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

Pierre Dubois   Rachel Griffith   Aviv Nevo

September 13, 2012

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 can explain some difference (e.g., the US-France difference in caloric intake), but generally cannot explain many of the compositional patterns by themselves. Instead, it seems like an interaction between the economic environment and differences in preference and eating habits is needed to explain cross country differences.

*We thank Martin Browning, Arthur Lewbel, Peter Neary and participants in several seminars for comments. We gratefully acknowledge support from USDA, ANR-ESRC Open Research Area 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. Indeed, cross country differences in outcomes are often cited as support for the health benefits of different diets.[1] Economists tend to attribute the difference across markets in food purchases to differences in relative 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 purchases of food for consumption at home between the US, UK and France. We show that US households purchase substantially 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 of expenditure is on fruits and vegetables. We also document substantial differences in relative prices and nutritional characteristics 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. The model we propose generalizes, in non-trivial ways, many of the commonly used demand models. We estimate the model using household-level (home scanner) data that document detailed food purchases for an extended period for participating households across 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 US households would purchase if faced with prices and food characteristics in France and the UK, and the nutritional content of the food baskets they purchase. This allows us to measure how much of

[1] For example, the Mayo Clinic web site promotes the benefits the Mediterranean diet, which incorporates the traditional eating style of countries bordering the Mediterranean Sea (http://www.mayoclinic.com/health/mediterranean-diet/CL00011). Guiliano (2005), a recent popular best seller, suggests similar benefits from French eating habits.

the differences in food purchases are due to differences in prices and nutritional attributes, 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.[2]

We find that, if faced with French relative prices and product attributes, the average US household would purchase substantially fewer calories, in fact 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 of the average US household's food basket to those seen in France, holding quantities fixed, 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. It turns out to 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 are not the whole story. Price differences mostly explain the large difference in caloric intake between the average French and US consumers. 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 the UK consumers have healthier purchasing patterns than US consumers despite the prices and product offering they face, not because of them.

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

There are several reasons to be interested in the differences across the three countries in food purchases. 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 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 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.[3]

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.[4] More generally, nutrition is well understood to be an important determinant of health outcomes, and poor health outcomes lead to high 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

[3] See also NHS (2009) for the UK and Obepi-Roche (2009) for France.

[4] See also, Abaluck (2011), Acs et al (2007a), 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) and Wardle (2007).

be to model demand for food at a much more aggregated level.[5] 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), but also on the quantity consumed of each product. This model nests commonly used models in characteristics space, such as 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). 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 exploit. In each country we have a large sample of many thousands of 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 the period 2005-2006. The raw data consists of millions of observations, which are purchases of specific food items by households on particular shopping occasions. 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 the expenditures on products to the nutritional content.

[5]For example, following this approcah 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.

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, which is the key to our identification, we need to 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.[6] Neither of these papers is able to provide 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.

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 to 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, which allows us to get 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

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

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 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 the same 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.[7]

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, Nieslen asks a random 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

[7] The 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.

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 analysis. To control for difference in size and composition across households we use an adult equivalence scale based on caloric needs.[8] 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). The Data Appendix provides details. The sample average of this measure is also similar across the countries.

Table 1 : Demographics

	France	UK	US
# of households	12,918	14,450	9,003
Household size	2.7	2.7	2.4
# of kids	0.6	0.7	0.5
Adult equivalent	2.2	2.1	2.0

Notes: numbers represent averages across households in the sample used in the subsequent analysis. Adult equivalent is a scale of caloric needs: we sum the daily caloric needs of each member of the household (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 with more aggregated purchase data and coarse data on expenditures and prices provides much less accurate estimates of preference parameters. Also having panel data on households' purchases 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 mostly rely on

[8] An 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.

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. 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 household in recording.

The act of 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 household 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. 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 suggest 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. While 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. Nevertheless, as we interpret the results and the cross country patterns we see in the data we will keep in mind that we do not include consumption outside the home and that calories consumed outside the home are the highest in the US.

2.2 Purchasing Patterns

We start by considering total food purchased.[9] 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, protein 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$.

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 mainly in the form of carbohydrates. French households purchase both a large amount and higher share of their calories in the form of protein and fat. Average spending on food is higher in France than in the US or UK.[10]

---

[9] We 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. We will assume that this effect is constant across the countries we examine.

[10] These broad comparisons also hold if, rather than looking at the mean, we look at the 25th, 50th or 75th percentile, of the distribution across consumers.

Table 2 : Mean Consumption Across Countries

	FR	UK	US
calories	1764.1	1818.7	2058.4
from carbohydrates	642.8 (37%)	839.7 (47%)	997.9 (49%)
from protein	280.3 (16%)	275.6 (16%)	261.7 (13%)
from fats	789.5 (46%)	655.5 (37%)	761.6 (38%)
carbohydrates (g)	171.4	223.9	266.1
protein (g)	70.1	68.9	65.4
fats (g)	87.7	72.8	84.6
expenditure ($)	4.95	4.42	4.52

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.[11] These figures imply 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 details what products are included in each category. In principle we can proceed to a much more disaggregated level, and we comment on a few examples where this helps to provide additional insight to our results.

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

[11] The 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.

and vegetables, both in dollar terms and as a fraction of overall expenditure. The average UK household spends less than US and French households on meats and more on grains, while the average US household spends less on dairy and more on drinks and prepared foods.

The next three columns present quantities, measured in kilograms.[12] The numbers in these columns present a slightly different picture than the expenditure numbers. The US and UK consumption patterns are now quite different. Even in categories where the expenditure shares were relatively similar, for example vegetables, the differences across countries in prices imply different quantities. 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, fruits 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, with the average UK household consuming more fruits, 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 fruits, vegetables, dairy and meat and less from prepared foods. The US and UK households look less similar now. For example, the expenditure share of prepared food is almost identical, and even quantities are not too dissimilar, yet the share of calories from prepared foods is higher in the US. 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.

[12] Here, 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.

Table 3: Expenditure and Quantity by Category

	Expenditure ($ per qtr)			Exp Shares (%)			Quantity (kilo per qtr)			Calorie Share (%)		
Category	FR	UK	US	FR	UK	US	FR	UK	US	FR	UK	US
Fruits	31.39	39.36	34.95	8.5	9.2	8.2	15.5	13.1	17.4	6.8	4.4	5.4
Vegetables	46.07	43.20	34.77	10.0	10.3	8.1	19.0	18.9	14.2	5.5	6.0	3.1
Grains	25.94	32.24	31.07	6.0	8.4	8.0	6.8	12.6	8.8	14.6	19.7	14.7
Dairy	76.78	50.03	39.14	16.6	12.6	9.7	25.9	25.8	20.7	17.4	12.7	9.6
Meats	153.45	75.01	81.17	31.4	18.3	19.3	14.7	10.5	14.8	16.9	13.2	16.5
Fats	15.63	7.70	8.42	3.3	2.0	2.0	3.1	2.0	2.3	12.6	6.8	7.1
Sugars	6.19	4.11	5.64	1.4	1.1	1.4	2.6	2.3	2.6	5.4	5.0	4.5
Drinks	28.41	23.08	41.75	6.1	5.8	10.3	45.5	16.7	50.3	3.3	2.0	6.1
Prepared	98.88	131.28	150.97	21.2	33.0	36.6	16.7	24.9	29.9	23.1	31.4	38.7

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 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 15% and 17%

if paying the UK and French prices. On the other hand, French expenditure would decrease by 13% and 3% if the French paid US or UK prices. The UK expenditure would decrease by 5% if paying US prices but would increase by 14% if paying French prices.

Table 4: Mean Prices by Category

	FR	UK	US
Fruits	2.09	3.20	2.11
Vegetables	2.53	2.32	2.63
Grain	3.94	2.65	3.72
Dairy	3.39	2.26	2.48
Meats	10.34	7.24	5.85
Fats	5.28	3.96	4.47
Sugar	2.79	2.37	4.43
Drinks	0.90	2.46	1.45
Prepared	6.10	5.40	5.12

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

The differences in relative prices are consistent with some of the difference 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 45).We have to be a little careful in this comparison, since it could be driven by composition effects.

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 of 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	65	71	3	4	2	8	7	1
Vegetables	39	30	50	20	10	13	76	21	7
Grain	210	188	227	34	31	38	95	23	36
Dairy	18	27	29	71	41	48	187	119	131
Meats	5	16	30	76	65	66	120	102	205
Fats	2	6	6	11	2	2	679	583	671
Sugar	305	346	345	3	3	0	0	1	0
Drinks	27	24	69	1	3	2	1	5	5
Prepared	126	104	194	24	19	22	127	82	117

Notes: Figures are means across all food products purchased in our sample, with each food product (UPC) having an equal weight. The 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), as opposed to differences in preferences. In principle, there are several ways we could approach the problem. We could model demand at a disaggregate product level, for example demand for soft drinks, and assess the importance of prices and various characteristics. This is a standard approach in the Industrial Organization literature (see, inter alia, Berry, Levinsohn and Pakes (1995) and Nevo (2000)). This approach will pin down preferences within narrowly defined product groups, but will not let us address questions of choice among product groups. Furthermore, since narrowly defined products (i.e., brands) are very different across countries, this approach will have problems matching products across the countries.

We therefore take a different approach. We model demand for food at home more generally. We could model demand at an aggregated level, for example at the level of the nine categories we used in the previous section. However, this would not take advantage of the detailed information in our data. Also, it does not 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. As we saw, there are differences across countries in the attributes of each product. We need a model of demand that allows for preferences to depend on characteristics, nutrients in our case, and products. We also need a demand model that can deal with the underlying richness of the data and the dimensionality problem it causes.

The model we use builds on Gorman (1956) and Lancaster (1966), where utility depends on the characteristics of the product. A special case of the characteristics model is the commonly used discrete choice model (McFadden (1974), Berry, Levinsohn and Pakes (1995)), where utility from each product depends on a small number of characteristics. Another commonly used model that builds on the characteristics model is 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 household[13] chooses from N products, where product n is characterized by C characteristics $\{a_{n1}, ..., a_{nC}\}$. 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 $\mathbf{A} \equiv \{a_{nc}\}_{n=1,..,N,c=1,..,C}$. 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:

$$\max_{x_i, \mathbf{y}_i} \; U(x_i, \mathbf{z}_i, \mathbf{y}_i; \eta_i)$$

$$s.t. \quad \sum\nolimits_{n=1}^{N} y_{in} p_n + p_0 x_i \leq I_i \;\; ; \qquad \mathbf{z}_i = \mathbf{A}' \mathbf{y}_i; \quad x_i, \; y_{in} \geq 0,$$

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

$$\max_{\mathbf{y}} \; U\left(\frac{R - \mathbf{p}'\mathbf{y}}{p_0}, \mathbf{A}'\mathbf{y}, \mathbf{y}\right)$$

$$s.t. \quad y_n \geq 0.$$

[13] As 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.

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

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

We will assume that second order conditions are satisfied, and show that this is the case when we specify the estimated utility function.

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 $\partial U/\partial y_n = 0$, 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.[14] 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.

[14]The equivalence is a bit tricker when there are both common characteristics and product specific ones. Suppose for example that we are in a discrete choice world where $y_n \in \{0, 1\}$ and $\sum_{n=1}^N y_n \leq 1$. Then if we believe that the utility function is $U(x, \mathbf{y})$, the utility maximization can be written as choosing among N options each with utility $U_n(x, f(a_n))$. This imposes an index restriction and is not in general equivalent to maximizing utility in characteristics space.

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, now $\frac{\partial U/\partial y_n}{\partial U/\partial x} = \frac{p_n}{p_0}$, 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. 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. Conversely, if the product has an attribute that has a negative marginal utility, $\partial U/\partial z_c < 0$, then the hedonic price is higher than the price and increases with the amount of the characteristic per unit that the product contains.

This model allows us to take advantage of both flexible models in product space and product characteristics to guide substitution patterns. By specifying a flexible functional form for utility, substitution patterns will be driven by the distance between products in characteristic space. In this sense the approach is closely related to the models proposed by Chan (2006), Pinkse, Slade and Brett (2002) and Davis (2006). The main difference is that our model comes from direct specification of the utility model, instead of a direct specification of flexible demand. This has the advantage of clearly relating demand parameters to utility parameters.

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. We do this in order to keep the model tractable, and to work at a level of products that are readily comparable across countries. We use a fairly aggregated level of nine food groups, but base the food group aggregates on solid foundations, aggregating them 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:

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

where $z_{ic} = \sum_{k,j} a_{kj,c} y_{ikj}$, $f_{ikj}\left(y_{ikj}\right)$ and $h_{ic}\left(z_{ic}\right)$ 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 $h_{ic}\left(z_{ic}\right) = z_{ic}^{\beta_c}$ 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 $f_{ikj}(y_{ikj}) = \lambda_{ikj} y_{ikj}^{\theta_{ij}}$.

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 $\mathbf{y_i} = \left(\mathbf{y_i^1}, .., \mathbf{y_i^J}\right)$ where $\mathbf{y_i^j} = (y_i^{j,1}, ..., y_i^{j,K_j})$, we assume that the utility function satisfies $U(x_i, \mathbf{z_i}, \mathbf{y_i}; \eta_i) = U(x_i, \mathbf{z_i}, W_1\left(\mathbf{y_i^1}\right), .., W_J\left(\mathbf{y_i^J}\right); \eta_i)$, where $W_j\left(.\right)$ 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 $\mathbf{A'y}$. 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 $\left(\mathbf{y}^1, .., \mathbf{y}^J\right)$ in the sets $Y(\mathbf{z_0}) = \left\{\mathbf{y}|\mathbf{A'y} = \mathbf{z}_0\right\}$ 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, $W_j(.)$, are defined over vectors $\widetilde{\mathbf{y}}_j$ of size N, where $\sum_{j=1}^{J}\widetilde{\mathbf{y}}_j = \mathbf{y}_j$. 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 the model in this paper, weak separability is generalized. However, the way weak separability is generalized in Blundell and Robin (2000) is different than our model.

Maximizing utility subject to the budget constraint will yield the following first order conditions[15]

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

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

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

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 further simplified to:

$$\sum\nolimits_k p_{kj}y_{ikj} = p_0\frac{\mu_{ij}\theta_{ij}}{\gamma_i} + \sum\nolimits_c p_0\frac{\beta_c}{\gamma_i}\sum\nolimits_k a_{kjc}y_{ikj}. \tag{1}$$

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 (1). 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,

[15] Without nutrients in the utility function, we know that global concavity of this Cobb-Douglas function with CES aggregates is obtained if $\theta_{ij}\mu_{ij} > 0$ (we assume that all λ_{ikj} are positive). But as nutrients also affect utility, we assume that the second order conditions are still satisfied, which can be checked (for example) easily if we had only one product.

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 $\gamma_i = 1$, which are innocent for the purpose of estimation. The normalization of the price of the outside good p_0 will have to be taken into account when we consider counterfactual experiments where the consumer is moved to another economic environment with potential different prices for other goods. We return to this point below.

Our estimating equation is

$$w_{ijt} = \sum_c \beta_c z_{ijct} + \delta_{ij} + \xi_{jt} + \varepsilon_{ijt} \tag{2}$$

where $w_{ijt} = \sum_k p_{ikjt} y_{ikjt}$, is the expenditure on food group j by consumer i at period t, and $z_{ijct} = \sum_k a_{kj,c} y_{ikjt}$ is the amount of nutrient c consumer i gets from group j at period 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 taste better 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 consumer 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 consumer-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, for all j and for simplicity assume that there is a single nutrient, say carbohydrates.

The estimating equation becomes

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

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}. 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 consumers 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 well-known 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 near where the consumer typically shops, and use this availability as an IV. Instead, we approximate it by computing for each household-category-time the (unweighted) average nutrient content of the set of products in the category purchased by the household in that quarter. The difference between this average and the endogenous variable z_{ijct} is that this average is not weighted by quantity. This average can be thought of 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.

Denote by $\mathcal{A}_{ijt}$ the choice set of products in category j for household i in period t. We use the average nutritional content of the choice set, $\omega_{ijct} = \frac{1}{\#\mathcal{A}_{ijt}} \sum_{k \in \mathcal{A}_{ijt}} a_{kj,c}$, as instrumental variables. Note, that these variables, one for each nutrient, will vary across periods, households and categories because of the variations in the choice sets $\mathcal{A}_{ijt}$. Our identifying assumption is that for $c = 1, ., C$

$$E\left(\varepsilon_{ijt} | \omega_{ijct}, \delta_{ij}, \xi_{jt}\right) = 0.$$

It requires that, conditional on household-category and category-period fixed effects (i) ε_{ijt} is not correlated with which products (UPCs) are bought, 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 ω_{ijct} are highly correlated with $z_{ijct} = \sum_{k \in \mathcal{A}_{ijt}} a_{kj,c} y_{ikjt}$, 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 (2), estimated separately for each country. An observation in the estimation is a household-category-quarter, where we define nine categories as in Section 2 and described in the Data Appendix. The dependent variable in all the regressions is the expenditure in dollars per adult equivalent for a household in a quarter and category. The nutrients we examine are carbohydrates, proteins and fats, all measured in kilograms. One might think that these nutrients would not be important in determining the choices at this level. However, the results seem to suggest they are statistically significant, 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	3.425	3.262	2.047	1.389	1.697	0.643
	(0.0830)	(0.015)	(0.014)	(0.183)	(0.127)	(0.230)
Proteins	39.40	27.25	25.07	21.10	14.16	27.78
	(0.417)	(0.059)	(0.084)	(0.619)	(0.571)	(0.937)
Fats	7.037	10.93	4.245	3.374	3.787	0.737
	(0.125)	(0.049)	(0.053)	(0.180)	(0.273)	(0.297)
Validity of IV				3427.1	4445.6	881.2
Weak IV				1148.8	1491.8	294.5
Observations	714,978	788,658	402,879	714,978	788,658	402,879
R-squared	0.701	0.625	0.606			

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 per household-category-quarter 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 matrix is full column rank, i.e. the model is identified. The weak IV test is the Cragg-Donald statistic.

The first three columns present estimates from fixed effects OLS regressions. All the regressions include household-category and category-time fixed effects. The first 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. All the coefficients are statistically significant and positive. While they are of somewhat similar magnitude, suggesting perhaps that we estimate a meaningful relationship, they do differ across countries in economically significant ways (as we will see below).

Because of the normalization we make, it is more meaningful to compare ratios of coefficients. The ratios suggest that tastes in the US and France are remarkably similar, with nearly identical ratios of proteins and fats to carbohydrates. The UK consumers, on the other hand, have a lower preference for protein and higher preference for fats. The differences across countries in the coefficients, and the ratios of coefficients (which are free of our normalizations) do not exactly track the differences in consumption. For example, in Table 2 we saw that the average US consumer purchased the most carbohydrates and the highest shares of calories from carbohydrates. On the other hand, the average French consumer purchased the largest fraction of calories from fat. Yet the relative coefficient on carbohydrates and fat in the US and France is almost identical. This difference could be suggestive of the importance of prices: even though preferences are similar, the observed outcomes are quite different due to the different economic environments.

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 in the household's choice set each quarter. 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.

As before all the coefficients are highly significant and the differences across countries substantial. Examining the ratio of coefficients confirms that the French have the highest relative preference for fats. The ratio of the fat coefficient to the carbohydrates coefficient is the highest in France and the lowest in the US. On the other hand, the ratio of protein to carbohydrates is almost three times higher in the US compared to France, and nearly five times higher than the UK. These results suggest that the differences in consumption of nutrients are explained by two factors: differences in the tastes for the food categories, which are captured by the fixed effects, and differences in the relative prices. For example, French households spend a significantly higher fraction of their total expenditure on meats and dairy relative to US households (31% and 17% vs.19% and 10%, see Table 3). This is due in part to the lower relative price of meat and dairy in the US, leading to a lower expenditure holding quantity fixed, and in part due to the different preferences for meat and dairy in the two countries.

In Table 7 we report the mean preferences for categories. These are computed by averaging the household-category and category-quarter fixed effects across households and quarters, within the nine categories. For each country these are,

$$\overline{\sigma}_j = \frac{1}{IT} \sum\nolimits_{it} \widehat{\sigma}_{ijt}$$

where

$$\widehat{\sigma}_{ijt} = w_{ijt} - \sum\nolimits_c \widehat{\beta}_c z_{ijct},$$

w_{ijt}, $\widehat{\beta}_c$ and z_{ijct} are the dependent variable, the estimated coefficients and the regressors from the regression defined in equation (2) for each country, and I and T are the total number of households and periods.

There is no reason that these cannot 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. We see that US households have a higher preference for prepared foods and drinks, both high in carbohydrates, while the French have a much higher preference for dairy and meat, higher in protein. This 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	20.75	30.94	27.44	26.50	35.08	30.60
Vegetables	24.77	24.75	26.60	35.37	33.89	27.80
Grains	-16.59	-21.21	-1.65	5.24	5.17	4.79
Dairy	-9.99	-0.34	6.02	31.78	25.88	7.88
Meat	25.11	3.08	8.90	85.53	40.27	10.53
Fats	-3.99	-6.57	1.41	6.12	2.70	7.18
Sugar	-1.95	-3.13	0.27	2.88	0.35	3.95
Drinks	22.78	19.23	34.84	25.72	21.11	38.93
Prepared	34.96	46.49	82.56	67.40	91.65	103.65

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 a household 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 the average US household purchase if faced with (average) French and UK prices and product attributes. We can view this as putting the average US household in France and the UK or as bringing the environment in France or the UK to the average US consumer. We will talk about the counterfactual as taking an American to a different country, but this is simply a stylistic exercise and not meant to literally simulate what an American would consume if in France. In all cases we look at how expenditure and nutrient patterns differ relative to purchases in the US and the purchases of the average household in France and the UK.

Defining the preferences of the average household requires some caution, both in what we attribute to preferences and how we compute the average. In defining preferences we always use the estimated slope coefficients, $\widehat{\beta}_1, ., \widehat{\beta}_C$ from the home country, i.e. the US coefficients in the simulations below. It is less clear how to think of the error term and the various fixed effects. In principle these could be treated as preference parameters, or we could treat them as unobserved country-specific attributes of the products. If they are preferences parameters then they should be imported with the household. However, if they are (unobserved) attributes of the products then we want to use the values of France or the UK and not the US, the household's home country. In reality they are probably a mixture of both. In the tables below we use the values from the household's home country, i.e., we treat them as preference parameters, but we discuss the alternative in the text.

We start by simulating the counterfactual quantities for each country using the preferences of the average household, $\overline{\sigma}_j^H$, prices and attributes in each country as a reference point. Given our utility function the simulated quantities are given by,

$$\widehat{y}_j^H = \max\left(0, \frac{\overline{\sigma}_j^H}{\overline{p}_j^H - \sum_c \widehat{\beta}_c^H \overline{a}_{jc}^H}\right) \qquad H \in \{US, FR, UK\} \tag{3}$$

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

Note that the simulation is for a household with average preferences. As is usually the case in non-linear models, the simulated quantity, $\widehat{y}_j^H$, will not equal the mean of observed purchases, $\bar{y}_j = \frac{1}{IT}\sum_{it} y_{ijt}^H$. Indeed, one can show that the simulated quantities using the average preferences, price and attributes, will tend to lead to lower quantities than the average observed quantities (reported in the tables above). Furthermore, the difference between the actual and simulated quantities could vary across the countries. In order to make the comparison meaningful, and to preserve some of the relationships we saw in the descriptive comparison across countries, we use average preferences, prices and attributes for a subset of households that are located around the mean total calories purchased across all households. We describe in the appendix exactly how we selected the households.

Note that the denominator of the second term in brackets is the hedonic price, which is an important determinant of the simulations. The hedonic prices are interesting because the solution to the consumer's problem is as usual, but using the hedonic prices instead of the actual prices. A higher nominal price could translate to a lower real hedonic price if the product is richer in nutrients.

When we consider preferences from the US and prices from France or the UK, we need to make an adjustment for differences in the price of the outside good between the two countries.

As we discussed above, for the purpose of estimation we normalized the price of the outside good, p_0, to 1. If we do not adjust the estimates then we are implicitly assuming that p_0 is the same across countries, which seems like a strong assumption. Instead, we use information from the Penn World Tables on the "price level of consumption" to proxy the price of the outside good and compute an adjustment for the (V) visited country, $\hat{\tau}^V = p_0^V / p_0^{US}$. The hedonic price in this case is given by,

$$\overline{p}_j^V - \hat{\tau}^V \sum_c \widehat{\beta}_c^{US} \overline{a}_{jc}^V. \tag{4}$$

In Table 8 we show these hedonic prices. We show the hedonic price for: a household with average US preferences facing US prices and attributes in column (1), average French preferences, prices and attributes in column (2) and average UK preferences, prices and attributes in column (4). In column (3) we show the hedonic price for a household with average US preferences, but facing French prices and attributes. The bottom row, labeled $\hat{\tau}$, indicates that the price of the outside good in France are 7.9% higher than in the US. In column (5) we show the same information for a household with the average US preferences facing UK prices and attributes, and $\hat{\tau}$ suggests that the price of the outside good is 8.9% higher in the UK than in the US. These prices will be informative when we consider the simulations below.

Table 8: Hedonic prices

	(1)	(2)	(3)	(4)	(5)
Fruits	1.86	1.78	1.80	2.86	2.89
Vegetables	2.10	1.93	1.82	1.84	1.62
Grain	0.73	0.85	0.67	0.50	0.40
Dairy	0.65	1.42	1.00	1.20	0.58
Meats	1.04	5.71	4.28	4.00	1.30
Fats	3.93	2.20	3.99	1.54	3.36
Sugar	3.66	1.60	2.22	0.80	1.60
Drinks	1.34	0.81	0.83	2.38	2.38
Prepared	3.50	4.14	4.16	3.81	3.44
prices in:	US	FR	FR	UK	UK
attributes in:	US	FR	FR	UK	UK
beta in:	US	FR	US	UK	US
$\hat{\tau}$	1	1	1.079	1	1.089

Notes: The hedonic price is given by equation (4). Prices are per kilo and are in US$ using an exchange rate of £1 = $1.80 and €1 =$1.25.

We consider three counterfactual scenarios.

Scenario A: the average US household purchases the same quantities of each good as at home, but the goods have the average attributes from France or the UK. In this case the simulated quantities are as in (3) and the amount of calories and nutrients are given by

$$\widehat{q}_{jc}^{A,V} = \widehat{y}_j^{US}\overline{a}_{jc}^{V} \qquad V \in \{FR, UK\}. \tag{5}$$

This scenario simulates the effect of the environment holding food choices constant, as such it mimics the ideas behind a Laspeyres price index.

Scenario B: preferences and attributes are those of the average US household but prices are as in France or the UK. In this case quantities are given by,

$$\widehat{y}_j^{B,V} = \max\left(0, \frac{\hat{\tau}^V \overline{\sigma}_j^{US}}{\overline{p}_j^V - \hat{\tau}^V \sum_c \widehat{\beta}_c^{US} \overline{a}_{jc}^{US}}\right) \qquad V \in \{FR, UK\} \tag{6}$$

and the amount of calories and nutrients are given by,

$$\widehat{q}_{jc}^{B,V} = \widehat{y}_j^{B,V}\overline{a}_{jc}^{US} \qquad V \in \{FR, UK\}. \tag{7}$$

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 prices and attributes are as in France or the UK. In this case quantities are given by,

$$\widehat{y}_j^{C,V} = \max\left(0, \frac{\hat{\tau}^V \overline{\sigma}_j^{US}}{\overline{p}_j^V - \hat{\tau}^V \sum_c \widehat{\beta}_c^{US} \overline{a}_{jc}^{V}}\right) \qquad V \in \{FR, UK\} \tag{8}$$

and the amount of calories and nutrients are given by,

$$\widehat{q}_{jc}^{C,V} = \widehat{y}_j^{C,V} \overline{a}_{jc}^V. \qquad V \in \{FR, UK\}. \tag{9}$$

This scenario simulates the total effect of the change in the economic environment, which can be broken up into components by comparing to Scenarios A and B.

The simulated quantities treat the category j as a single product, i.e., $K_j = 1$. In the simulation, prices and 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 simulation. 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 simulation 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 import preferences from one country to another we would need to average the parameters and choices they imply.

4.2.1 An American in Paris

We start by considering the purchasing behavior of a household with preferences of the average US consumer facing French prices and product attributes. Table 9 shows the predicted change in calories and macronutrients purchased by the average US household under scenarios A, B and C described above.

In column (1) we show the simulated purchases by a household with average US preferences, and in column (5) by a household with average French preferences. For reasons we discussed above, these figures are slightly lower than actual average calories purchased in the US or France. Columns (3)-(5) show the simulated purchases for scenarios A-C.

Table 9: An American in Paris

	(1)	(2)	(3)	(4)	(5)
scenario:		A	B	C	
Calories	1969.6	1947.5	1732.6	1796.6	1774.8
Carb (cal)	981.1	824.5	1054.8	901.3	655.6
% cals	50	42	61	50	37
Prot (cal)	236.9	278.2	151.1	193.1	293.1
% cals	12	14	9	11	17
Fat (cal)	751.6	844.8	526.7	702.2	826.1
% cals	38	43	30	39	47
% difference compared to column (1)					
Calories		-1.14	-13.68	-9.63	
Carb (cal)		-18.99	6.98	-8.85	
Prot (cal)		14.82	-56.81	-22.71	
Fat (cal)		11.04	-42.70	-7.03	
% difference compared to column (5)					
Calories		8.87	-2.44	1.21	
Carb (cal)		20.48	37.84	27.26	
Prot (cal)		-5.37	-93.98	-51.80	
Fat (cal)		2.22	-56.84	-17.64	
attributes in:	US	FR	US	FR	FR
prices in:	US	US	FR	FR	FR
sigma in:	US	US	US	US	FR
beta in:	US	US	US	US	FR
$\hat{\tau}$	1	1	1.079	1.079	1

Notes: Figures are per adult equivalent per day, cols (1) and (5) are simulated using equation (3), col (2) using (3) and (5), col (3) using (6) and (7), col (4) using (8) and (9), and information on the subset of households described in the previous section.

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. Moving to column (3), where we use French prices but keep nutrients as in the US, has a very substantial impact, leading the average US household to substantially reduce the calories they purchase. Column (4) considers the average US household facing French prices and product attributes, and we see that the level of calories purchased is very similar to the average French household - just 1.2% higher, and nearly 10% lower than when they faced US prices and product attributes.

In comparing the total calories purchased by the average US and French household it appears that prices explain almost all of the difference.

However, a conclusion that prices explain all the observed difference between US and French consumption, 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 protein and fats than the average French household, even when faced with French prices and attributes. Looking at column (2) we see that changing attributes alone leads the average US household to move in the direction of the average French household, but prices move them more in the other direction. The bottom two sections of the table show the percentage change in the average US household's purchases under each of the scenarios from first the average US household in the US and second the average French household in France.

When comparing the macronutrients purchased by US and French households it appears that preferences play an important role. When faced with the French environment the US consumer actually purchases less fat and protein relative to purchases at home, thus making purchases even less similar than that of the average French consumer. Carb purchases are reduced, but still the US consumer purchases almost 30% more carbs than the French consumer. Thus, at the aggregate level preferences seem to be playing an important role, but looking at food categories both attributes and prices have an important impact.

Consideration of Tables 5 and 8 shows us in part where these results come from. Comparing columns (1) and (3) of Table 8 we can see that the hedonic price of many of the large food categories for a household with US preferences is higher in France 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. But this is not true of all categories, Fruits, Vegetables, Sugar and Drinks are all cheaper, and so the average US household increases the quantity purchased in these categories. But 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 prices, but the difference in product attributes means that this represents a reduction in calories. This is also true for Sugar and Drinks. For Dairy products it goes in the other direction, the higher hedonic price means that the average US consumer purchases less Dairy when faced with French prices, but the difference in product attributes means that this represents an increase in calories. This is also true for Fats. 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 in France, and this leads to a big reduction in Meat purchased, and in addition 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 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 protein from Meats, 27.4% from Prepared foods, 14.5% from Dairy and 13.9% from Grains. When faced with French prices and attributes they obtain only 14% of protein 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 An American in London

We now consider a household with the average US preferences facing UK prices and attributes. Table 10 is laid out as Table 9, in column (1) we show the simulated purchases by a household with average US preferences, and in column (5) by a household with average UK preferences. For reasons we discussed above, these figures are lower than actual average calories purchased in the US or the UK. Columns (3)-(5) show the simulated purchases for scenarios A-C.

These results are quite different to those for the average US household in France. Comparing columns (1) and (2) we see that using US quantities and UK nutrient content leads to a substantial reduction in calories of over 9%. Products in the UK tend to have less carbohydrates and 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 protein.

Comparing columns (1) and (3) we see that the effect on the average US household of facing UK prices is to purchase substantially more calories, and increase purchase of all the nutrients. This is the opposite of the impact of product attributes. Column (4) combines the two effects. When facing UK prices and attributes The average US consumer would purchase over 7% more calories than in the US, and over 20% more than a household with the average UK preferences facing UK prices and attributes.

Table 10: An American in London

	(1)	(2)	(3)	(4)	(5)
scenario:		A	B	C	
Calories	1969.6	1800.7	2322.8	2134.7	1700.8
Carb (cal)	981.1	839.5	1221.2	1058.3	763.9
	0.50	0.47	0.53	0.50	0.45
Prot (cal)	236.9	269.3	276.1	303.6	272.1
	0.12	0.15	0.12	0.14	0.16
Fat (cal)	751.6	692.0	825.4	772.8	664.9
	0.38	0.38	0.36	0.36	0.39
% difference compared to column (1)					
Calories		-9.38	15.20	7.73	
Carb (cal)		-16.87	19.66	7.29	
Prot (cal)		12.00	14.19	21.95	
Fat (cal)		-8.61	8.95	2.75	
% difference compared to column (5)					
Calories		5.55	26.78	20.33	
Carb (cal)		9.01	37.45	27.82	
Prot (cal)		-1.04	1.48	10.39	
Fat (cal)		3.92	19.45	13.97	
attributes in:	US	UK	US	UK	UK
prices in:	US	US	UK	UK	UK
sigma in:	US	US	US	US	UK
beta in:	US	US	US	US	UK
tau	1	1	1.089	1.089	1

Notes: Figures are per adult equivalent per day, cols (1) and (5) are simulated using equation (3), col (2) using (3) and (5), col (3) using (6) and (7), col (4) using (8) and (9), and information on the subset of households described in the previous section.

Again, Tables 5 and 8 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 protein 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.

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 also differ in important ways. 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 (Lancaster, 1966, McFadden, 1974). Our model allows to depart 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.

This allows us to simulate counterfactual quantities purchased by households with preferences from one country but facing the economic environment of another country. We use these to learn about the relative importance of preferences versus the economic environment.

Our main findings are as follows. First, when we consider an average American household facing the economic environment in France or the UK, we find that the economic environment (prices and product attributes) explains the difference in calories purchased with France, but preference explain the difference with the average UK household. Second, even when the total calories purchased do not change, where these calories come from might change quite a lot. This suggests that, in general, its 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 consumer generally improves her purchasing habits, (while we have not shown this in the above tables) the same is true for the average UK consumer. The UK environment, on the other hand, generates worse outcomes for the average US consumer (and the average French consumer as well). Indeed, the reason the UK consumers purchase less calories than US consumers 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 nest 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. Furthermore, the basic methods can be used to explore demand at a lower level of aggregation.

6 Reference

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

Acs, Z., A Cotten and K. Stanton (2007a) "The infrastructure of obesity" in Acs and Lyles (eds) Obesity, Business and Public Policy, Edward Elgar

Acs, Z., L. Henderson, D. Levy, A. Lyles and K. Stanton (2007b) "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 (2008) "Table de composition nutritionnelle des aliments Ciqual 2008"
http://www.afssa.fr/TableCIQUAL/.

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

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, Murray and 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 (Jan., 2000), 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., & Krueger, A. B. (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

Chan T. (2006) "Estimating a Continuous Hedonic Choice Model with an Application to Demand for Soft Drinks" *RAND Journal of Economics*, Vol. 37, No. 2, Summer 2006.

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

Court, Andrew T. (1939) "Hedonic Price Indexes with Automotive Examples". In The Dynamics of Automotive Demand, ed. Charles F. Roos 99-1 17. New York General Motors.

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

Davis, Peter J., (2006) "The Discrete Choice Analytically Flexible (DCAF) Model of Demand for Differentiated Products," CEPR Discussion Papers 5880

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

Drewnowski, A.(2004), "Obesity and the Food Enviorment: Diety 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 Eenergy 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 Popkin, B.M. (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

Elston, J., K. Stanton, D. Levy and Z. Acs (2007) "Tax solutions to the external costs of obesity" in Acs and Lyles (eds) Obesity, Business and Public Policy, Edward Elgar

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, amd R. Jeffery (2001), "Environmental Influences on Eating and Physical Activities," *Annual Review of Public Health*, 22:309-35.

Goldman D., D. Lakdawalla and Zheng (2009) "Food Prices and the Dynamics of Body Weight" NBER Working paper 15096

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.

G*ortmaker, Steven, Boyd Swinburn, David Ley, Rob Carter, Particia Mabry, Diane Finegood, Teryr Huang, Tim Marsh and Marjory Moodie (2011) "Changing the future of obesity: science, policy, and action" The Lancet, 378: 838-847*

Griffith, R., Leibtag, E., Leicester, A. and Nevo A. (2009) "Consumer shopping behavior: how much do consumers save?" *Journal of Economic Perspectives*, Vol 23, No. 2, 99-120

Griffith, R. and O'Connell, M. (2009) "The Use Of Market Research Data For Research Into Nutrition" *Fiscal Studies*, vol. 30, no. 3/4, pp. 339-365

Griffith, R. and O'Connell, M. (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, 1961. Collected in Griliches (1971).

Guiliano M. (2005), French Women Don't Get Fat. Random House.

HMSO (1991) Report of the Panel on Dietery 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. (Jan., 1985), 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 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

NHS (2009) " Statistics on obesity, physical activity and diet: England, February 2009" http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles/obesity/statistics-on-obesity-physical-activity-and-diet:-england-february-2009

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.

Obépi-Roche (2009) "Enquête épidémiologique nationale sur le surpoids et l'obésité", http://www.roche.fr

Philipson, Tomas J. and Posner, Richard A. (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., C. Dai, L. Helmchen, J. Variyam, (2004) "The Economics of Obesity: A Report on The Workshop Held at USDA Economic Research Services" USDA, Washington, DC.

Philipson T., and R. Posner, (2011) “Economics of Obesity: A Decade of Research”, forthcoming, *Journal of Economic Literature*

Pinkse J., M. Slade, C. Brett (2002) "Spatial Price Competition: A Semiparametric Approach", *Econometrica*, Vol. 70, pp. 1111-1153

Rosen, S. (1974) “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” *Journal of Political Economy*, 34-55.

Seale, J., Regmi, A., Bernstein, J. (2003) "The International Evidence on Food Consumption Patterns" USDA Technical Bulletin Number 1904

Swinburn, B., Sacks, G. and Ravussin, E. (2009) "Increased food energy supply is more than sufficient to explain the US epidemic of obesity" *The 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.

Wardle, J. (2007) "Eating behavior and obesity" *Obesity Reviews*, 8, 73-75

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 protein by 4, and grams of fat by 9.

When we do the simulations we use the mean values from households that form a symmetric interval whose width is such that the average of total calories purchased of households in this subset is the closest to the average of the full sample. Using averages from this subset of households has the advantage that it keeps the average simulated amounts closer to the overall average observed in the population, because it is less influenced by outliers.

A.1 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 purchased[16]. 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).

A.2 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 one of the nine food groups.

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

[16]This collection and matching of nutritional characteristics with the Kantar panel was done by Pierre Dubois and research assistant Yohann Chiu in spring-summer 2010.

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.

A.3 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. It there was no match we used the average nutrients in the Gladson data within product module (PM)[17], 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 case 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

[17] Nielsen classifies each product into a very narrowly defined group called product module. There are over 600 of these groups.

we manually fill in the using information from the USDA National Nutrient Database for Standard Reference.

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. In order to facilitate comparison across countries we further aggregate these into nine broad product categories. The category definitions we use are as follow

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
Fats	oils, butter, margarine, and lards
Sugar	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 by:

Table A2: Caloric needs by age and gender

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

Source: HMSO (1991).