive Demand*, Ed. Charles F. Roos 99-117. New York General Motors.

Cutler D., E. Glaeser and J. Shapiro (2003) "Why Have Americans Become More Obese?" *Journal of Economic Perspectives* 17 (3), 93-118

Deaton A. (1997) *The Analysis of Household Surveys: A microeconomic approach to development*, The John Hopkins University Press: Baltimore Maryland

Drewnowski A.(2004) "Obesity and the Food Environment: Dietary Energy Density and Diet Costs," *American Journal of Preventive Medicine*, 27(3 Suppl):154-62.

Drewnowski A. and S. Specter (2004) "Poverty and Obesity: the Role of Energy Density and Energy Costs," *The American Journal of Clinical Nutrition*, 79(1):6-16.

- Drewnowski A., P. Monsivais, M. Maillot and N. Darmon (2007) “Low-Energy-Density are Associated with Higher Diet Quality and Higher Diet Costs in French Adults,” *Journal of the American Dietetic Association*, 107(6):1028-32.
- Duffey K.J. and B.M. Popkin (2011) “Energy Density, Portion Size, and Eating Occasions: Contributions to Increased Energy Intake in the United States, 1977-2006” *PLOS Medicine*, Vol.8 Issue 6
- Einav L., E. Leibtag and A. Nevo (2010) “Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?” *Quantitative Marketing and Economics*, volume 8, Number 2, 207-239
- Epple D. (1987) “Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products” *Journal of Political Economy*, 95, 59-80.
- Finkelstein E. and L. Zuckerman (2008) *The Fattening of America: how the economy makes us fat, if it matters, and what to do about it*, John Wiley & Sons: New Jersey.
- French S., M. Story, and R. Jeffery (2001) “Environmental Influences on Eating and Physical Activities,” *Annual Review of Public Health*, 22:309-35.
- Goldman D., D. Lakdawalla and Y. Zheng (2009) “Food Prices and the Dynamics of Body Weight” NBER Working paper 15096, Forthcoming Economic Aspects of Obesity. Chicago: University of Chicago Press
- Gorman T. (1956) “A possible procedure for analysing quality differentials in the egg market”. Reprinted in *The Review of Economic Studies* (1980)
- Gorman T. (1959) “Separable utility and aggregation” *Econometrica* 27, 469-481.
- Gortmaker S., B. Swinburn, D. Ley, R. Carter, P. Mabry, D. Finegood, T. Huang, T. Marsh and M. Moodie (2011) “Changing the future of obesity: science, policy, and action” *The Lancet*, 378: 838-847
- Griffith R. and M. O’Connell (2009) “The Use Of Market Research Data For Research Into Nutrition” *Fiscal Studies*, Vol. 30, no. 3/4, pp. 339-365
- Griffith R. and M. O’Connell (2010) “Public Policy Towards Food Consumption” *Fiscal Studies*, Vol. 31, no. 4, pp. 481-507
- Griliches Z. (1961) “Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change.” In *The Price Statistics of the Federal Government*.
- HMSO (1991) Report of the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Policy (COMA) 1991, The Stationary Office, London
- Lakdawalla D. and T. Philipson (2002) “The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination,” NBER Working Papers 8946
- Lakdawalla D. and T. Philipson (2009) “The Growth of Obesity and Technological Change”, *Economics and Human Biology*, 7, 283-293
- Lancaster K. (1966) “A New Approach to Consumer Theory” *Journal of Political Economy*, 74 (2)
- Leicester A. and Z. Oldfield (2009) “Using Scanner Technology to Collect Expenditure Data” *Fiscal Studies*, Vol. 30, No. 3-4, December 2009, 309-337
- Lewbel A. (1985) “A Unified Approach to Incorporating Demographic or Other Effects into Demand Systems” *The Review of Economic Studies*, Vol. 52, No. 1., pp. 1-18.

- Lewbel A. and K. Pendakur (2011) "Generalized Random Coefficients With Equivalence Scale Applications" Boston College Working Paper
- Lu Y. et D. Goldman (2010) "The Effects of Relative Food Prices on Obesity – Evidence from China: 1991-2006", NBER Working Paper 15720
- Maillot M., N. Darmon, F. Vieux and A. Drewnowski (2007) "Low-Energy-Density are High Nutritional Quality Are Each Associated with Higher Diet Costs in French Adults," *American Journal of Clinical Nutrition*, 86(6):690-696.
- McFadden D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, eds., *Frontiers of Econometrics*, New York, Academic Press.
- National Academy of Sciences (2013) "US Health in International Perspective: Shorter Lives, Poorer Health" http://www.nap.edu/catalog.php?record_id=13497
- National Research Council (2011) Crimmins, Eileen M., Samuel H. Preston, and Barney Cohen (eds), 2011, "Explaining Divergent Levels of Longevity in High-Income Countries", Panel on Understanding Divergent Trends in Longevity in High-Income Countries
- Nevo, A. (2000) "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9(4), 513-548, 2000.
- Philipson T. J. and R. A. Posner (2003) "The Long-Run Growth in Obesity as a Function of Technological Change" *Perspectives in Biology and Medicine* - Volume 46, Number 3 Supplement, Summer 2003, pp. S87-S107.
- Philipson T.J. and R.A. Posner (2008) "Is the obesity epidemic a public health problem? a review of Zoltan J. Acs and Alan Lyles's Obesity, Business and Public Policy", *Journal of Economic Literature*, 46(4), 974-982
- Philipson T. and R. Posner, (2011) "Economics of Obesity: A Decade of Research", forthcoming, *Journal of Economic Literature*
- Rosen S. (1974) "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 34-55.
- Seale J., A. Regmi, J. Bernstein (2003) "The International Evidence on Food Consumption Patterns" USDA Technical Bulletin Number 1904
- Swinburn B., G. Sacks and E. Ravussin (2009) "Increased food energy supply is more than sufficient to explain the US epidemic of obesity" *American Journal of Clinical Nutrition* Vol. 90 No 6 pp. 1453-1456
- UK Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*, Report to the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Policy, London: HMSO.
- US Department of Agriculture (2010) "Away from Home: Percentage of Selected Nutrients Contributed by Foods Eaten Away from Home, What We Eat in America, NHANES 2007-1008," <http://www.ars.usda.gov/Services/docs.htm?docid=18349>.

A Data Appendix

In this appendix we provide details on the construction of the data. The data are collected by market research firms in each of the three countries. In France and the UK, data are collected by Taylor Nelson and Sofres (TNS), now a part of the Kantar Group, as part of their Worldpanel. In the US, data is collected by Nielsen as part of their Homescan data.

The data are collected in a similar manner in all countries. A panel of households use scanners in the home to record all food purchases brought into the home. Participants scan each bar code and record quantity of items purchased; they also record the store of purchase. This information is transmitted to the market research firm. Prices are obtained either directly from the store, if the retailer is part of Nielsen’s store level data, or from the information the participant records. TNS also uses information on till (cashier) receipts to confirm prices and special offers. Thus for each item purchased we know exactly what was bought (as denoted by the barcode or UPC), the quantity purchased, the price paid, and exactly when and where it was bought.

Each participating household collects information on all products with a barcode. Items without barcodes, often called “random weights” items are not recorded by all households. These items include some fruit, vegetables, meat and deli items. The way information on random-weight items is collected varies across countries, as described below.

Information on household demographics are collected through an annual questionnaire.

These data are matched with information on the nutrient content of each food item; this information is collected in a slightly different manner in each country, as explained below. Macronutrients are converted from grams to calories by multiplying grams of carbohydrates by 3.75, grams of proteins by 4, and grams of fats by 9.

Income is measured in brackets for each country. We used the middle values of intervals to approximate household income and then used household equivalence scales to equalize incomes per adult equivalent.

A.1 US data

Each year there are roughly 61,000 participating households out of which a subsample of roughly 15,000 record random-weight purchases. Nielsen monitors the recording and drops households it feels are unreliable. The reliable panel, often called the “static” panel has roughly 40,000 household in total of which 8,000 a year report random-weight purchases.

We start with the static panel and drop quarters where a household did not report expenditure in five or more of the nine food groups.

The Nielsen data does not have nutritional information, this information was collected by Gladson. The Gladson data records information for about 400,000 items, as specified by the barcode or Universal Product Code (UPC). For each item they record essentially everything that is on the box, including the nutritional label, as well as attributes of the box, such as dimensions and weight. To match the Gladson data with Homescan we followed the following steps. About 60% of the UPCs in Homescan had a direct match in the Gladson data. If there was no match we used the average nutrients in the Gladson data within product module (PM)¹⁴, size type, brand, product, flavor, and formula (as defined by Nielsen). This adds roughly another 8% match to a total of slightly over 68%. Many of the remaining items do not match because they are from store brands, which Gladson does not record. For these cases we average within PM, size type, product, variety, type, formula, and style (i.e., drop the brand requirement). This matches another 25 percentage points for a total of roughly 93% match. The rest of the information is mostly for random-weight items, which we manually fill in using the information from the USDA National Nutrient Database for Standard Reference.

¹⁴Nielsen classifies each product into a narrowly defined group called product module. There are over 600 of these groups.

A.2 French data

Each year there are approximately 14,500 participating households. We drop observations that are outliers (below the first percentile or above the 99th, and also households who purchase less than 5 of the food categories over the 9 in a quarter).

Each household is asked to record only certain random-weight categories. For example, households are asked to report purchases of random-weights of either “fruits and vegetables” or “meats and fish”, but not both. We use an imputation method based on the household observable demographics to impute the value of purchase at the quarter level from other households reporting their purchase.

We directly collected nutritional characteristics on macronutrients (calories, proteins, fats and carbohydrates) and matched these with the products purchased¹⁵. The nutritional information come from several sources. We primarily used information collected directly from labels and public sources such as the CIQUAL database (from the public French Information Center on Food Quality) from Afssa (2008) “Table de composition nutritionnelle des aliments Ciquial 2008” <http://www.afssa.fr/TableCIQUAL/>.

A.3 UK data

Each year there are approximately 25,000 participating households. We drop observations that are outliers (below the first percentile or above the 99th). We drop quarters where a household did not report expenditure in more than five of the nine food groups.

In the UK all households record purchases of all random-weight items.

TNS collects information on the characteristics of all individual products, which includes their nutritional content (as shown on the packaging), from a variety of sources including manufacturer databases and from the packages directly.

¹⁵This collection and matching of nutritional characteristics with the food purchases data was done in 2010 by Pierre Dubois and research assistant Yohann Chiu who we thank for his work.

A.4 Definition of Food Groups

The specific food products purchased in each country are very different, in fact even the food categories used by market research firms are different, reflecting differences in the types of food purchased. We therefore classify products into 52 categories used by the USDA, making sure that each product is always classified in the same category whatever the country. In order to facilitate comparison across countries we further aggregate these into nine broad product categories. The category definitions we use are given in Table A1.

Table A1: Definition of food groups

Name	Main items
Fruits	fresh, canned or frozen fruit as well as fruit juices
Vegetables	fresh, canned or frozen vegetables and starchy food
Grain	flour, cereals, dry and fresh pasta, rice, couscous, breakfast cereals, and breads
Dairy	milk, cream, cheese, and yogurt
Meats	beef, pork, lamb, veal, poultry, as well as bacon, ham, sausages, eggs and all fish and seafood, whether fresh, smoked, frozen or canned; nuts
Oils	oils, butter, margarine, and lards
Sweeteners	sugar, syrup, honey and artificial sweeteners
Drinks	sodas, water, coffee, tea and other beverages
Prepared	all commercially prepared items, whether sweet savory, frozen, canned or deli

A.5 Adult Equivalence

We construct a household equivalence scale based on daily caloric requirement of all household members divided by 2500. Daily Caloric Requirement of individual household members is given in Table A2.

Table A2: Caloric needs by age and gender

	Age Categories							
Gender	4-6	7-10	11-14	15-18	19-50	51-59	60-74	75 plus
Female	1545	1740	1845	2110	1940	1900	1900	1810
Male	1715	1970	2220	2755	2550	2380	2330	2100

Source: HMSO (1991).

A.6 Hedonic prices

As we can see in equation (4) the results in the simulations depend on three key inputs: $\hat{\tau}^{V_5}$, $\bar{\sigma}_j^{V_3}$, and $\bar{p}_j^{V_2} - \hat{\tau}^{V_5} \sum_c \hat{\beta}_c^{V_4} \bar{a}_{jc}^{V_1}$. The first two can be found in the respective column for each simulation and in Table 7, by looking up the appropriate column. The third are the hedonic prices displayed below. The layout of the columns follows that of Tables 8 and 9.

Table A3: Hedonic prices for French food in the US

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
scenario:		A	B	C	D	E	
attributes (V_1)	US	FR	US	FR	FR	FR	FR
prices (V_2)	US	US	FR	FR	FR	FR	FR
category effects (V_3)	US	US	US	US	US	FR	FR
nutrient pref (V_4)	US	US	US	US	US	US	FR
price adjustment (V_5)	1	1	1	1	1.079	1.079	1
Fruits	1.92	1.92	1.89	1.93	1.92	1.92	1.90
Vegetables	2.53	2.53	2.41	2.42	2.41	2.41	2.33
Grain	2.89	2.89	3.04	2.96	2.89	2.89	2.57
Dairy	1.11	1.11	1.89	1.63	1.50	1.50	1.13
Meats	2.36	2.36	6.81	6.25	5.93	5.93	5.10
Oils	3.35	3.35	4.07	3.93	3.83	3.83	3.78
Sweeteners	3.18	3.18	1.46	1.57	1.46	1.46	1.54
Drinks	1.44	1.44	0.76	0.81	0.81	0.81	0.81
Prepared	2.45	2.45	3.36	3.08	2.84	2.84	3.43

Notes: The hedonic price is given by $\bar{p}_j^{V_2} - \hat{\tau}^{V_5} \sum_c \hat{\beta}_c^{V_4} \bar{a}_{jc}^{V_1}$. Prices are per kilo and are in US\$ using an exchange rate of £1 = \$1.80 and €1 = \$1.25.

Table A4: Hedonic prices for British food in the US

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
scenario:		A	B	C	D	E	
attributes (V_1)	US	UK	US	UK	UK	UK	UK
prices (V_2)	US	US	UK	UK	UK	UK	UK
category effects (V_3)	US	US	US	US	US	UK	UK
nutrient pref (V_4)	US	US	US	US	US	US	UK
price adjustment (V_5)	1	1	1	1	1.089	1.089	1
Fruits	1.92	1.92	3.01	3.02	3.00	3.00	2.95
Vegetables	2.53	2.53	2.21	2.19	2.18	2.18	2.07
Grain	2.89	2.89	1.79	1.97	1.91	1.91	1.44
Dairy	1.11	1.11	0.84	1.03	0.93	0.93	1.10
Meats	2.36	2.36	3.77	3.23	2.87	2.87	3.54
Oils	3.35	3.35	2.85	2.91	2.81	2.81	1.59
Sweeteners	3.18	3.18	0.95	1.04	0.92	0.92	0.74
Drinks	1.44	1.44	2.39	2.40	2.39	2.39	2.35
Prepared	2.45	2.45	2.75	2.31	2.03	2.03	2.92

Notes: The hedonic price is given by $\bar{p}_j^{V_2} - \hat{\tau}^{V_5} \sum_c \hat{\beta}_c^{V_4} \bar{a}_{jc}^{V_1}$. Prices are per kilo and are in US\$ using an exchange rate of £1 = \$1.80 and €1 = \$1.25.