HEDG Health, Econometrics and Data Group
THE UNIVERSITY of York

HEDG Working Paper 07/13

Expenditure dispersion and dietary quality: evidence from Canada

Timothy Beatty

May 2007

york.ac.uk/res/herc/hedgwp

Expenditure Dispersion and Dietary Quality: Evidence from Canada

Timothy K.M. Beatty[1]
Reader
Department of Economics
University of York

Abstract

This paper examines the relationship between the way in which a household spreads their food expenditure over time and the dietary quality of the food they purchase. I find that households who make more frequent, smaller food purchases buy healthier foods than households who make fewer, larger purchases. These households are more likely to purchase foods with a lower share of total calories from fats, saturated fats and a larger share of calories from fruits and vegetables. The analysis is extended using quantile regression. The effect of expenditure dispersion is found to be largest amongst households with poor diets i.e. those households with diets high in saturated fats and low in fruits and vegetables.

Keywords: Dietary Quality, Health, Expenditure Dispersion, Quantile Regression.

JEL Codes: D12, I19

CONTACT INFORMATION:

Timothy K.M. Beatty
Department of Economics
University of York
Heslington, York, U.K.
YO10 5DD
tb526@york.ac.uk
www.timothybeatty.name

[1] Thanks to participants at the 2006 AAEA annual meetings for helpful discussions. Funding for this research was provided the Canada Research Chair and the SSHRC standard research grant program and is gratefully acknowledged.

Expenditure Dispersion and Dietary Quality: Evidence from Canada

Introduction

Diet is a factor in many chronic and acute diseases, notably obesity, diabetes, cardiovascular disease, hypertension and various cancers (W.H.O. (2003)). Direct costs of medical care and treatment for diseases strongly linked to diet have been estimated in excess of $200 billion per year in the United States (Klurfeld and Krestch (2001)). This paper is the first to investigate whether there exists a link between the way in which a household spreads expenditure on food over time and the nutritional quality of the food they purchase. Specifically, I consider whether households who purchase most of their food on a small number of purchase occasions buy food that differs in the share of calories from fats, saturated fats, protein, carbohydrates and fruits and vegetables, from households who spread their food-at-home expenditure more evenly over a larger number of purchase occasions. Results suggest that expenditure dispersion and dietary quality are closely linked. This may provide a possible lever for policy makers.

Given the considerable costs associated with low quality diets, it is not surprising that a considerable amount of research has gone into identifying socio-demographic characteristics, such as gender, education and employment status, which correlate with dietary quality (Horton and Cambell (1991, Adelaja, Nayga and Lauderbach (1997, Nayga (1997), Irala-Estevez, et al. (2000, Nayga (2000, Raper, Wanzala and Nayga (2002), Variyan, Blaylock and Smallwood (2002), Ricciuto, Tarasuk and Yatchew (2006)). In short, this literature has found that income, schooling, health knowledge, being white and being female are positively correlated with a better quality diet. Several authors have considered the relationship between food insecurity, poverty and measures of nutrient availability, for example Rose (1999) and more recently Bhattacharya, Currie and Haider (2004). Bhattacharya, Currie and Haider (2004) find that food insecurity and poverty are predictive of several measures of dietary quality for adults and the elderly, but not for children. Previous work on dietary quality has been extremely important in identifying groups that might benefit from policy intervention. However, it has been less informative in identifying observable behaviors that correlate with low quality diets, i.e. behaviors that may be amenable to policy intervention. In contrast, this paper focuses on the role of a single observable behavior, the way in which a household spends its food budget over time, as a predictor of several indicators dietary.

In absence of specific research, nutritionists have recommended that low-income households should concentrate their purchases to take advantage of quantity

discounts. For example, the Dieticians of Canada, suggest purchasing foods in bulk and discourage extra shopping trips, (Lynch (1997)). Indeed, the Center for Nutrition Policy and Promotion at the U.S.D.A., Hogbin, Davis and Escobar (1999), suggest purchasing foods for several meals at once, in an effort to avoid "impulse" buys that may be less healthy. Hersey, et al. (2001) consider a range of shopping behaviors, such as making a list and planning meals ahead, and find they are positively related to nutrient availability in a household. In related work, several authors from the Economics Research Service (E.R.S.) of the U.S.D.A. have looked at how the structural characteristics of food markets affect food choice, dietary quality and health. Kaufman, et al. (1997) finds that low-income households tend to purchase in bulk in order to lower total food expenditures. A priori, the health consequences of this type of expenditure behavior are not obvious. Feather (2003) suggests that lack of access to larger stores may limit access to nutritious food for low-income households.

This paper makes several contributions. First, I propose a novel and easy to interpret means of measuring the way in which a household spends its food budget over time, based on the entropy principle. I then explore the relationship of this measure of expenditure dispersion to household income and total food expenditure by means of a simple semiparametric model. I then investigate the link between expenditure dispersion and measures of dietary quality. Finally, I take a quantile regression approach to completely characterize the role of expenditure dispersion on several measures of dietary quality. I find that households that spread their expenditure more evenly over time are more likely to have a better quality diet and the effect of expenditure dispersion is found to be largest amongst households with the worst diets.

Empirical Approach

This paper asks whether the way in which a household spends its food budget over time is predictive the nutritional quality of the food purchased. Households trade off between present costs associated with producing a healthy diet, the utility of current food consumption and future health benefits. Factors that impact the utility associated with health, such as income, or factors that affect the health production process itself, such as education, (a more educated household may generate health at lower cost) enter the decision process as demand shifters. Expenditure dispersion is treated as another exogenous input into the health production function.

I now lay out an empirical strategy for identifying the relationship between the way in which a household spends its food budget over time and the nutritional quality of the resulting food-at-home expenditure. The analysis proceeds in three

steps: First I propose a measure of expenditure dispersion. Second I relate this measure to food-at-home expenditure and household income. Finally I consider the correlation between expenditure dispersion and the dietary quality of household food expenditure at the conditional mean and at several conditional quantiles.

Analyzing the predictive power of expenditure dispersion in explaining dietary quality requires a data set that covers both a large enough set of food items for measures of dietary quality to be computable and a long enough survey period for role of purchase frequency to be observable. The data for our analysis is drawn from the 1996 Family Food Expenditure Survey (FOODEX), (Statistics Canada (1999)). Appendix A contains a detailed description of the data used in this analysis.

Expenditure Dispersion

How should they way in which a household spends its food budget over time be measured? Intuitively, one would like a metric that captures the degree to which a household concentrates their expenditure on a single purchase occasion versus spreading their expenditure evenly over the entire sample period. The metric would take its maximum when expenditures are evenly spread across the sample period and take its minimum when purchases occur only on a single day.

A well-known metric that has these properties is the entropy function (Shannon (1948)). Theil (1967) suggested using the entropy function as an inequality index. A measure of expenditure dispersion constructed in this way can thus be interpreted as the inequality of daily food expenditure shares over the (in this case two-week) sample period. For a given household, I write the entropy of expenditure over the sample period as

$$H(\mathbf{x}) = -\sum_{d=1}^{D} (x_d / X) \ln(x_d / X), \tag{1}$$

where x_d is food-at-home expenditure on day d , D denotes the length of the survey period and X is total food expenditure over the entire sample period. I adopt the convention that $(0)\ln(0) \equiv 0$. As a result, the entropy function takes on a minimum at zero, when all purchases are made in a single day and is has a maximum value of Log(D), when households spread their expenditure evenly over the two-week sample period. To ease interpretation, I normalize the entropy metric H($\mathbf{x}$) by dividing by Log(D), such that the inequality of food expenditure is bounded between zero and one. Figure 1 presents the empirical density function of the normalized entropy of expenditure for the households in our sample. The normalized measure of expenditure dispersion has a mean of 0.43 and a standard deviation of 0.15.

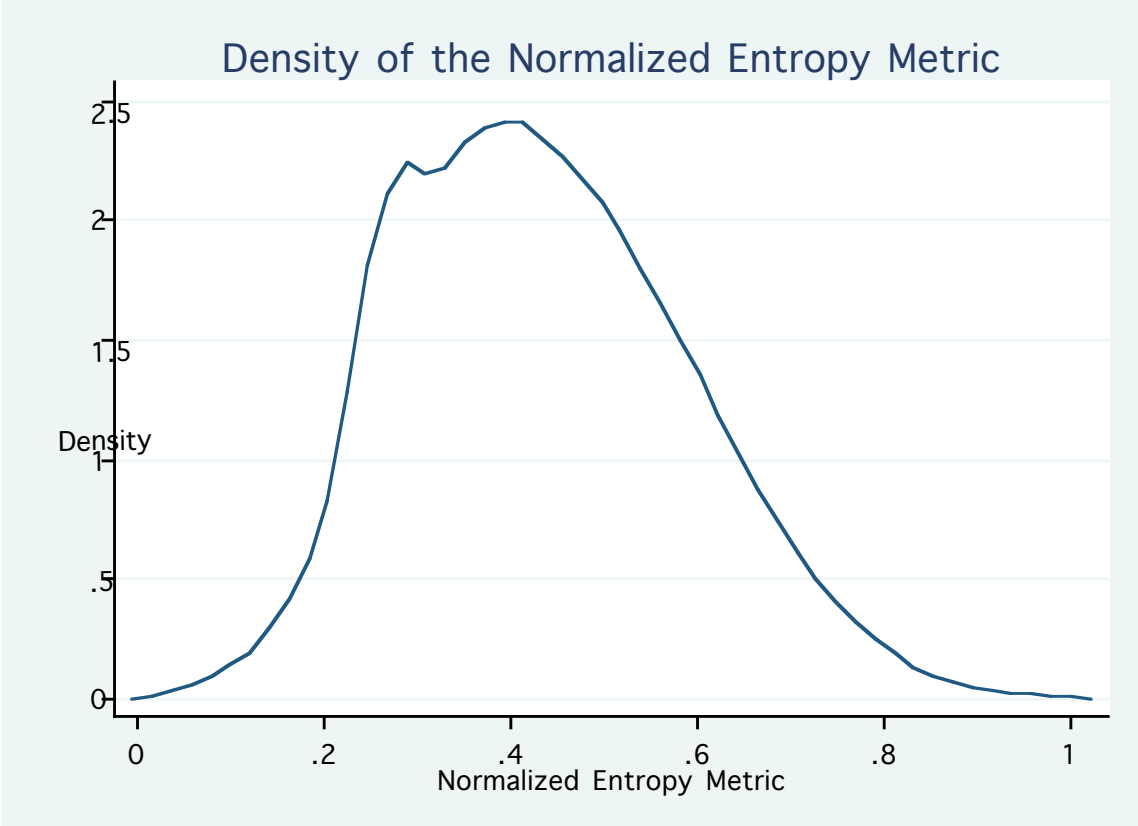

Figure 1. Empirical Density of the Entropy of Food Expenditure

An alternative measure of the way in which households allocate their budget over time is a count of number of days with nonzero expenditure. As one would expect this measure is correlated with the entropy score. However, the share of days with positive expenditure provides a less fine measure of expenditure dispersion. For example, a household that spent 95% of total food expenditure on a single day and spread the remaining 5% over the remaining sample period would, by this metric, have a dispersed expenditure pattern, whereas their expenditure is actually relatively concentrated.

Expenditure Dispersion, Household Income and Total Food Expenditure

Given a credible measure of how households spread their expenditure over time, I now consider the relationship between this measure and household income and total food expenditure while controlling for household structure. I begin by considering the reduced form relationship between shopping intensity, food expenditure and household income. Following previous work on shopping intensity (McKenzie and Schargrodsky (2005)) I specify a reduced form semiparametric model of expenditure dispersion as,

$$E(H_i \mid \text{Income, Expenditure, S}) = f\left(\log(\text{Income})\right) + g\left(\log(\text{Food Expenditure})\right) + \sum_{j=1}^{J} \theta_j S_{i,j} \quad (2)$$

where *f()* and *g()* are functions to be estimated, H is the normalized entropy of expenditure over the sample period, S are the demographic variables described above and are used to control for the effect of household structure on purchase expenditure dispersion.

If the value of health is a positive function of income (Grossman (1972)), then the estimate of the relationship between expenditure dispersion and income will provide some cautious initial evidence of the relationship between expenditure dispersion and dietary quality. In addition, ceteris paribus, the benefits of shopping for low prices will also be increasing in total food expenditures.

The semiparametric model (2) can readily estimated using the penalized spline approach described by Ruppert, Wand and Carroll (2003). In particular, I use the mixed model estimation approach as described in Ngo and Wand (2004). I provide a brief overview of this estimation approach in the appendix. The results of estimating (2) are reported in Figure 1 and Table 1. Figure 2 shows the nonparametric estimates of $\hat{f}()$ and $\hat{g}()$ along with a pointwise 2 standard deviation confidence band. Table 1 reports the estimates of the parametric coefficients[2].

[2] I estimate the model using the **nlme** (Pinheiro, et al. (2006)) library in R (www.r-project.org).

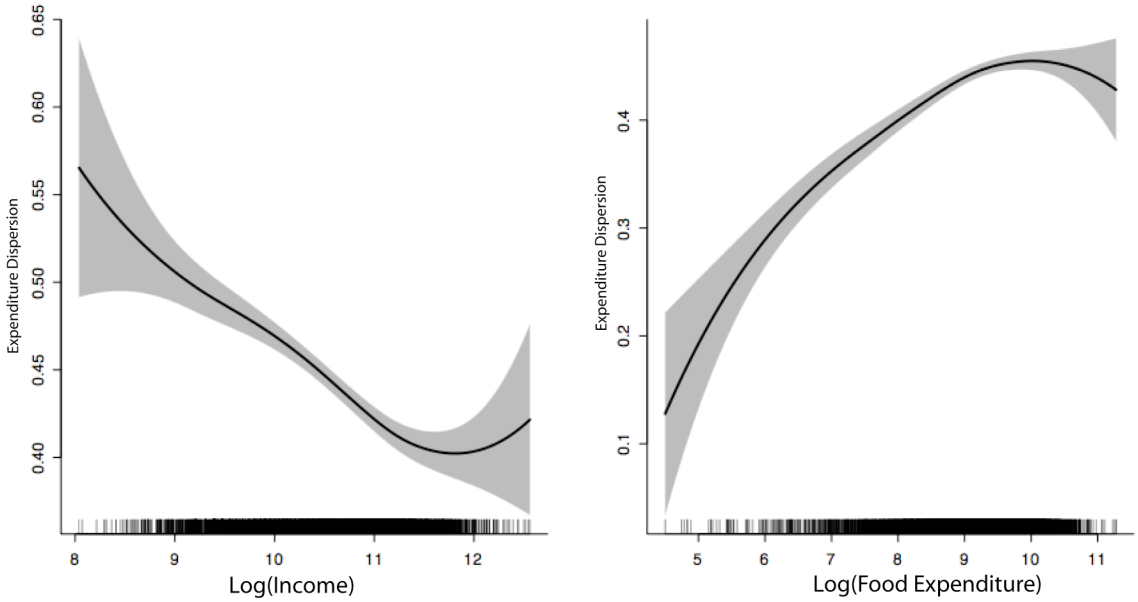

Figure 2 Semiparametric Estimates

Figure 2 shows that, as economic intuition suggests, expenditure dispersion is decreasing almost linearly in the logarithm of total household income. As noted above this may provide some weak evidence that expenditure dispersion and health are positively associated. Expenditure dispersion is increasing in the logarithm of total food expenditure, but at a decreasing rate. Again this is consistent with the notion that the returns to shopping activities are higher for households whose expenditure on food is larger.

Table 1. Explaining Expenditure Dispersion

Variable	Estimate	S.E.	P-Value
Log Household Size	0.0491	0.0067	0.00
Share of Household LT 15	0.0131	0.0131	0.32
Share of Household GT 65	-0.0245	0.0080	0.00
Age	0.0022	0.0002	0.00
Female	0.0005	0.0035	0.89
Completed High School	0.0059	0.0061	0.34
Some Post Secondary	0.0088	0.0075	0.24
Other Post Secondary	0.0059	0.0070	0.40
Completed University	0.0190	0.0076	0.01
Single Adult Household	-0.0104	0.0063	0.10
Constant	0.5050	0.1699	0.00

Amongst the household characteristics modeled parametrically, the logarithm of household size, the age of the respondent, having completed university and a being a single adult household were statistically significant. Larger household

sizes were positively correlated with statistically significant and economically large increases in expenditure dispersion. Given that households may face absolute storage constraints, this seems plausible. The dummy "Completed University" may capture a taste preference for fresh foods that need to be purchased more frequently.

These results are broadly consistent with, McKenzie and Schargrodsky (2005) who examine the response of food expenditure to the 2002 economic crisis in Argentina. In response to a decline in total expenditure, they found that while expenditure on food fell slightly, shopping frequency increased. In addition, households shopped at a wider variety of stores than before and purchased lower quality goods. Finally, the estimated effects of the demographic controls are broadly consistent with earlier work from Blaylock (1989).

Indicators of Dietary Quality

I now consider the role of expenditure dispersion as a predictor of several measures of dietary quality proposed by the W.H.O. (2003). This is accomplished by regressing the calorie shares on expenditure dispersion and a host of household demographic variables. In addition, I consider the effect of expenditure dispersion over the complete conditional distribution of dietary quality indicators via a series of quantile regressions. This is important because policy intervention will tend to focus on households with the worst diets.

Table 2 provides summary statistics of these measures for the respondents in the FOODEX. For the purposes of this research measures of dietary quality consist of the share of expenditure on food-at-home calories from the three macronutrients: fats, carbohydrates and proteins[3]. To provide a richer measure of the nutritional quality of expenditure on food-at-home, I consider two additional measures, the share of calories from saturated fats and the share of calories from fruits and vegetables. According to the W.H.O. diets high in fats and particularly diets high in saturated fats can lead to cardiovascular disease. Diets high in carbohydrates are thought to be relatively healthier, but this measure fails to distinguish between a diet high in complex carbohydrates, such as fruits and vegetables, and a diet rich in simple sugars. For this reason, I consider the additional measure, the share of calories from fruit and vegetable expenditure. In addition, diets rich in fruits and vegetables are thought to decrease the risk of

[3] Note the adding up condition, in that the share of calories from each of the macronutrients (Fats, Proteins and Carbohydrates) sum approximately to one.

cancer. Taken together, these measures provide a reasonable overview of the quality of a household's diet.

Table 2. Indicators of Dietary Quality

Variable Name	Mean	Std. Dev.	Min	Max
Share of Calories from Fats	35.14%	0.0966	0.0188	0.8541
Share of Calories from Saturated Fats	11.76%	0.0392	0.0025	0.4449
Share of Calories from Protein	17.63%	0.0557	0.0111	0.9154
Share of Calories from Carbohydrates	47.31%	0.1085	0.0000	0.9781
Share of Calories from Fruits and Vegetables	9.45%	0.0837	0.0000	0.7599

Table 3 shows that at the mean, approximately 35% of a household's caloric expenditure consists of fats, of which approximately one third were saturated fats. This is above the W.H.O. recommended levels of 15% to 30% of calories from total fat and above the recommended 10% of total calories from saturated fats. The largest share of calories was on average from carbohydrates, at roughly 47%, below the W.H.O. recommendation of 55%-75%. Amongst the macronutrients, the smallest share of calories was from protein at approximately 18%, slightly above the W.H.O. guidelines of 10-15%. Finally, the average share of calories from fruits and vegetables was 9.5%.

Finally Table 4, provides several quantiles of the empirical distribution of the measures of dietary quality.

Table 3. Selected Quantiles of Dietary Quality Indicators

Variable Name	Q10	Q50	Q90
Share of Calories from Fats	23.51%	34.83%	47.24%
Share of Calories from Saturated Fats	7.39%	11.34%	16.44%
Share of Calories from Protein	11.75%	16.90%	24.26%
Share of Calories from Carbohydrates	33.76%	47.27%	61.03%
Share of Calories from Fruits and Vegetables	1.56%	7.34%	19.71%

Table 4 provides additional information beyond the means and variances reported in Table 3. For example, using the W.H.O. guidelines, roughly half the sample consumes too high a share calories from fats and over half the sample consumes to high a share of calories from saturated fats. Simply looking at the mean of the share of calories from fruits and vegetables misses the fact that at the low end, households are consuming an extremely low share of calories from fruits and vegetables and an extremely high share of calories from fats and saturated fats and thus face elevated risk of heart disease and cancer.

Dietary Quality: Conditional Mean

I now return to the central question, is expenditure dispersion predictive of the dietary quality of food at home expenditure? Using the entropy of daily food expenditure over the sample period as a metric to quantify expenditure dispersion, I consider the relationship between this measure and dietary quality at the conditional mean and subsequently at the conditional quantiles.

A reduced form, conditional mean prediction equation for the share of at home food expenditure calories y_i from a given source (e.g. the share of calories from carbohydrates) can be written for household i

$$E(y_i \mid H, S, D) = \alpha + \gamma H_i + \sum_{j=1}^{J} \theta_j S_{i,j} + \sum_{c=1}^{C-1} \delta_c D_{i,c} \,, \tag{3}$$

where H is the normalized entropy of expenditure over the sample period, S are demographic variables and D are the province/quarter cluster dummies. I estimate (3) using ordinary least squares and report robust standard errors. All results use sampling weights provided by Statistics Canada.

As with many household expenditure surveys, prices are not directly observed. Rather than use unit values, which are known to be endogenous, Deaton (1988, Crawford, Lainsey and Preston (2003), prices are assumed constant within a geographic cluster at a point in time. Using dummy variables for each province/quarter captures the effect of differences in price structures between clusters. Ideally one would use smaller clusters, but for confidentiality purposes, the data is geocoded at the provincial level. Given Canada's developed food markets and largely urban population, it would be surprising if relative food prices were extremely different within a province (or even between provinces) for a given quarter. I test the sensitivity of the specification to the use of cluster dummies in a subsequent section.

I now describe the results of regressing five measures of dietary quality on expenditure dispersion and a set of demographic variables[4]. In addition, I include quarter and province dummies and their interactions to capture the effects of changes in relative prices between geographic/time clusters. In the interests of exposition, these coefficients are suppressed. Table 4 reports the results for the share of total calories from fat.

Table 4. Share of Calories from Fat

Variable	Estimate	Robust S.E.	P-Value

[4] Given that the explanatory variables are the same for all measures, there is no advantage to estimating them simultaneously.

Entropy of Food Expenditure	-0.0067	0.0040	0.09
Log Income	-0.0012	0.0035	0.72
Log Total Food Expenditure	0.0049	0.0030	0.10
Log Household Size	-0.0087	0.0065	0.18
Share of Household LT 15	-0.0098	0.0114	0.39
Share of Household GT 65	-0.0038	0.0077	0.62
Age	-0.0001	0.0002	0.74
Female	-0.0012	0.0031	0.71
Completed High School	-0.0018	0.0060	0.77
Some Post Secondary	-0.0062	0.0066	0.35
Other Post Secondary	-0.0084	0.0064	0.20
Completed University	-0.0263	0.0070	0.00
Single Adult Household	-0.0165	0.0070	0.02
Quarter, Province and Quarter/Province Interactions			
Constant	0.3639	0.0403	0.00

The measure of expenditure dispersion is negative and significant at the 10% level. What this tells us is that households with dispersed expenditure, that is to say households who spread their food expenditure more evenly over the sample period, have a lower share of their calories from fats than households who concentrate their food expenditure on a smaller number of purchase occasions. At the conditional mean, this increased dispersion is associated with an improvement in the nutritional quality of food-at-home expenditure.

As regards the other socio-demographic variables of interest, the logarithm of total food expenditure has a positive and significant (at the 10% level) impact on the share of calories from fat. Relative to the omitted category, "Less than secondary school", dummies that represent progressively higher educational attainment have progressively greater in magnitude coefficients. Interestingly, households with only one adult have a lower share of calories from fats.

Table 5 presents the results of estimating (3), where the dependent variables is the share of at home food expenditure calories from saturated fats.

Table 5. Share of Calories from Saturated Fats

Variable	Estimate	Robust S.E.	P-Value
Entropy of Food Expenditure	-0.0042	0.0017	0.01
Log Income	0.0018	0.0013	0.18
Log Total Food Expenditure	0.0002	0.0010	0.84
Log Household Size	-0.0036	0.0026	0.17
Share of Household LT 15	0.0073	0.0045	0.10
Share of Household GT 65	0.0030	0.0030	0.31
Age	0.0000	0.0001	0.65

Female	-0.0009	0.0012	0.48
Completed High School	-0.0012	0.0024	0.61
Some Post Secondary	-0.0015	0.0028	0.59
Other Post Secondary	-0.0034	0.0026	0.20
Completed University	-0.0113	0.0027	0.00
Single Adult Household	-0.0033	0.0029	0.25
Quarter, Province and Quarter/Province Interactions			
Constant	.1028	.01499	0.00

The entropy of food expenditure has a negative and statistically significant impact on the share of calories from saturated fats. The effect is significant at all conventional levels. This echoes the previous result concerning total fats. Households who concentrate their food expenditure on a smaller number of purchase occasions, purchase food for consumption at home with a larger share of calories from saturated fats. Their food purchases are therefore less healthy.

When considering the effect of the demographic variables, the logarithm of income has a significant and positive impact on the share of calories from saturated fats. As before, the effect of a female head of household is negative and significant. Finally, the share of calories from saturated fat is decreasing in educational attainment. Less educated households have food at home purchases that are higher in saturated fat calorie shares than more educated households.

Table 6. Share of Calories from Protein

Variable	Estimate	Robust S.E.	P-Value
Entropy of Food Expenditure	-0.0004	0.0023	0.88
Log Income	0.0070	0.0019	0.00
Log Total Food Expenditure	-0.0004	0.0019	0.83
Log Household Size	-0.0104	0.0042	0.01
Share of Household LT 15	0.0001	0.0071	0.98
Share of Household GT 65	-0.0056	0.0041	0.17
Age	0.0000	0.0001	0.79
Female	0.0005	0.0018	0.80
Completed High School	-0.0082	0.0035	0.02
Some Post Secondary	-0.0086	0.0040	0.03
Other Post Secondary	-0.0099	0.0038	0.01
Completed University	-0.0094	0.0041	0.02
Single Adult Household	-0.0065	0.0042	0.12
Quarter, Province and Quarter/Province Interactions			
Constant	0.1373	0.0228	0.00

At the conditional mean, the entropy of food expenditure does not have a statistically significant impact on the share of calories from protein. The share of calories from protein is increasing in the logarithm of household income. This may be due to a higher expenditure on meat. However the share of calories from protein is decreasing the logarithm of household size and decreasing in the share below 65. The effect of education is consistent across levels of educational attainment. Relative to "Less than High School", all other households have a slightly lower share of calories from protein.

Table 7 reports the results of (3) for the share of calories from carbohydrates.

Table 7. Share of Calories from Carbohydrates

Variable	Estimate	Robust S.E.	P-Value
Entropy of Food Expenditure	0.0061	0.0045	0.18
Log Income	-0.0057	0.0040	0.16
Log Total Food Expenditure	-0.0049	0.0031	0.11
Log Household Size	0.0168	0.0073	0.02
Share of Household LT 15	0.0100	0.0128	0.43
Share of Household GT 65	0.0093	0.0084	0.27
Age	0.0002	0.0002	0.37
Female	0.0006	0.0035	0.87
Completed High School	0.0099	0.0068	0.14
Some Post Secondary	0.0169	0.0076	0.03
Other Post Secondary	0.0185	0.0073	0.01
Completed University	0.0387	0.0079	0.00
Single Adult Household	0.0222	0.0079	0.01
Quarter, Province and Quarter/Province Interactions			
Constant	0.4989	0.0434	0.00

Expenditure dispersion has a positive effect on the share of food-at-home calories form carbohydrates at the conditional mean. Given that in this sample the mean household purchased a lower share of calories from carbohydrates than the W.H.O. recommendation, an increase in the entropy of food expenditure is associated with a higher quality diet. However the effect is not statistically significant at conventional levels.

As concerns the socio-demographic variables, the share of food-at-home calories from carbohydrates is increasing in the logarithm of total household size. As was the case for fats and saturated fats, education has is positively correlated with higher quality food expenditure. Finally the coefficient on the dummy for households with a single adult is positive and significant on the share of calories from carbohydrates.

Table 8 presents the results for the share of calories from fruits and vegetables.

Table 8. Share of Calories from Fruits and Vegetables

Variable	Estimate	Robust S.E.	P-Value
Entropy of Food Expenditure	0.0066	0.0037	0.08
Log Income	0.0062	0.0034	0.07
Log Total Food Expenditure	-0.0070	0.0024	0.00
Log Household Size	-0.0213	0.0059	0.00
Share of Household LT 15	0.0186	0.0101	0.07
Share of Household GT 65	-0.0025	0.0069	0.71
Age	0.0009	0.0002	0.00
Female	0.0006	0.0029	0.85
Completed High School	0.0004	0.0059	0.95
Some Post Secondary	0.0084	0.0068	0.22
Other Post Secondary	0.0018	0.0063	0.78
Completed University	0.0189	0.0070	0.01
Single Adult Household	-0.0124	0.0066	0.06
Quarter, Province and Quarter/Province Interactions			
Constant	0.0614	0.0345	0.08

At the conditional mean, the expenditure dispersion is predictive of improved dietary quality. The entropy of food expenditure is positively and significantly, at the 10% level, correlated with the share of calories from fruits and vegetables. This shows that households who spread their expenditure evenly over the sample period purchase food-at-home with larger share of calories from fruits and vegetables.

The age of the head of household has a statistically significant positive impact on the share of calories from fruits and vegetables. Food expenditure and income have signs that are of approximately equal magnitude, but opposite in sign. The share is increasing in the logarithm of income, but decreasing in the logarithm of total food expenditure. As before, the measure of dietary quality is increasing in education, but in this case the effect is only significant for those who have completed university. Note that in this case, the dummy variable on single adult households is negative and statistically significant.

In general, expenditure dispersion appears to be a useful predictor of measures of dietary quality for foods at home. Specifically, an increase in expenditure dispersion is correlated with a diet that is statistically significantly lower in fats and saturated fats, and higher in fruits and vegetables. Among the other variables of interest education and gender are found to be important. Relative to

the omitted category, "Less than Secondary Education", higher levels of education are in most cases associated with higher dietary quality. The effects of income and total food expenditure are mixed. These estimates are consistent with much of the previous literature.

Conditional Quantile Regression

It seems plausible that the effect of purchase frequency may differ over the conditional distribution of our measures of dietary quality. In other words, expenditure dispersion may disproportionately affect households with low or high dietary quality. Table 3 suggests this may be the case. As a concrete example, the 90th quantile is the smallest share of calories from a given source, say saturated fats, such that at least 90% of households have lower shares. This is important from a policy perspective. Policy interventions need to focus on households with "low" dietary quality. Thus if purchase dispersion is to be a useful lever for changing household dietary quality it needs to be effective for the group most affected.

To this end, I extend the analysis above by estimating conditional quantile regressions of the previously described indicators of dietary quality as a function of the entropy of food expenditure and other demographic variables. This is not the first study to use quantile regression to look at diet, as previously noted Variyan, Blaylock and Smallwood (2002) use quantile regression to study the impact of demographic variables changed over the conditional distribution of several measures of dietary quality. Indeed, they find considerable variation over the conditional distributions considered.

I now provide a brief, and necessarily incomplete, discussion of quantile regression. For a gentle introduction to quantile regression see Koenker and Hallock (2001), and for a more comprehensive treatment see Koenker (2005). The τ th conditional quantile of the share of calories from a given source as a linear function of K explanatory variables can be written:

$$Q_y(\tau \mid x) = x^{\mathrm{T}}\beta(\tau), \tag{4}$$

where $x_i = (H_i; S_i; D_i)$ the vector of explanatory variables for household i and $\beta(\tau) = (\gamma(\tau); \theta(\tau); \delta(\tau))$ is the parameter vector, as described in equation (3). For the purpose of reporting results, I focus on the 10^{th}, the 50^{th} and the 90^{th} conditional quantiles of the dietary quality measure.[5] In the interests of

[5] Estimates were obtained with R (www.r-project.org) using the quantreg library, (Koenker (2006)) and as above results are obtained using sampling weights provided by Statistics Canada.

exposition, I present only the estimates of the entropy measure of expenditure dispersion.

Table 9 reports the parameter estimates for the 10^{th}, 50^{th} and 90^{th} conditional quantile regressions of the fat share on the measure of expenditure dispersion the standard error and the corresponding p-values.

Table 9. Quantile Regression: Fat

Quantile	Estimate	Std. Err.	P-Value
Q10	0.0044	0.0415	0.32
Q50	-0.0025	0.0037	0.51
Q90	-0.0268	0.0054	0.00

The key result presented in Table 9 is that, for those households at the top of the conditional distribution of the share of calories from fat, i.e. those whose diet contains the most fat and is therefore generally considered to be least healthy, the effect of expenditure dispersion is negative and significant at all conventional levels. What this result implies is that households for whom the effect of expenditure dispersion is the largest are those households with the highest share of calories from fats. As a result these are also the households whose food choices places them at greatest risk from cardiovascular diseases.

Table 10 reports the coefficient estimates on the entropy of expenditure dispersion of the quantile regressions for saturated fats.

Table 10. Quantile Regression: Saturated Fats

Quantile	Variable	Std. Err.	P-Value
Q10	-0.0006	0.0015	0.73
Q50	-0.0034	0.0017	0.04
Q90	-0.0086	0.0026	0.00

Expenditure dispersion has a significant and negative impact on share of calories from saturated fats at the median and the 90^{th} quantile, but is not significantly different from zero at the 10^{th} quantile. Note that the magnitude of the coefficient at the 90^{th} quantile is roughly two and a half times the size of the coefficient at the median. This implies that the impact of expenditure dispersion is increasing as we move up the conditional distribution of the share of calories from saturated fats. Again, we see that the effect of expenditure dispersion is greatest for those at greatest risk of cardiovascular disease.

Table 11 reports coefficient estimates for the 10^{th}, 50^{th} and 90^{th} quantiles of expenditure dispersion on the share of calories from protein.

Table 11. Quantile Regression: Protein

Quantile	Variable	Std. Err.	P-Value
Q10	0.0067	0.0022	0.03
Q50	0.0032	0.0020	0.12
Q90	-0.005	0.0041	0.22

Recall from Table 6, that at the conditional mean, the effect of expenditure dispersion on share of calories from protein was not significantly different from zero. Table 11 shows that the effect is not statistically significant at the median at conventional levels, but is significant in the lower tail of the conditional distribution. For those at the low end of the conditional distribution, observing concentrated expenditures would lead one to predict that the household would have a greater share of calories from protein. For those households at the top of the conditional distribution, an increase in expenditure dispersion is associated with a lesser share of calories from protein. However the effect is not statistically significant.

Table 12 summarizes the results of the quantile regression analysis for the share of calories from carbohydrates.

Table 12. Quantile Regression: Carbohydrates

Quantile	Variable	Std. Err.	P-Value
Q10	0.0171	0.0073	0.02
Q50	0.0030	0.0045	0.50
Q90	-0.0004	0.0047	0.93

In contrast to the results for fat and saturated fat, the effects of purchase dispersion is positive and significant for households at the low end of the conditional distribution of share of calories from carbohydrates. At the 10^{th} quantile, the coefficient on the expenditure dispersion is positive and significantly different from zero. This says that an increase in expenditure dispersion is associated with a larger share of calories from carbohydrates. Given that households at the 10^{th} quantile are well below recommended levels of share of calories from carbohydrates, an increase in expenditure dispersion is associated with an increase in the nutritional quality of food-at-home expenditure.

Finally, Table 13 reports the quantile regression results for the share of calories from fruits and vegetables.

Table 13. Quantile Regression: Fruits and Vegetables

Quantile	Variable	Std. Err.	P-Value
Q10	0.0097	0.0014	0.00
Q50	0.0105	0.0028	0.00
Q90	-0.0048	0.0067	0.47

Consistent with the results above concerning carbohydrates, an increase in expenditure dispersion is positively and significantly correlated with the share of calories from fruits and vegetables. The estimates are significant at all conventional levels for the 10^{th} and the 50^{th} quantiles. Given that diets rich in fruits and vegetables are considered healthy, expenditure dispersion is positively associated with improved dietary quality amongst those households whose diet is of lower quality. This is important for practical policy purposes; these are the households that policy is most likely to target. The magnitude of the effect is broadly comparable at the 10^{th} and 50^{th} conditional quantiles.

Sensitivity Analysis

One important potential source of bias in the analysis presented above is evidence of respondent fatigue, noted by Ahmed, Brzozowski and Crossley (2006), between the first and second weeks of the survey. Respondent fatigue will yield less dispersed expenditures than would otherwise be observed. This suggests that the metric of expenditure dispersion will be too small. This would imply that the results above overstate the impact of expenditure dispersion on Fats and Saturated Fats and understate the impact of Carbohydrates and Fruits and Vegetables. To assess the magnitude of the bias, I construct a set of weights such that for each household/province/income group, expenditure is the same on average between the first and second weeks. The data was then weighted and the analysis described above was rerun. Results were virtually identical between the two specifications.

In the preceding analysis, dummy variables capture the relative price structure in a given location during a given quarter. I test the robustness of this approach by using information on the average prices of 10 major foods groups provided by Statistics Canada. These average prices are computed for each quarter and each province. When average prices are used identification due to variation between relative prices over time and space. Results between the two specifications were very similar, but because the averages were highly collinear, estimated standard errors were larger and the resulting estimates of the effects of demographic

variables were somewhat less significant in general. However, the effect of expenditure dispersion was very similar between the two specifications.

Note that in the conditional mean regressions, both the logarithm of income and the logarithm of total food expenditure are included. A natural alternative specification is the share of income given to food. This specification can be accommodated and tested in the framework above by a simple restriction on the coefficients: $\beta \log(X/M) = \beta_X \log X - \beta_M \log M$ when $\beta = \beta_X = -\beta_M$. In all cases, I reject the null hypothesis that these coefficients are equal for the conditional mean regressions, at all conventional significance levels.

Finally, as noted above, the entropy of daily food expenditure over the sample period is not the only possible metric for measuring the dispersion of expenditure. Alternatively one could use the number of days with nonzero expenditure to capture a similar, but not identical phenomena. Results with this alternative specification were broadly consistent with the entropy approach, but were considerably less precise.

Conclusion

This paper establishes that the way in which a household disperses their expenditure on food over time is predictive of several important measures of the nutritional quality of their food-at-home expenditure. I find that expenditure dispersion is an increasing function of total food expenditure and a decreasing function of income. In addition, I find that expenditure dispersion is a statistically significant and economically important factor in predicting several measures of dietary quality. Households with more dispersed expenditures appear to have relatively healthier food expenditure. This is an important factor in understanding the incidence of diseases that have dietary causes. It is equally important in understanding the health impacts of food assistance programs such as the USDA Food Stamp program. One important policy implication is that the USDA affects the diets of food stamp recipients as a consequence of the frequency of food stamp payments. A simple quantile regression yields additional insights into the effects of purchase frequency across the conditional distributions of the measures of dietary quality. This approach reveals that the effect of expenditure dispersion is consistently significant at "low" levels of the food-at-home quality indicators, but not significant at "high" levels of the food-at-home quality indicators.

These results will be of interest to policy makers as they focus on specific behaviors that may be amenable to policy intervention. For example, policy makers might increase the frequency of income assistance payments to encourage

household to avoid concentrating their purchases. It is interesting to note that the conventional wisdom amongst dieticians is that increasing expenditure dispersion will lead to lower quality diets. The theory being that a series of small expenditures will result in households consuming unhealthy "convenience" foods. Whilst the results of a single study are never definitive, the results presented above call this standard recommendation into question. More research on the topic is clearly needed. Furthermore, while our dependent variables are measures of dietary quality, they are not measures of health. The link between expenditure frequency and direct measures of health should be the focus of future research.

References

Income Statistics Division. 1999. 1996 Food Expenditure Survey Public-Use Microdata Files. Statistics Canada.

World Health Organization. 2003. Diet, Nutrition and the Prevention of Chronic Diseases.

Adelaja, A., R. M. Nayga, and T. Lauderbach. 1997. "Income and Racial Differentials in Selected Nutrient Intakes." *American Journal of Agricultural Economics* 79:1452-1460.

Ahmed, N., M. Brzozowski, and T. F. Crossley. 2006. "Measurement Errors in Recall Food Consumption Data." London, Institute for Fiscal Studies.

Bhattacharya, J., J. Currie, and S. Haider. 2004. "Poverty, Food Insecurity and Nutritional Outcomes in Children and Adults." *Journal of Health Economics* 23:839-862.

Blaylock, J. 1989. "An Economic Model of Grocery Shopping Frequency." *Applied Economics* 21:843-852.

Brumback, B., D. Ruppert, and M. P. Wand. 1999. "Comment on 'Variables Selection and Function Estimation in Additive Nonparametric Regression Using Data-Based Prior' by Shively, Kuoh and Wood." *Journal of the American Statistical Association* 94:794-797.

Crawford, I., F. Lainsey, and I. Preston. 2003. "Estimation of Household Demand Systems with Theoretically Compatible Engel Curves and Unit Value Specifications." *Journal of Econometrics* 114:221-241.

Deaton, A. 1988. "Quality, Quantity and Spatial Variation of Price." *American Economic Review* 78:418-443.

Eilers, P. H. C., and B. D. Marx. 1996. "Flexible Smoothing with B-Splines and Penalties." *Statistical Science* (11):89-121.

Feather, P. M. 2003. "Valuing Food Store Access: Policy Implications for the Food Stamp Program." *American Journal of Agricultural Economics* 85(1):162-172.

Grossman, M. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80(2):223-255.

Hersey, J., J. Anlike, C. Miller, R. M. Mullis, S. Daugherty, C. R. Bray, P. Dennee, M. Sigman-Grant, and H. O. Thomas. 2001. "Food Shopping Practices Are Associated with Dietary Quality in Low-Income Households." *Journal of Nutrition Education and Behavior* 33:S16-S26.

Hogbin, M., C. Davis, and A. Escobar, U.S.D.A. 1999. Preparing Nutritious Meals at Minimal Cost.

Horton, S., and C. Cambell. 1991. "Wife's Employment, Food Expenditures and Apparent Nutrient Intake: Evidence from Canada." *American Journal of Agricultural Economics* 73(784-794).

Irala-Estevez, J., G. M, J. L, U. Oltersdorf, R. Prattala, and M. A. Martinez-Gonzalez. 2000. "A Systematic Review of Socio-Economic Differences in Food Habits in Europe: Consumption of Fruits and Vegetables." *European Journal of Clinical Nutrition* 54(9):706-714.

Kaufman, P. R., J. M. MacDonald, S. M. Lutz, and D. M. Smallwood. 1997. "Do the Poor Pay More for Food? Item Selection and Prices Differences Affect Low-Income Household Food Costs." Working paper 759.

Klurfeld, D. M., and M. Krestch, H. N. N. P. Agricultural Research Service. 2001. Human Nutrition National Program.

Koenker, R. 2005. *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker, R. 2006. "Quantreg: Quantile Regression." R package version 4.02 Edition.

Koenker, R., and K. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives* 15(4):143-156.

Lynch, S. 1997. "Eating Value for Your $." vol. 2006, Dietitians of Canada.

McKenzie, D., and E. Schargrodsky. 2005. "Buying Less, but Shopping More: Changes in Consumption During a Crisis." Working paper 092, BREAD.

Nayga, R. M. 1997. "Impact of Sociodemographic Factors on Perceived Importance of Nutrtition in Food Shopping." *Journal of Consumer Affairs* 31(1):1-9.

Nayga, R. M. 2000. "Schooling, Health Knowledge and Obesity." *Applied Economics* 32:815-822.

Ngo, L., and M. P. Wand. 2004. "Smoothing with Mixed Model Software." *Journal of Statistical Software* 9(1).

Pinheiro, J., D. Bates, S. Debroy, and D. Sarkar. 2006. "Nlme: Linear and Nonlinear Mixed Effects Models." R package version 3.1-68.1 Edition.

Raper, K. C., M. N. Wanzala, and R. M. Nayga. 2002. "Food Expenditures and Household Demographic Composition in the Us: A Demand Systems Approach." *Applied Economics* 34:981-992.

Ricciuto, L., V. Tarasuk, and A. Yatchew. 2006. "Socio-Demographic Influences on Food Purchasing among Canadian Households." *European Journal of Clinical Nutrition* 60:778-790.

Robinson, G. K. 1991. "That Blup Is a Good Thing: The Estimation of Random Effects." *Statistical Science* 6:15-51.

Rose, D. 1999. "Econonmic Determinants and Dietary Consequences of Food Insecurity in the United States." *The Journal of Nutrition* 129:517S-520S.

Ruppert, D., and R. Carroll. 1997. "Penalized Regression Splines." Cornell University.

Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. *Semiparametric Regression.* Cambridge: Cambridge University Press.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27(July,October):379-423,623-656.

Theil, H. 1967. *Information and Economic Theory.* Amsterdam: North Holland.

Variyan, J. N., J. Blaylock, and D. M. Smallwood. 2002. "Characterizing the Distribution of Macronutrient Intake among U.S. Adults: A Quantile Regression Approach." *American Journal of Agricultural Economics* 84(2):454-466.

APPENDIX A

Data: Expenditure Dispersion

The FOODEX is a two-week survey conducted by Statistics Canada. Respondents keep a detailed diary of all food expenditures. These are classified into 201 food categories. Over the course of a twelve-month period, slightly more than ten thousand households were surveyed. For each household, I observe expenditure and quantity information on each of 201 food categories. Note, I use the 1996 version of the survey, as it was the last version of the survey in which detailed demographic and household income information was collected. Finally, Statistics Canada has derived a set of household weights for use with the publicly available microdata files that take into account of the survey design and non-response. When weighted, the sample is generally representative of the Canadian population. In all subsequent analysis the results incorporate these weights.

Of the existing food expenditure data series, the FOODEX is perhaps the best suited to examining the relationship between indicators of dietary quality in food at home and expenditure dispersion. In contrast to another frequently used expenditure survey, the Consumer Expenditure Survey (CEX)), the FOODEX collects information on physical quantities. This allows us to map from quantities purchased to nutrients purchased. In contrast to the CFSII, which is based on two (CSFII_94) or three (CFSII_89) 24-hour diaries, the FOODEX builds a detailed picture of food expenditure patterns over a two-week sample period. The length of the survey period and the fact that expenditures are recorded on a daily basis allows us to study how a household spends its food budget over time and to construct a novel measure of expenditure dispersion.

The Food Bureau of Agriculture and Agri-Food Canada provides a link from the 1996 FOODEX to the Canadian Nutrient File (CNF) that contains information

on 28 different nutrients for each food purchased. In all cases, nutrient quantities are adjusted for losses due to preparation. While the FOODEX is the best available survey data for answering the question posed, it is not without flaws. First, the ideal data set would allow us to link a measure of food expenditure dispersion to the quality of food actually consumed by a household. It should be emphasized that the FOODEX captures food expenditure foods rather than food consumption. If food purchases and food consumption are close on average then food expenditure can be used to talk about diet. However, the possibility of bias exists if for example high-income households are more wasteful than low-income households. As a result, some caution must be exercised in drawing conclusions about overall dietary quality from the quality of food purchased for consumption at home. Note also that the FOODEX categories describing expenditure on foods consumed away from home are not sufficiently detailed to link to nutritional measures. As a result our measures of quality do not capture this important source of calories. As a consequence, results below apply only to indicators of calories for foods at home purchases. If the quality of food households eat outside the home resembles the quality of food households eat inside the home, then the results below can be used to make inferences about the overall quality of food expenditure.

One of the great strengths of the FOODEX, particularly for this research, is the lengthy diary period. However, the length of the survey is not without cost. Ahmed, Brzozowski and Crossley (2006) show that respondents in the FOODEX demonstrate signs of survey fatigue. There is evidence of a drop off in recorded expenditure, of on average 10 percent, between the first and second weeks of the survey. This will induce some measurement error in the metric of expenditure dispersion. This effect was tested in the sensitivity analysis following the main empirical findings. Finally, as with all expenditure surveys, a number of households are excluded, due to concerns about data validity. In particular, I exclude households that did not complete the entire two-week diary period, households who purchased only a single food item, household who reported spending more than 80% of income on food and households reporting a negative income. Excluding these households yields a useable sample of 7516 households. Table 14 provides a summary of the demographic variables used in the subsequent analysis.

Table 14. Summary of Demographic Shifters

Variable Name	Mean	Std. Dev.	Min	Max
Log Income	10.6164	0.6961	8.0392	12.5602
Log Total Food Expenditure	9.1529	0.8200	4.4998	11.2763
Log Household Size	0.8266	0.5674	0.0000	1.7918

Share of Household LT 15	0.1179	0.1863	0.0000	0.6667
Share of Household GT 65	0.1944	0.3797	0.0000	1.0000
Age	48.3659	15.3396	24.0000	80.0000
Female	0.4788	0.4995	0.0000	1.0000
Not Completed High School	0.1193	0.3242	0.0000	1.0000
Completed High School	0.3928	0.4884	0.0000	1.0000
Some Postsecondary	0.1325	0.3390	0.0000	1.0000
Other Postsecondary	0.1999	0.3999	0.0000	1.0000
Completed University	0.1554	0.3623	0.0000	1.0000
Single Adult Household	0.2468	0.4311	0.0000	1.0000

In the regression analysis I retain a reasonably standard set of demographic variables that are thought to influence dietary quality (Ricciuto, Tarasuk and Yatchew (2006)). First, I include the logarithms of total household income and total household food expenditure. I then include a number of variables that capture the demographic structure of the household, the logarithm of total household size, the share of the household less than fifteen years of age and the share of the household greater than sixty-five years of age. In addition, I include a dummy variable for households headed by a single adult. In most studies of dietary quality, education is found to have a significant impact and as a result I include dummy variables for five levels of educational attainment: Less than High School, Completed High School, Some Postsecondary Education, Completed Other Postsecondary and Completed University.

APPENDIX B

There are a number of ways to estimate semiparametric models as described by (2). I employ a parsimonious approach known as penalized regression splines (p-splines) that is relatively common in the statistical literature, but is somewhat less well known in econometrics. In its present form, this approach was first proposed by Eilers and Marx (1996) and Ruppert and Carroll (1997)[6]. In the interests of exposition, I describe a simplified version of (2) with only single variable being modeled semiparametrically. The extension to two semiparemtric variables is straightforward.

The smooth functions $f(\cdot)$ or $g(\cdot)$ can be written using a cubic radial basis spline. The cubic degree radial basis spline model (sometimes called a thin plate spline) for the logarithm of total food expenditure, for household i can be written

$$f(\ln X_i) = \gamma_0 + \gamma_1 \ln X_i + \sum_{k=1}^{K} \mu_k \left|\ln X_i - \kappa_k\right|_+^3 , \tag{5}$$

[6] For a textbook length treatment of this approach see Ruppert, Wand and Carroll (2003).

where, $\kappa_1 < \kappa_2 < \ldots < \kappa_K$, denote the knot points and the functions $\left|\ln X_i - \kappa_k\right|_+^3$ are the cube of the absolute value of the difference between a value of the log of total food expenditure and a given knot point. Following the recommendation of Ruppert, Wand and Carroll (2003) the number of knots is chosen according to $K = \min(0.25 \times \text{ number of unique } X_i, 35)$ and are evenly spaced over the range of $\ln X$.

Recasting the estimation problem in matrix form, write the vector of unit values $\mathbf{v} = \left[V_1 \ldots V_N\right]^{\mathrm{T}}$, define the design matrices

$$\begin{aligned} \mathbf{X} &= \left[1, \ln X\right]_{1 \le i \le N} \\ \mathbf{Z} &= \left[\left|\ln X_i - \kappa_1\right|^3, \ldots, \left|\ln X_i - \kappa_K\right|^3\right]_{1 \le i \le N} \end{aligned}, \tag{6}$$

coefficient vectors $\mathbf{g} = [\gamma_0, \ldots, \gamma_p]^{\mathrm{T}}$, $\mathbf{m} = [\mu_1, \ldots, \mu_K]^{\mathrm{T}}$ and error term $\mathbf{e} = \left[\varepsilon_1 \ldots \varepsilon_N\right]^{\mathrm{T}}$. The estimation problem can be concisely written as

$$\mathbf{v} = \mathbf{X}\mathbf{g} + \mathbf{Z}\mathbf{m} + \mathbf{e}. \tag{7}$$

Note that if one wanted, equation (7) can be fit using ordinary least squares. However, this can result in overfitting the component being modeled nonparametrically. In order to avoid this, the influence of the extended basis function $\mathbf{Z}$ needs to be constrained in some way. Following Ruppert, Wand and Carroll (2003), based on earlier work by Robinson (1991) and Brumback, Ruppert and Wand (1999), this is accomplished by writing $\mu_k \sim N\left(0, \sigma_\mu^2\right) \forall k$. In other words, by modeling the parameters on the extended basis function as random with mean zero and finite variance. The result is a fit where the degree of smoothness is a function of σ_μ^2. Note that ordinary least squares is the special case where the variance term, σ_μ^2, is infinite.

More formally, given (7) assuming

$$\mathrm{E}\begin{pmatrix} \mathbf{m} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{8}$$

and

$$\mathrm{Cov}\begin{pmatrix} \mathbf{m} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \sigma_\mu^2 \mathbf{I} & 0 \\ 0 & \sigma_\varepsilon^2 \mathbf{I} \end{pmatrix}, \tag{9}$$

the log likelihood function can be written

$$\ell\left(\mathbf{g}, \mathbf{\Omega}\right) = -\frac{1}{2}\left(n \log(2\pi) + \log\left|\mathbf{\Omega}\right| + \left(\mathbf{v} - \mathbf{X}\mathbf{g}\right)^{\mathrm{T}} \mathbf{\Omega}^{-1} \left(\mathbf{v} - \mathbf{X}\mathbf{g}\right)\right), \tag{10}$$

where $\mathbf{\Omega} = \mathrm{Cov}(\mathbf{v}) = \sigma_\mu^2 \mathbf{Z}\mathbf{Z}^{\mathrm{T}} + \sigma_\varepsilon^2 \mathbf{I}$.

Incorporating additional parametric and nonparametric covariates is simply a matter of appending additional columns to the $\mathbf{X}$ matrix and adding the corresponding parameters to the vector $\mathbf{g}$. E.g. $\tilde{\mathbf{X}} = [\mathbf{X} \mid \mathbf{S}]$ and $\tilde{\mathbf{g}} = [\mathbf{g} \mid \theta_1 \ldots \theta_J]$, where as before, $\mathbf{S}$ is a matrix of demographic variables, $\mathbf{D}$ is a matrix of cluster dummies with parameters θ_j.