Demand for Nutrients: A cross-country comparison

Pierre Dubois, Rachel Griffith, and Aviv Nevo

February 2012

Broad motivation

- Increase in diet-related disease across developed economies
- Obesity
  - linked to hypertension, high cholesterol, coronary heart disease, type 2 diabetes, psychological disorders and various types of cancer
  - Estimates of the cost of obesity are in the (many) billions.
  - Obesity rates differ widely across countries
    - France: 14.5%
    - UK: 23.6%
    - US: 30.0%
- Poor nutrition not just about obesity
  - Excess saturated fat, sugar and salt consumption
  - Low levels of fibres, such as fruit, vegetables and other unprocessed foods

Broad motivation

- What role does the economic environment play?
- Are people making sub-optimal food choices?
  - information/cognition failures
  - externalities
  - other failings in markets
- Role for government intervention?
- What sorts of policies might be effective?

Specific aims

- In this paper, our goals are:
  - provide cross-country descriptive statistics
  - separate the cross-country differences into differences due to the economic environment and other factors
- We exploit a combination of several unique data sets to study cross-country differences
- For France, UK and the US we have:
  - a panel of household purchases gathered using home scanners
  - product level nutritional information

Outline of talk

- Data sets and descriptive statistics
- A model of demand for food and nutritional characteristics
- Estimation using data from France, UK, US
- Simulations

Data

- In each country data collected by households using home scanning devices
  - for market research firms (TNS and AC Nielsen)
  - participating households record all food purchased
  - exact date and location of purchase
  - specific product (barcode) level quantity and price
  - in total hundreds of millions of transactions
- Detailed demographic information
- Nutritional information:
  - information contained on the nutritional label on the back of the package, very detailed

Data comparison across countries

- The products, and even food categories, differ widely across countries
- We classify products into 52 categories
- Further aggregate into 9 broad product categories, which we focus
  - Fruits, Vegetables, Grains, Dairy, Meats, Fats, Sugars, Drinks, Prepared foods
- Focus on macro-nutrients: carbohydrate, protein, fat
- Time period 2005-2006

Data caveats

- Many many advantages of these data
  - very detailed, well measured, comparable
  - observe household over time
  - etc., etc.
- But also some potential concerns:
  - some slight differences across country in: nutritional information, "random weight" items
  - recording error, sample selection
  - consumption outside the home
  - purchase versus consumption
- We try to make as comparable as possible, we provide corroborative evidence where possible

Descriptive statistics: demographics

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
Male age	53.2	56.1	53.6
Female age	51.5	48.8	53.1

Averages across households in the sample

Descriptive statistics: equivalent scale

- Data are household level; household composition varies
- We equivalise the data
  - construct total household caloric needs
  - e.g. girl age 4-6 requires 1545 calories per day
  - boy age 11-14 requires 2220, etc.
  - divide by 2500 to get "adult equivalent"

Descriptive statistics: aggregate purchases

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%)
expenditure ($)	4.95	4.42	4.52

Average per person per day using an adult equivalent scale

Descriptive statistics: expenditure shares by category

Table 3: Expenditure and quantity by Category

	Exp Shares (%)			Quantity (kilo per qtr)		
Category	FR	UK	US	FR	UK	US
Fruits	8.5	9.2	8.2			
Vegetables	10.0	10.3	8.1			
Grains	6.0	8.4	8.0			
Dairy	16.6	12.6	9.7			
Meats	31.4	18.3	19.3			
Fats	3.3	2.0	2.0			
Sugars	1.4	1.1	1.4			
Drinks	6.1	5.8	10.3			
Prepared	21.2	33.0	36.6			

Average per person per quarter using an adult equivalent scale, conditional on strictly positive expenditure

Descriptive statistics: expenditure and quantity by category

Table 3: Expenditure and quantity by Category

	Exp Shares (%)			Quantity (kilo per qtr)		
Category	FR	UK	US	FR	UK	US
Fruits	8.5	9.2	8.2	15.5	13.1	17.4
Vegetables	10.0	10.3	8.1	19.0	18.9	14.2
Grains	6.0	8.4	8.0	6.8	12.6	8.8
Dairy	16.6	12.6	9.7	25.9	25.8	20.7
Meats	31.4	18.3	19.3	14.7	10.5	14.8
Fats	3.3	2.0	2.0	3.13	2.0	2.3
Sugars	1.4	1.1	1.4	2.6	2.3	2.6
Drinks	6.1	5.8	10.3	45.5	16.7	50.4
Prepared	21.2	33.0	36.6	16.7	24.9	29.9

Average per person per quarter using an adult equivalent scale, conditional on strictly positive expenditure

Descriptive statistics: 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

Descriptive statistics: nutrition by category

Table 5: Nutritional Content by Category, in calories

	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

Model overview

- Key challenge: how do we take advantage of the richness of the data?
- Option 1: Estimate demand the "usual" IO way
  - is the disaggregated choice relevant for the big picture?
  - can we generalize?
- Option 2: Use more aggregate product definition
  - How to use nutrient information?
- We follow the second approach and offer a "new" demand system that combines model in product and characteristics space

The model

- The consumer chooses from N products
- Product n is characterized by C characteristics $\{a_{n1}, ..., a_{nC}\}$
- The utility of consumer i with demographics τ_i is $U(x_i, \mathbf{z}_i, \mathbf{y}_i; \tau_i)$
  - x_i is the numeraire; $\mathbf{z}_i$ characteristics, $\mathbf{y}_i$ quantities consumed
- Define the $N \times C$ matrix $\mathbf{A} \equiv \{a_{nc}\}_{n=1,..N,c=1,..,C}$

The consumer's problem:

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

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

The model (cont)

Following standard arguments (and dropping the i subscripts)

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

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

Assuming that $\{y\}_{n=1}^{N}$ are continuous then the FOC if $y_n > 0$ is

$$\sum_{c=1}^{C} a_{nc}\frac{\partial U}{\partial z_c} - \frac{\partial U}{\partial x}\frac{p_n}{p_0} + \partial U/\partial y_n = 0$$

Discussion of the model

- With linear technology, if $U(x, \mathbf{z})$, at most C products consumed
- If $y_n \in \{0, 1\}$ and $\sum_{n=1}^{N} y_n = 1$, then standard discrete choice model.
- If $U(x, \mathbf{y})$ then standard model in product space
- In general the model can:
  - rely on characteristics to guide substitution patterns, without assuming discrete choice
  - "augment" this substitution by relying on flexible functional forms

Discussion of the model

- In our model demand depends on the hedonic prices of each good instead of prices
- If two goods have the same price, but one has more of a characteristic that the consumer values positively, they will adjust downwards the hedonic price - the good is more valuable to them

The model (cont)

- Assume J categories, each with K_j products
- Functional form (for now):

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

where $z_{ic} = \sum_{kj} a_{kjc} y_{ikj}$

- Further assume
  - $h_{ic}\left(z_{ic}\right) = z_{ic}^{\beta_c}$
  - $f_{ikj}(y_{ikj}) = \lambda_{ikj} y_{ikj}^{\theta_{ij}}$ (CES)
  - Alternative: $f_{ikj}(y_{ikj}) = \lambda_{ikj} y_{ikj} - y_{ikj} \ln y_{ikj}$ (Logit)
- Taking FOC and sum over k for a given j

$$\sum_k p_{kj} y_{ikj} = p_0 \frac{\mu_{ij}\theta_{ij}}{\gamma_i} + \sum_c p_0 \frac{\beta_c}{\gamma_i} \sum_k a_{kjc} y_{ikj}$$

Defining products

- In principle, products could be defined very narrowly
- However this creates several problems
  - Is this a good model for the choice between narrowly defined products?
  - Different characteristics can be at play at different levels
  - We need to make the estimates transferable across countries
- Therefore, focus on J "categories" (for now the 9 we showed above) each with K_j mutually exclusive products

Estimating equation

- Assume one characteristic unobserved. Let

$$p_0 \frac{\mu_{ij}\theta_{ij}}{\gamma_i} + p_0 \frac{\beta_1}{\gamma_i} \sum_k a_{kj1} y_{ikjt} = \delta_{ij} + \xi_{jt} + \varepsilon_{ijt}$$

- Normalize, $\gamma_i = 1$ and let $\tilde{\beta}_c = p_0 \beta_c$

$$w_{ijt} = \sum_c \tilde{\beta}_c z_{ijct} + \delta_{ij} + \xi_{jt} + \varepsilon_{ijt}$$

- where
  - $w_{ijt} = \sum_k p_{ikjt} y_{ikjt}$, $z_{ijct} = \sum_k a_{kjc} y_{ikjt}$
  - δ_{ij} HH-cat FE; ξ_{jt} cat-qtr FE

Identification and Instruments

- The error term includes individual preferences for specific categories, category specific seasonal effect, promotional activities and random noise
- The variable, z_{ijct}, is likely to be correlated with these
- We include HH-category fixed effects and category-quarter effects
- To further control for endogeneity we will instrument using the nutrient content of **available** products
  - we assume that observed characteristics are independent of unobserved ones
- Denote by $\mathcal{A}_{ijt}$ the choice set of products in category j for household i in period t.
- $\omega_{ijct} = \frac{1}{\#\mathcal{A}_{ijt}} \sum_{k \in \mathcal{A}_{ijt}} a_{kjc}$,
- Our identifying assumption is that

$$E\left(\varepsilon_{ijt} | \omega_{ijct}\right) = 0.$$

Demand estimates

Table 6: Demand Estimates: preferences for nutrients

	OLS - Fixed Effects			IV - Fixed Effects		
	FR	UK	US	FR	UK	US
Carbs	3.425	3.262	2.047	1.389	1.697	0.643
	(0.0830)	(0.015)	(0.014)	(0.183)	(0.127)	(0.230)
Protein	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)
Obs	714,978	788,658	402,879	714,978	788,658	402,879
R-squared	0.701	0.625	0.606			

Category preferences

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

Average of the household-category and category-quarter fixed effects

Counterfactuals

- We want to simulate what consumers would buy if faced with a different economic environment, e.g that of other countries
- Need to "export" preferences
  - For nutrients it's clear (just take the coefficients)
  - What to do with the product effects?
    - Could reflect preferences, the environment, or a mixture of both
- Can simulate at each data point and average, or compute at the average
- What to do about the price of the outside good (that was normalized in estimation)?
- Choice of products within category (i.e. average characteristics)

Simulation

Start with simulated quantities in each country as reference point

$$\widehat{y}_j^H = \max\left(0, \frac{\overline{\sigma}_j^H}{\overline{p}_j^H - \sum_c \widehat{\beta}_c^H \overline{a}_{jc}^H}\right) \qquad H \in \{US, FR, UK\} \tag{1}$$

where $\overline{\sigma}_j^H$: mean unobserved "preferences"; $\overline{p}_j^H$: mean price; $\overline{a}_{jc}^H$: mean characteristic.

Simulation

- When taking an American to Paris we need to make an adjustment for differences in the price of the outside good between the two countries
- We use information from the Penn World Tables on the "price level of consumption" to proxy the price of the outside good
- Compute an adjustment for the (V)isited country, $\hat{\tau}^V = p_0^V / p_0^{US}$. The hedonic price in this case is given by,

$$\overline{p}_j^V - \hat{\tau}^V \sum_c \widehat{\beta}_c^{US} \overline{a}_{jc}^V. \tag{2}$$

Hedonic prices

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: units are US$ per 1 kilogram

Simulated scenarios

We consider a US consumer in France and the UK; three counterfactual scenarios.
Scenario A: the average US household purchases the same quantities as at home, but the goods have the average attributes from France or the UK. Simulated quantities are as above, calories and nutrients given by

$$\widehat{q}_{jc}^{A,V} = \widehat{y}_j^{US} \overline{a}_{jc}^{V} \qquad V \in \{FR, UK\}. \tag{3}$$

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

Simulated scenarios

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,

$$\widehat{y}_j^{B,V} = \max\left(0, \frac{\hat{\tau}^V \overline{\sigma}_j^{US}}{\overline{p}_j^V - \hat{\tau}^V \sum_c \widehat{\beta}_c^{US} \overline{a}_{jc}^{US}}\right) \qquad V \in \{FR, UK\} \tag{4}$$

and the amount of calories and nutrients are given by,

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

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).

Simulated scenarios

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,

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

and the amount of calories and nutrients are given by,

$$\widehat{q}_{jc}^{C,V} = \widehat{y}_j^{C,V} \overline{a}_{jc}^V. \qquad V \in \{FR, UK\}. \qquad (7)$$

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.

An American in Paris

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
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

An American in London

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
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

Concluding comments

- Documented differences in food purchases across US, UK and France
- Estimated a demand model and used it to simulate behavior across countries

Concluding comments

- Counterfactual simulations to learn about the relative importance of preferences versus the economic environment:
  - an American in France - the economic environment (prices and product attributes) explains the difference in calories purchased
  - an American in the UK- preference explain the difference
  - even when the total calories purchased do not change, where these calories come from might change quite a lot, suggesting it is the interaction of preference, prices and attributes that explains the cross country differences
  - we can rank "healthiness" of preferences and the environment
  - French environment generally encourages healthier purchasing habits
  - UK environment generates worse outcomes, the reason the UK consumers purchase less calories than US consumers is because of their preferences and *despite* their environment, not because of it.

Concluding comments

- Future work :
  - robustness of the results
  - more flexible functional forms
  - individual heterogeneity
  - additional counterfactuals, e.g. within country look at food deserts...