We Are What We Eat

What Do We Eat? Why?

Aviv Nevo

Northwestern University and NBER

FTC Conference

Motivation

- Obesity is linked to disease such as 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: e.g., France: 14.5%, UK: 23.6% ,US: 30.0%
  - states: e.g., Miss, 34%, Alabama, 32.2%, CT 22.5%, CAL, 24%
  - demographic groups
- A key cause of obesity is caloric intake, as well as other poor nutritional habits.

General Questions

- What's going "wrong"?
  - prices?
  - product offering?
  - preferences?
  - behavioral?
- Role for government intervention?
- What sorts of policies might be effective?

General Strategy

- Today I focus (mainly) on cross country differences
- Goals:
  - provide cross-country descriptive statistics
  - separate the cross country differences into differences due to the economic environment and other factors
- For France, UK and the US we have:
  - a panel of household purchases gathered using home scanners
  - product level nutritional information
- Estimate demand and simulate couterfactual consumption

Data

- Data from US, UK and France collected using home scanning devices
  - participating HH record all food purchased
  - exact date and location of purchase
  - UPC level quantity and price
  - In total hundreds of millions of transactions
- Detailed demographic information
- Nutritional information:
  - In UK comes with the data;
  - In France collected "by hand"; only macro nutrients
  - In US comes from Gladson; very detailed

Data Matching

- The products, and even categories, differ widely across countries;
- We therefore classified the products into 52 categories used in the past by the USDA
- We further collapsed these into 9 broad product categories, which we focus on today

Descriptive Statistics

Demographics

	France	UK	US
# of households	12,918	16,108	9,002
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	55.0	53.6
Female age	51.5	47.7	53.1

Averages across households in the sample

Aggregate Purchases

Purchases Across Countries

	FR	UK	US
calories	1770.9	1885.4	2062.4
from carbohydrates	647.6 (37%)	879.4 (48%)	998.5 (49%)
from protein	280.6 (16%)	257.9 (14%)	261.7 (13%)
from fats	790.9 (46%)	698.9 (38%)	765.0 (38%)
carbohydrates (g)	172.7	234.5	266.3
protein (g)	70.2	64.5	65.4
fats (g)	87.9	77.7	85.0
expenditure ($)	4.96	4.08	4.53

Average per person per day using an adult equivalent scale

Aggregate Purchases by city

Purchases Across Cities

city	Calories	Carbs	Protein	Fat	Exp
Chicago	2053.0	271.3	66.1	85.7	4.50
LA	1781.8	232.5	59.2	74.8	4.47
SF	1804.3	229.5	62.2	77.2	4.95
Philly	2084.6	277.8	66.7	86.2	4.74
Mid-west	2241.7	294.2	72.7	94.2	4.54
Southern	2260.4	296.5	68.0	97.4	4.27

Expenditure and quantity

Expenditure and quantity by Category

Category	Expenditure ($ per qtr)			Exp Shares (%)			Quantity (kilo per qtr)		
	FR	UK	US	FR	UK	US	FR	UK	US
Fruits	30.66	32.95	34.82	8.7	8.3	8.2	15.25	11.34	17.24
Vegetables	40.28	34.53	34.77	8.8	9.0	8.1	18.01	16.21	14.18
Grains	26.06	30.17	30.96	5.9	8.4	7.9	6.85	14.67	8.78
Dairy	78.32	41.00	38.66	16.8	11.2	9.6	26.24	21.68	20.53
Meats	152.36	64.67	81.23	30.9	17.0	19.3	14.71	8.87	14.76
Fats	15.63	8.30	9.49	3.3	2.3	2.3	3.13	2.30	2.41
Sugars	6.19	3.91	5.64	1.4	1.1	1.4	2.56	2.13	2.61
Drinks	28.27	25.06	41.86	6.0	6.8	10.3	45.49	16.99	50.45
Prepared	108.60	135.10	151.08	23.1	36.8	36.7	18.39	27.61	30.07

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

Prices

Prices by Category

	FR	UK	US
Fruits	2.08	3.14	2.13
Vegetables	2.31	2.18	2.63
Grain	3.95	2.20	3.73
Dairy	3.41	2.31	2.47
Meats	10.28	7.38	5.85
Fats	5.28	3.84	4.67
Sugar	2.79	2.40	4.42
Drinks	0.89	2.32	1.44
Prepared	6.05	5.03	5.12

Notes: units are US$ per 1 kilogram

Descriptive Statistics

Nutritional Content by Category

	carbohydrates			protein			fat		
	FR	UK	US	FR	UK	US	FR	UK	US
Fruits	13.3	10.9	19.0	1.2	0.6	0.5	1.7	0.5	0.1
Vegetables	8.7	9.4	13.2	2.6	2.7	3.3	0.7	2.5	0.8
Grain	56.2	43.9	60.9	8.6	7.8	9.6	10.6	6.0	4.0
Dairy	4.9	5.0	7.9	17.7	15.3	12.3	20.7	19.0	13.9
Meats	1.7	4.4	8.1	19.2	18.8	16.4	12.8	14.1	22.9
Fats	0.4	0.6	1.6	2.7	0.3	1.2	75.4	81.3	68.0
Sugar	81.2	82.4	92.0	0.9	1.1	0.1	0.1	0.1	0.1
Drinks	5.9	10.2	18.4	0.2	1.4	0.5	0.1	0.7	0.6
Prepared	31.5	30.7	51.4	6.2	6.1	5.6	14.7	11.3	13.0

Average in grams of nutrients (carbohydrates, proteins, fats) per 100 grams of food.

Model Overview

- Key challenge: how do we take advantage of the richness of the data?
- Option 1: Estimate demand the "usual" 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:

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

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

The Model (cont)

Following standard arguments (and dropping the i subscripts)

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

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

Assuming that $\{y\}_{n=1}^{N}$ are continuous then the FOC are

$$\begin{aligned} \sum_{c=1}^{C} a_{nc} \frac{\partial U}{\partial z_c} - \frac{\partial U}{\partial x} p_n + r_n &= 0 \quad \text{if} \quad y_n > 0 \\ \sum_{c=1}^{C} a_{jc} \frac{\partial U}{\partial z_c} - \frac{\partial U}{\partial x} p_n + r_n &< 0 \quad \text{if} \quad y_n = 0 \end{aligned}$$

where $r_n = \partial U / \partial y_n$

Discussion of the Model

- With linear technology, if $U(x, \mathbf{z})$, at most C products consumed;
- Define $r_n = \xi_n$ or $\xi_{nt} + \varepsilon_{int}$ (with consumer and time variation)
- If $y_n \in \{0, 1\}$ and $\sum_{n=1}^{N} y_n = 1$, then standard discrete choice model.
- If discrete choice but continuous options then hedonics
- If $U(x, \mathbf{y})$ then standard model in product space;
- In general the model can:
  - rely on characteristics to guide substitution patterns, w/o assuming discrete choice;
  - "augment" the substitution by relying on flexible functional forms;

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 $z_{ic} = \sum_{k,j} a_{kj,c} y_{ikj}$, $f_{ikj}\left(y_{ikj}\right)$ and $h_{ic}\left(z_{ic}\right)$

- 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 summing k for a given j

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

Estimating Equation

- Assume one characteristic unobserved. Let

$$\frac{\mu_{ij}\theta_{ij}}{\gamma_i} + \frac{\beta_1}{\gamma_i} \sum_k a_{kj,1} \times y_{ikjt} = \delta_{ij} + \xi_{jt} + \varepsilon_{ijt}$$

- Normalize, $\gamma_i = 1$

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

- where
  - $w_{ijt} = \sum_k p_{ikjt} y_{ikjt}$, $z_{ijct} = \sum_k a_{kj,c} y_{ikjt}$
  - δ_{ij} HH-cat FE; ξ_{jt} cat-qtr FE
- We avoid having to define quantities

Identification and Instruments

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

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

Demand Estimates

Demand Estimates: preferences for nutrients

	OLS - Fixed Effects			IV - Fixed Effects		
	FR	UK	US	FR	UK	US
Carbs	3.642	2.775	2.056	0.976	1.739	0.593
	(0.080)	(0.014)	(0.0138)	(0.202)	(0.112)	(0.234)
Protein	39.69	27.93	25.10	22.01	20.91	27.37
	(0.397)	(0.056)	(0.0840)	(0.681)	(0.467)	(0.942)
Fats	7.169	10.01	4.225	3.290	3.443	0.761
	(0.108)	(0.037)	(0.052)	(0.181)	(0.152)	(0.294)
Observations	714,359	888,739	405,296	714,359	888,739	405,296
R-squared	0.705	0.636	0.473			

Category Preferences

Demand Estimates: preferences for categories

	OLS - Fixed Effects			IV - Fixed Effects		
	FR	UK	US	FR	UK	US
Fruits	19.82	27.11	27.32	26.42	29.14	30.63
Vegetables	23.54	19.54	26.59	32.31	24.33	27.94
Grains	-17.61	-26.51	-1.72	6.56	-7.92	5.38
Dairy	-9.83	-0.79	6.09	32.47	13.92	8.29
Meat	22.80	6.83	8.90	81.94	25.86	11.56
Fats	-4.34	-8.51	2.11	6.28	2.40	7.79
Sugar	-2.46	-2.01	0.25	3.85	0.14	4.08
Drinks	20.63	20.22	34.90	25.31	21.90	39.19
Prepared	36.99	42.41	83.02	75.16	79.71	105.22

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

Counterfactuals

- Our goal is to simulate what consumers would buy in 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.

Simulation

$$\widehat{y}_{ijt}^{H,V} = \max\left(0, \frac{\tau\overline{\sigma}_j^V}{\overline{p}_{ijt}^H - \tau\sum_c \widehat{\beta}_c^V \overline{a}_{ijct}^H}\right) \quad H, V \in \{FR, UK, US\}.$$

where τ is a scaling factor that accounts for the differences in p_0 across countries

An American in Paris (and London)

	US	FR	UK	FR	UK
calories	1806	1422	1690	1715	1862
carbs (g)	241	185	217	225	239
protein (g)	52	37	50	46	57
fat (g)	77	64	75	77	82
scale factor		1	1	1.1	1.05

per day adult equivalent. US parameters (i.e., "preferences")

An American in Paris (and London)

	US	FR	UK	FR	UK
Fruits	16.2	16.8	10.6	18.8	11.2
Vegetables	13.1	15.5	18.2	17.5	19.5
Grains	6.6	5.2	5.6	7.9	6.8
Dairy	11.7	6.6	11.8	8.7	13.9
Meat	9.9	2.4	5.2	3.0	6.2
Fats	2.0	1.9	2.4	2.2	2.6
Sugar	1.1	1.8	2.5	2.1	2.6
Drinks	29.0	48.5	18.7	53.9	19.7
Prepared	29.5	24.5	32.4	28.1	35.0
scale factor		1	1	1.1	1.05

Concluding Comments

- Documented differences in food purchases across US, UK and France
- Estimated a demand model and used it to simulate behavior across countries
- The (preliminary) results suggest that economic factors play a key role in the observed cross-country differences
- Future:
  - robustness, generalizations, extensions, etc.
  - US cross region differences; exploit migration data
  - differences across demographic groups