The socioeconomic gradient in diet quality

Rachel Griffith, Martin O'Connell and Kate Smith

Institute for Fiscal Studies

April 2013

Motivation

Well established relationship between health outcomes and socioeconomic status

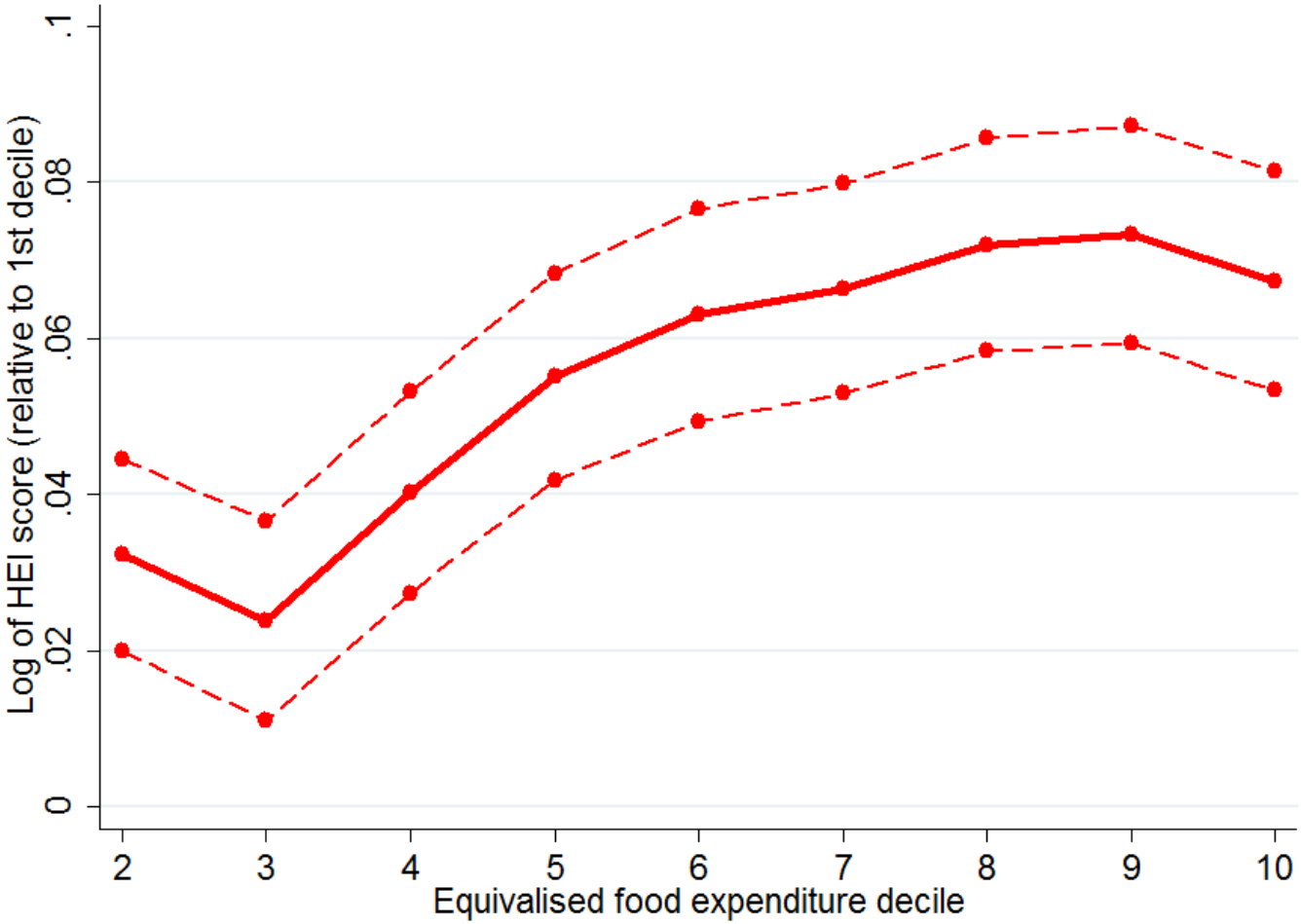
Lower income, less well educated groups tend to have poorer health outcomes

Many of these health outcomes related to diet; not limited to obesity, also concern about excess salt, sugar, saturated fat etc, we focus on these

Correlation between income and diet quality could be driven by the constraints households face (income, prices, time), different preferences, information

MOTIVATION

FIGURE: *Difference in mean HEI across expenditure deciles*



Contribution

Estimate a model of demand for food:

- Panel data allows us to capture unobserved heterogeneity across households
- Detailed data on household specific prices and the nutritional composition of foods

Use this model to estimate the relative impact of different factors, including prices and income, on diet quality

Demand for food

Assume preferences:

- are defined over food groups
- take form leading to Quadratic Almost Ideal Demand System (QUAIDS)

Augment basic framework with household specific preferences, allowing for unobserved heterogeneity

and household specific prices to reflect variation in the prices faced by different households

Demand equations

Model decision of household h in period t over how to allocate total monthly food expenditure, y_{ht}, over food groups indexed $j \in \{1, ..., J\}$

w_{hjt} denotes the share of its period t food expenditure when faced with prices $p_{ht} = (p_{h1t}, ..., p_{hJt})$

$$w_{hjt} = \alpha_{hjt} + \sum_k \gamma_{jk} \ln p_{hkt} + \beta_j \ln \left(\frac{y_{ht}}{\Gamma(p_{ht})} \right) + \frac{\lambda_j}{\Pi(p_{ht})} \left[\ln \left(\frac{y_{ht}}{\Gamma(p_{ht})} \right) \right]^2 + \epsilon_{hjt}$$

where

$$\ln \Gamma(p_{ht}) = \alpha_0 + \sum_j \alpha_{hjt} \ln p_{hjt} + \frac{1}{2} \sum_j \sum_k \gamma_{jk} \ln p_{hjt} \ln p_{hkt}$$

$$\ln \Pi(p_{ht}) = \sum_j \beta_j \ln p_{hjt}$$

Identification

Use period 2006-2009: period of fluctuations in real total food expenditure and variation in relative prices of different foods

Endogeneity of **food expenditure**: instrument for y_{ht} with total non-food grocery expenditure

Endogeneity of **prices**: instrument for p_{ht} using a price constructed with household's long run average purchase weights

Data

Data on all purchases of fast-moving consumer goods that are brought into the home by a representative sample of UK households

- Household records all purchases using handheld scanner
- Including expenditure and transaction level prices on disaggregate products (at barcode level)

Panel of 10,841 households; average length of time in the panel is 41 (of 48) months; 430,238 observations

Data include details of nutritional content of each individual food product

Food types

Table: *Mean expenditure shares, by food type*

Food type and main items	Expenditure share
Fruit: fruit, including fruit juices	8.8%
Vegetables: fresh, canned or frozen vegetables	11.0%
Grains: flour, cerals, pasta, rice, breads	8.7%
Dairy: milk, cream, yogurt	8.8%
Cheese: cheese, oils, butter, margarine	5.8%
Red meat: beef, lamb, pork, nuts, eggs	11.2%
Poultry and fish: poultry, seafood	7.5%
Drinks: fizzy drinks, tea, coffee, water	5.2%
Prepared (sweet): ice cream, cakes, cookies etc.	11.1%
Prepared (savoury): ready meals, soups, snacks	22.0%

Expenditure coefficient estimates

Across all households

VARIABLES	(1) Fruit	(2) Vegetables	(3) Grains	(4) Dairy	(5) Cheese
$\ln(y_{ht}/\Gamma(p_{ht}))$	0.02212***	-0.00788***	-0.01615***	0.04436***	-0.01620***
	(0.00277)	(0.00163)	(0.00288)	(0.00461)	(0.00308)
$\frac{1}{\Pi(p_{ht})}\ln(y_{ht}/\Gamma(p_{ht}))^2$	-0.00321***	0.00018	0.00100***	-0.00607***	0.00141***
	(0.00032)	(0.00019)	(0.00033)	(0.00052)	(0.00035)
	(6) Meat	(7) Poultry	(8) Drinks	(9) PrepSweet	(10) PrepSav
$\ln(y_{ht}/\Gamma(p_{ht}))$	-0.04993***	0.00705***	0.06087***	-0.01505***	-0.02920***
	(0.00683)	(0.00183)	(0.00776)	(0.00264)	(0.00285)
$\frac{1}{\Pi(p_{ht})}\ln(y_{ht}/\Gamma(p_{ht}))^2$	0.00591***	-0.00053**	-.00521***	0.00322***	0.00329***
	(0.00078)	(0.00021)	(0.00088)	(0.00030)	(0.00033)
HH fixed effects	Yes	Yes	Yes	Yes	Yes
Time effects	Yes	Yes	Yes	Yes	Yes
Observations	430,238	430,238	430,238	430,238	430,238
No of households	10,841	10,841	10,841	10,841	10,841

Contrast with "basic" approach

Basic model:

- Use cross-sectional variation in expenditures to identify shape of Engel curves
- Replaces household specific term in α_{hjt} with a vector of observable household characteristics
- Less precise measures of prices - common across households, rather than household specific

Expenditure elasticities

for all households

	Full model	Basic model
Fruit	0.92	0.87
Vegetables	0.94	1.10
Grains	0.92	0.66
Dairy	0.87	0.67
Cheese	0.94	0.96
Red meat	1.04	1.26
Poultry and fish	1.03	1.30
Drinks	1.25	1.36
Prepared (sweet)	1.13	0.79
Prepared (savoury)	1.00	1.05

▸ Price elasticities

Expenditure elasticities

for all households

	Full model	Basic model
Fruit	0.92	0.87
Vegetables	0.94	1.10
Grains	0.92	0.66
Dairy	0.87	0.67
Cheese	0.94	0.96
Red meat	1.04	1.26
Poultry and fish	1.03	1.30
Drinks	1.25	1.36
Prepared (sweet)	1.13	0.79
Prepared (savoury)	1.00	1.05

▸ Price elasticities

Expenditure elasticities

for all households

	Full model	Basic model
Fruit	0.92	0.87
Vegetables	0.94	1.10
Grains	0.92	0.66
Dairy	0.87	0.67
Cheese	0.94	0.96
Red meat	1.04	1.26
Poultry and fish	1.03	1.30
Drinks	1.25	1.36
Prepared (sweet)	1.13	0.79
Prepared (savoury)	1.00	1.05

▸ Price elasticities

VARIATION IN DIET QUALITY

Translate predictions about food purchasing behaviour into implied diet quality

Use the 'Healthy Eating Index' (HEI) measure developed by the USDA

▸ HEI components

Use model to assess the relative contributions of differences across household in: expenditure, prices, preferences

Find that prices and income are not the main determinants of the variation in diet quality across the expenditure distribution

So what is driving it?

The intercept

The intercept of the share demand equation is given by:

$$\alpha_{hjt} = \sum_k \alpha_{jk} z_{htk} + \eta_{hj}$$

Observed time varying hh characteristics: z_{htk}

Write η_{hj} as

$$\eta_{hj} = \sum_l \alpha_{jl} x_{hl} + v_{hj}$$

Observed time invariant hh characteristics: x_{hl}

Unobserved time invariant hh characteristics: η_{ht}

Can consistently recover the α_{jl}s if we assume $\mathbb{E}(v_{hj}|x_h) = 0$

What explains the gradient?

Next steps

Plan is to better understand what the α_{hjt} term represents

Preliminary analysis looks at the HEI we observe in the data

Investigate how relationship with expenditure changes when we account for demographics and household attitudes:

$$\ln(HEI)_h = \beta_0 + \sum_l \beta_l x_{hl} + \epsilon_h$$

Aiming to capture differences in cognitive ability, access to information, difference in tastes, time constraints

Two examples: **education** and **attitude towards health**

What explains the gradient?

Two example factors

1. Education: proxied by a measure of social class.

Households grouped into 6 categories (A, B, C1, C2, D, E), based on occupation of the head of household.

Class A contains the most well-educated households, and class E the least.

What explains the gradient?

Two example factors

2. Attitude towards health: households asked 16 questions which we use to proxy their attitude towards health. e.g

- "I try to lead a healthy lifestyle"
- "My diet is very important to me"
- "I restrict how much sugary food I eat"

Based on answer ('Strongly agree', 'Agree', 'Neither agree nor disagree', 'Disagree' or 'Strongly disagree') given a score of 1–5.

Group households into quintiles based on total score across all questions.

▸ Distribution of scores

What explains the gradient?

VARIABLES	(1) $\ln(HEI)$
Log of total food exp.	**0.055***** (0.004)
Class B	
Class C1	
Class C2	
Class D	
Class E	
Second quintile of health concern	
Third quintile of health concern	
Fourth quintile of health concern	
Top quintile of health concern	
Observations	13,430

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

What explains the gradient?

VARIABLES	(1) ln(*HEI*)	(2) ln(*HEI*)	
Log of total food exp.	**0.055*** **	**0.052*** **	**-5.5%**
	(0.004)	(0.004)	
Class B		-0.006	
		(0.020)	
Class C1		-0.028	
		(0.020)	
Class C2		-0.054***	
		(0.020)	
Class D		-0.068***	
		(0.020)	
Class E		-0.063***	
		(0.020)	
Second quintile of health concern			
Third quintile of health concern			
Fourth quintile of health concern			
Top quintile of health concern			
Observations	13,430	13,430	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

What explains the gradient?

VARIABLES	(1) ln(*HEI*)	(2) ln(*HEI*)		(3) ln(*HEI*)	
Log of total food exp.	**0.055*** **	**0.052*** **	**-5.5%**	**0.037*** **	**-32.7%**
	(0.004)	(0.004)		(0.004)	
Class B		-0.006		-0.005	
		(0.020)		(0.020)	
Class C1		-0.028		-0.023	
		(0.020)		(0.020)	
Class C2		-0.054***		-0.048**	
		(0.020)		(0.020)	
Class D		-0.068***		-0.058***	
		(0.020)		(0.020)	
Class E		-0.063***		-0.058***	
		(0.020)		(0.020)	
Second quintile of health concern				0.053***	
				(0.005)	
Third quintile of health concern				0.096***	
				(0.005)	
Fourth quintile of health concern				0.114***	
				(0.005)	
Top quintile of health concern				0.133***	
				(0.005)	
Observations	13,430	13,430		13,430	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Summary

Correlation between quality of diet and total food expenditure could be driven by a number of factors

Estimate a model of food demand to separate out these effects

Find evidence that differences in prices and income are not the main determinants of the socioeconomic gradient in diet quality

Have begun to look at what may be driving this gradient; the next steps are to incorporate this more fully into the model

Healthy Eating Index (HEI)

Component	Score	Low limit (per 1000 kcals unless stated)	High limit (per 1000 kcals unless stated)
Total fruit	5	0	120g
Whole fruit	5	0	60g
Total vegetables	5	0	165g
Dark green/orange veg	5	0	60g
Total grains	5	0	75g
Whole grains	5	0	32.5g
Milk	10	0	260g
Meat	10	0	70g
Oils	10	0	12g
Sodium	10	>2g	<0.7g
Saturated fat	10	>15% of energy	<7% of energy
Calories from SoFAAS	20	>50% of energy	<20% of energy
Total	100		

▸ Back: SE group and diet quality

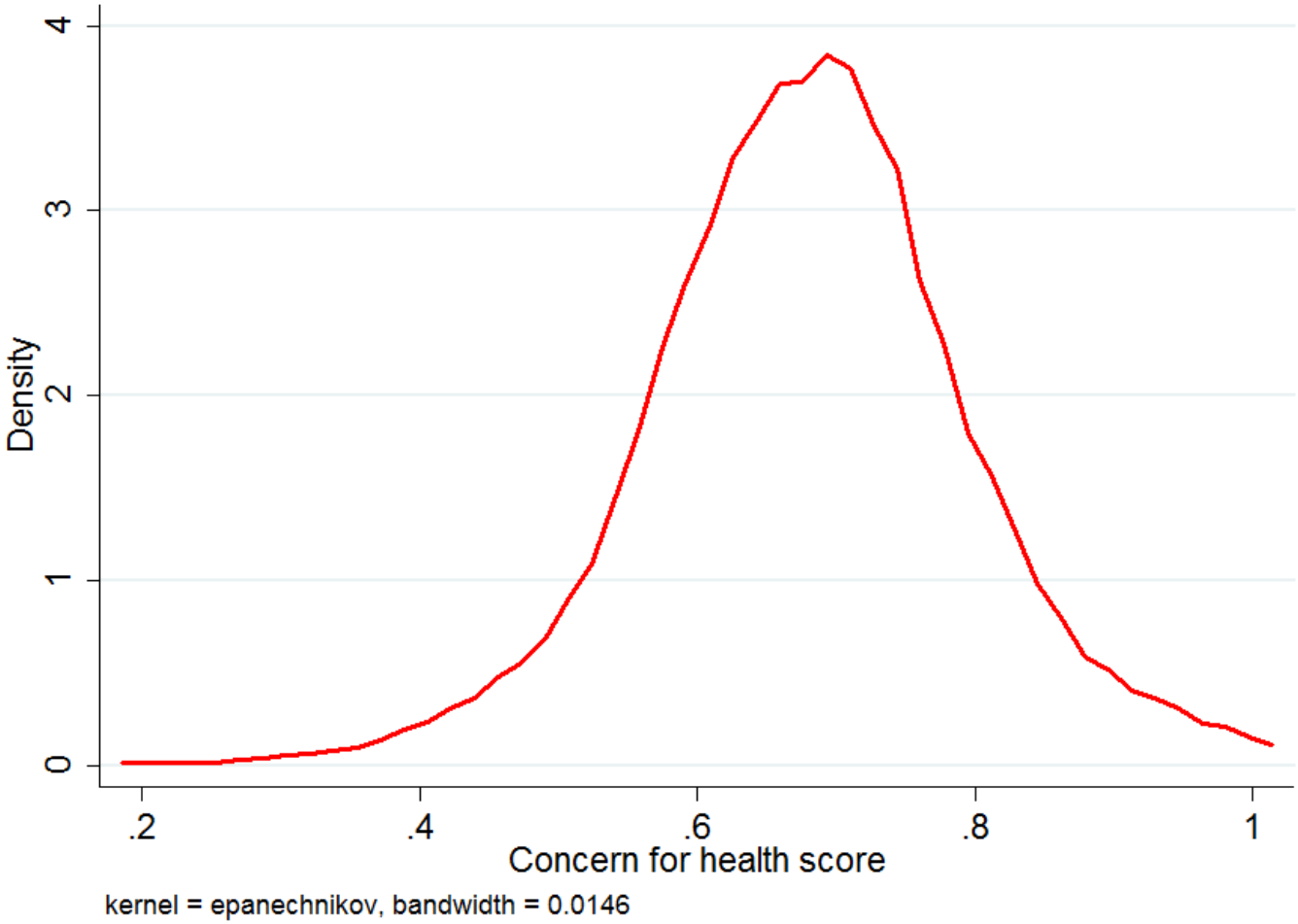
Price elasticities

	Fruit	Vegetables	Grains	Dairy	Cheese	Meat	Poultry	Drinks	Sweet	Savoury
Fruit	**-0.669**	-0.007	-0.039	-0.036	-0.053	-0.043	-0.043	-0.082	-0.021	-0.020
Vegetables	-0.009	**-0.867**	-0.022	-0.018	-0.049	-0.063	-0.026	-0.018	0.007	0.007
Grains	-0.040	-0.018	**-0.711**	-0.024	-0.065	-0.057	-0.029	0.005	-0.008	-0.022
Dairy	-0.041	-0.018	-0.027	**-0.833**	0.002	0.006	-0.022	-0.092	0.003	-0.008
Cheese	-0.031	-0.024	-0.039	0.008	**-0.618**	-0.057	-0.025	-0.015	-0.021	-0.014
Meat	-0.041	-0.055	-0.061	0.023	-0.096	**-0.746**	-0.029	0.047	-0.005	-0.042
Poultry	-0.024	-0.010	-0.014	-0.002	-0.023	-0.022	**-0.809**	-0.031	-0.007	-0.015
Drinks	-0.021	0.010	0.026	-0.019	0.015	0.026	-0.007	**-1.066**	-0.001	0.002
Sweet	0.000	0.025	0.013	0.034	-0.011	0.002	0.007	0.010	**-1.099**	0.015
Savoury	-0.041	0.019	-0.048	-0.008	-0.046	-0.085	-0.047	-0.014	0.020	**-0.907**

Notes: Numbers reported are expenditure weighted elasticities across all households. Element (i, j) gives the change in share of food type j with respect to the price of food type i.

▸ Back: Expenditure elasticities

HEALTH SCORE DISTRIBUTION



Back