Fat Chance: Modelling the Socio-Economic Determinants of Dietary Fat Intake in China

Bhavani Shankar
Department of Agricultural & Food Economics
University of Reading
PO Box 237
Reading RG6 6AR
UK
Phone: +44 118 378 7422
Email: b.shankar@reading.ac.uk

Contributed Paper prepared for presentation at the International Association of Agricultural Economists Conference, Beijing, China, August 16-22, 2009

Fat Chance: Modelling the Socio-Economic Determinants of Dietary Fat Intake in China

Introduction

The Chinese economy has registered remarkable growth in the last few decades, with annual GDP growth rates averaging 9-10% in the course of 1980-2005 and household incomes rising steadily. The proportion of the absolutely poor in China came down from 80% in 1978 to less than 12% in 1998, while the proportion of the extremely poor was reduced from 20% to 6% during the same period (Du, *et. al.*, 2004). Concomitantly with economic growth, Chinese households have also enjoyed significant nutritional improvements.

At the same time, some worrying trends have been detected. China is now often held up as a classic illustration of the 'double burden of malnutrition' wherein undernutrition in some segments of the population co-exists with obesity and disease related to overnutrition in other segments (Popkin, *et. al.*, 1995). "Overnutrition" can bring with it massive health costs as witnessed in the developed world. Annual mortality from coronary heart disease and stroke, "non-communicable" diseases to which overnutrition is a significant contributor, are estimated to reach 800,000 and 3 million respectively by 2030 in China (Chen, 1995). Dietary patterns are increasingly based on significant consumption of animal fats, edible oils and processed foods, while some traditional foods have started to disappear from diets. Therefore, attention is increasingly being paid to the economic analysis of the demand for specific groups of food products in China and/or the analysis of macronutrients in the form of energy, fat and carbohydrates (*e.g.* Du, *et. al.*, 2004; Yen, *et. al.*, 2004; Fang and Beghin, 2004).

Increasing fat content of Chinese diets has been flagged as an especially troubling trend. Based on the WHO definition of a high-fat diet as one in which energy from fat is in excess of 30% of total energy, Du *et. al* (2004) showed that in the eight years from 1989 to 1997, the average incidence of high-fat diets increased from 13.8% to 37.5%. The most obvious downside of elevated dietary fat intake of course is the contribution that it can make to heart disease, stroke and various cancers. However, Bray and Popkin (1998) have made the argument that dietary fat intake is also a significant contributor to obesity levels. The conventional wisdom holds that obesity is only related to energy balance, *i.e.* the intake and expenditure of calories, and that diet composition issues, such as fat composition, are not relevant to obesity. However, Bray and Popkin (1998) conducted a cross-country regression that found the prevalence of overweight to be strongly related to the proportion of dietary energy from fat. They argue that higher fat intake leads to higher energy consumption due to passive overconsumption and the low thermic effect of fat. An implication of their finding is that fat composition in Chinese diets may be imposing a set of second order costs on China by contributing to obesity, in addition to the first order effects on non-communicable diseases.

Under these circumstances, determining the socio-economic and demographic determinants of fat intake in China becomes an important research question. Du *et. al.* (2004) modelled the probability that energy from fat in Chinese adults diets exceeded the 30% of total energy threshold using a probit model. Controlling for a variety of socio-economic variables, they focused on the income elasticity, finding that a doubling of income would lead to a 40% increase in the incidence of high-fat diets. We continue investigation of the relationship between key socio-economic drivers, particularly income, and fat intake here, but go beyond the previous work by Du, *et. al.* (2004) by specifying and estimating a richer model that allows the marginal effects of socio-economic drivers to vary across the fat intake distribution. We describe in the next section how quantile regression is particularly well suited to the analysis of nutrition problems, where the focus of over or under nutrition begs special attention to the tails of conditional intake distribution.

2. Regression approaches to determinants of nutrient intake

A vibrant literature exists on econometric modelling of the determinants of nutrient intakes. Calorie intake, being closely related to hunger satiation, is the most frequently modelled variable. One strand of the literature subsumes this within traditional demand modelling approaches, and estimates systems of demand equations using broad groupings of food categories, and indirectly derives calorie demand elasticities using appropriate conversion factors. While estimation of demand systems enables full appreciation of cross and own price and other effects in determining nutrient outcomes, nutrient values of food products within broad categories can vary tremendously, and the use of aggregated groups can result in substantial bias in estimation due to the fact that intra-group substitutions are ignored (Behrman and Deolalikar, 1987). Therefore the other strand undertakes direct estimation, expressing nutrient value as a function of income, prices and other socioeconomic variables (*e.g.* Abdulai and Aubert, 2004; Behrman and Deolalikar, 1987; Behrman and Wolfe, 1984). This is the approach taken in this paper.

Most previous regression-based approaches in the literature have modelled either the conditional mean of nutrient intake, or the probability of intake exceeding a threshold as functions of covariates However, conditional mean modelling treats various parts of the conditional distribution of intake identically, and constrains the marginal effects of independent variables to be the same throughout the distribution of nutrient intake, which is a significant shortcoming in this particular problem setting. From a public health and nutrition policy perspective, particular parts of nutrient intake distributions are likely to be of more interest than others. In the fat intake case, the behaviour of those in the upper tail of the conditional intake distribution, *e.g.* those in the upper 10%, is likely to be of more interest than those in the lower reaches of the distribution. It is natural to be more concerned about the ‘worst performers’ in any performance related grouping. Most importantly, the effect of an independent variable may reasonably be hypothesised to vary across the distribution of the dependent variable. For instance, given a specification of other conditioning variables, it is likely that the effect of increased income on fat intake for those currently in the bottom 10% of the intake distribution will be different from those in middle, or in the top 10%.

Limited dependent variable approaches, such as probit regressions, express probabilities of exceeding cutoff points as a function of covariates. By doing so, they offer a potential way to train focus on particular segments of the intake distribution. For example, as in Du, *et. al.* (2004), the cutoff may be 30% of energy intake and a probit regression would model the effect of a covariate on the probability of energy from fat exceeding this cutoff. However, this common practice of reducing continuous information on intakes into information on a small number of categories involves a loss of information. For instance, logistic regression involving a category such as ‘more than 30%’ would treat a data point with intake of 31% identically to a data point with intake of 80%. Note also that, although logistic regression can help obtain focus on segments of the distribution of interest, the effect of a covariate on the underlying distribution remains constant throughout the distribution of the dependent variable.

Quantile[1] regression (QR) in this context would allow the impact of the explanatory variable to vary along the whole range of nutrient intake. Koenker and Basset (1978) noted that a set of causal variables could have myriad effects on the distribution of the dependent variable. They proposed that conditional quantiles of the dependent variable be estimated as linear functions of covariates, whereas simple linear regression expresses only the conditional mean as a linear function of covariates. By allowing conditional functions to be defined at any chosen quantile, QR leads allows the effect of a given set of covariates to flexibly vary across the distribution of the dependent variable. Unlike the standard applications of logistic regression methods on intake data, no sacrifice of information is entailed, while assumptions about the form of the parametric distribution (such as

[1] ‘Quantile’ is general terminology for what may be referred to as percentile, decile, quartile, etc in specific cases.

the logistic distribution in logistic regressions) are avoided. Classical linear regression reduces to a special case of QR where the effects of covariates are constrained to be the same across the distribution of the dependent variable. QR methods have gained popularity among economists and ecologists over the last decade. They hold particular promise in applications to nutrition problems where dietary excess and/or inadequacy questions beg particular attention to the tails of distributions, although there seem to be only a small number of applications so far (*eg.*, Variyam, *et. al.*, 2002; Gustavsen and Rickertsen, 2002). Accessible introductions to QR methods are available in Koenker and Hallock (2001), and Cade and Noon (2003).

3. Data and Methods

The data used in this study derive from the China Health and Nutrition Survey (CHNS)[2], a large Sino-American collaborative longitudinal data collection exercise with numerous research outputs (http://www.cpc.unc.edu/projects/china/publication provides a listing). The dietary portion of the CHNS recorded food consumption in great detail over a 3-day period in each wave. Extensive individual as well as household-specific information was collected, and CHNS also makes available a carefully constructed income variable for each household. Supplementary community-level food price information is also available and is used here in conjunction with the individual and household-specific information. These data were intersected with nutrient conversion information from China Food Composition (FCT) Tables 2002 produced by the Institute for Nutrition and Food Safety, China CDC (Chinese Centre for Disease Control and Prevention). The FCT provides conversion values for over 1500 food items consumed in China, enabling a good degree of precision in calculating nutrient values. The data used here are from the 2000 wave, and is at the individual level. From the original dataset, all individuals between the ages of 20 to 45 were included in our estimation sample, which is consistent with Du, *et. al.* (2004). After deleting observations with missing information, data on 2612 individuals were used in analysis.

[2] We are grateful to the CHNS data archive for use of the data. However, any opinions expressed here, and any shortcomings in this research are our own.

Percentage of calories derived from fat was specified as the dependent variable. The dataset contained estimates of both individual as well as household income. While an argument can be made for the inclusion of both income variables in estimation, the correlation between these incomes was in excess of 90%, preventing the inclusion of both in practice. It can be argued that in the closely knit family settings in developing countries, household incomes are more relevant than individual incomes in determining consumption and nutritional outcomes. Hence household income was the income variable used in the analysis. Demographic variables included Family Size (number of people in the family), Education (completed years of education of individual), Gender (Male: 0, Female:1) and Age. An 'Urban' dummy was also included (0 if located in a rural area, 1 if in urban area). In addition to income and demographics, data on a set of community-level prices for 11 key food products were included. For every food in community price dataset, the CHNS survey could not find a price recorded in at least some local communities. To account for the non-availability of prices for specific foods in some communities, the price variables in the regression were specified as appearing multiplicatively with dummies indicating availability.

A regression was first estimated using Ordinary Least Squares to provide a basis for comparison with the quantile regression. The functional form for all regressions was initially specified as quadratic in all continuous independent variables (*i.e.*, income, household size, education and age) in order to allow for nonlinear effects of these variables on fat composition. The quadratic terms for both the education and age variables were consistently found to be insignificantly different from zero in all regressions, and were subsequently dropped. The continuous variables were centred at their medians to assist interpretation.

In the QR, 73 different conditional quantile functions were estimated for % of energy from fat, starting with the 0.05 (5%) quantile and proceeding in 0.0125 increments until the 0.95[3] (95%)

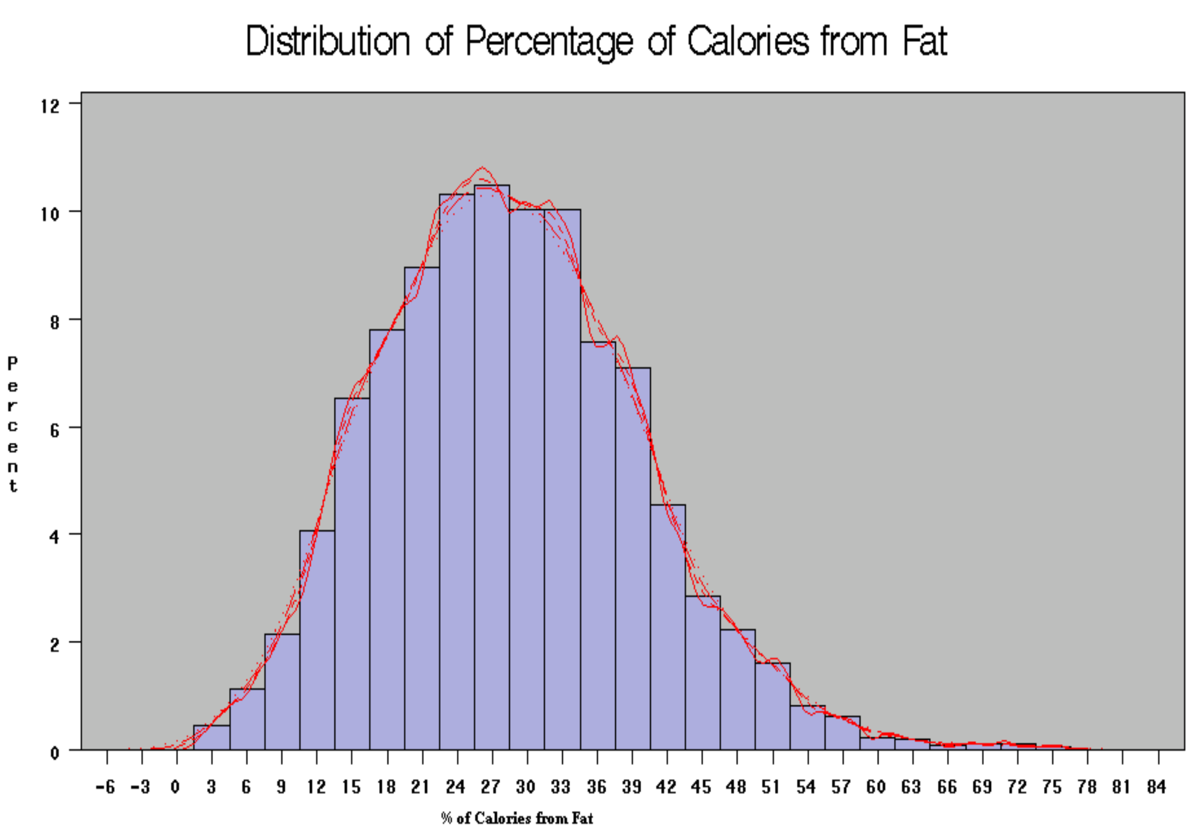
[3] The starting and finishing points were set at 0.05 and 0.95 because extreme quantiles are occasionally known to encounter problems with stable confidence interval estimation.

quantile (*i.e.*, 0.05, 0.0625, 0.75, 0.875, 0.1…0.95). A Markov Chain Marginal Bootstrap (MCMB) (He and Hu, 2002) was implemented to compute confidence intervals for every quantile estimate.

Results

Figure 1 below shows the histogram of percentage of calories from fat in the sample. It also shows fitted kernel density estimates (estimated at multiple bandwidths in case the estimate proves sensitive to bandwidth specification). Clearly, excess fat composition of diets is a significant problem given the proportion of sample exceeding the WHO threshold of 30% of total energy intake. Some positive skewness is evident, with some individuals consuming more than 75% of energy intake in the form of fat. 42% of the sample has fat intake contributing more than the threshold 30% of energy intake.

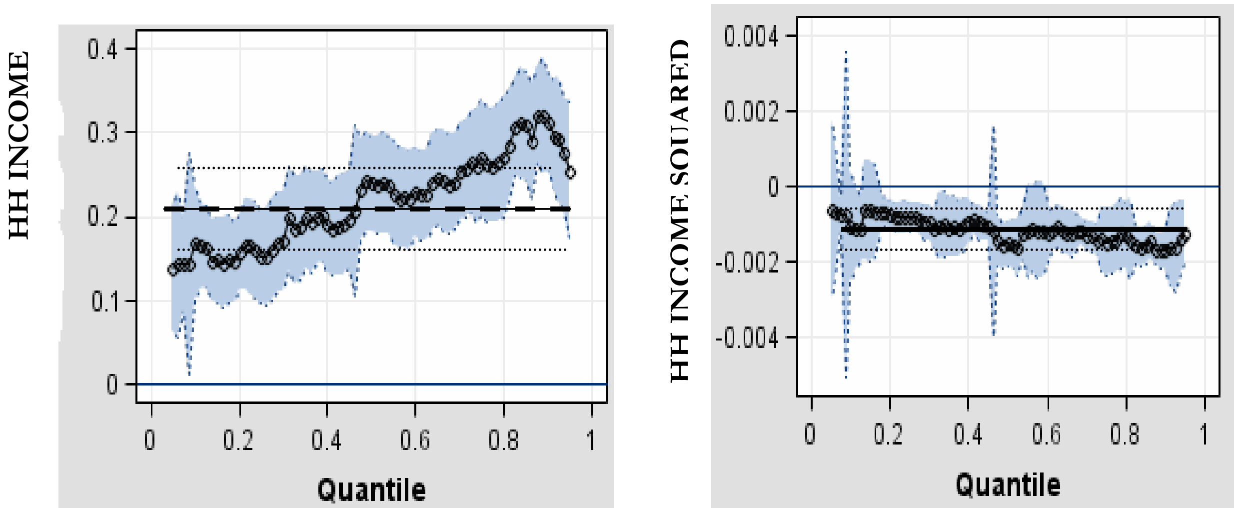
Figure 1. Distribution of % of Calories from Fat



With 19 independent variables, and estimates available for 73 different conditional quantiles, presentation of results necessarily has to be selective. Graphs plotting estimates against quantiles for individual variables can adequately capture and summarise the cross-quantile variation in estimates. We employ such graphs for a selection of independent variables, selected either because of their importance in explaining nutrient intake in previous literature or because they demonstrate interesting cross-quantile variation. These graphs (Figures 2 to 5 below) show the quantile estimates along with the bootstrapped 95% confidence intervals shown as shaded areas around the estimates. In addition, the OLS estimate and the 95% confidence interval are also shown as dashed/dotted straight lines superimposed on the QR graphs.

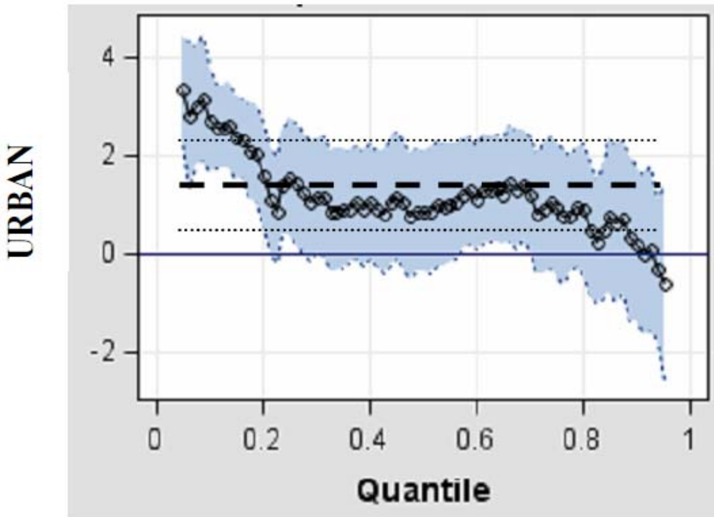
It is readily apparent from the set of figures 2 to 9 that in most cases at least some portion of the QR estimates lie outside the OLS confidence intervals. This suggests that the simple conditional mean shift implied by the OLS model is not plausible. The first covariate we discuss is income, the most discussed variable in developing country nutritional intake studies. The estimates for income and its square are shown in figure 2. The OLS as well as QR regression results in Figures 3 and 4 at any given quantile suggest that increasing income generally increases the proportion of energy from fat, but that this effect levels off as income gets larger. Since the continuous variables have been centred to the median, the parameter estimates of the level (non-squared) terms indicate the marginal effect at the sample median. Thus at the median income in the sample, the OLS results indicate that every 1000 Yuan ($120 in 2000) increase in household income results in increased fat contribution to energy intake of 0.22%. This reinforces Du *et. al.*'s (2004) finding that as the Chinese get more affluent, increasing amounts of energy are derived from fat. However, the QR results paint a more nuanced picture, and show that a more worrying situation emerges when estimates are allowed to vary across conditional quantiles. At the extreme, for the 0.05 conditional quantile, the computed effect of every 1000 Yuan increase from the median income level is only 0.14%, while at parts of the higher end of the conditional intake distribution (i.e. where the 'overnutrition' worry is high), the effect is higher than 0.22% OLS effect. For example, at the 0.875 conditional quantile, a 1000 Yuan increase from the median income produces a 0.32% increase in fat contribution to energy intake.

Figure 2. OLS and Quantile Regression Estimates: Household income and income squared



Urban Chinese diets are known to have transformed more significantly than rural diets. In fact, recent clinical data has shown that more rapidly westernizing urban diets in China have contributed to increases in body-weight, BMI, waist circumference, cholesterol levels and blood glucose compared to rural areas. The urban effect is seen clearly in Figure 3 below. The OLS estimate implies that an otherwise identical individual would get 1.3% more of their energy from fat living in an urban location rather than a rural one. However, the QR results show that this location effect dies off at the upper reaches of the distribution. Beyond the 0.7 conditional quantile, the QR results in fact do not predict a statistically significant urban vs rural location effect. This suggests that at the most unhealthy levels of fat density conditional on covariates, the problem is as much a rural one as it is urban.

Figure 3. OLS and Quantile Regression Estimates: Urban vs rural



Price effects on nutrient demands in developing countries are usually important, but also inevitably complex (Behrman, 1995). With a range of substitutions occurring between foods, only part of which may be evident due to aggregation problems, signs may be counter to immediate intuition, and interpretations may not always straightforward. Nevertheless, we illustrate key price effects on fat density of diets in figures 4, 5 and 6 below. Eggs are a moderately important source of dietary fat in China. Figure 4 shows the effects of variation in egg prices on fat density of diets. The OLS results suggest that as egg prices rise, consumers substitute for egg consumption in such a way that energy derived from fat is lowered. At the higher tail of the conditional distribution where fat levels are at their unhealthiest, the price effects are seen to be strongest, more than double the magnitude of the conditional mean effect predicted by OLS. Rice prices, on the other hand, have little effect on fat density, whether measured by OLS or QR methods. The confidence intervals of the OLS estimate as well as the QR estimates for rice price largely fall on either side of zero, indicating a lack of statistical significance. Figure 5 illustrates the effects of soybean oil price on fat density of energy intake. Ng. *et. al.* (2008) have found that edible oil pricing is a key element in nutritive health, and that declining edible oil prices have contributed to elevated fat intakes. This is borne out by the OLS as well as QR estimates for soybean price effects. The QR estimates in this case are largely within the confidence interval of the OLS estimate, but it is worth noting that in contrast to the case with egg prices, the influence of edible oil prices are lowest at the higher end of the intake distribution.

Figure 4. OLS and Quantile Regression Estimates: Egg Price

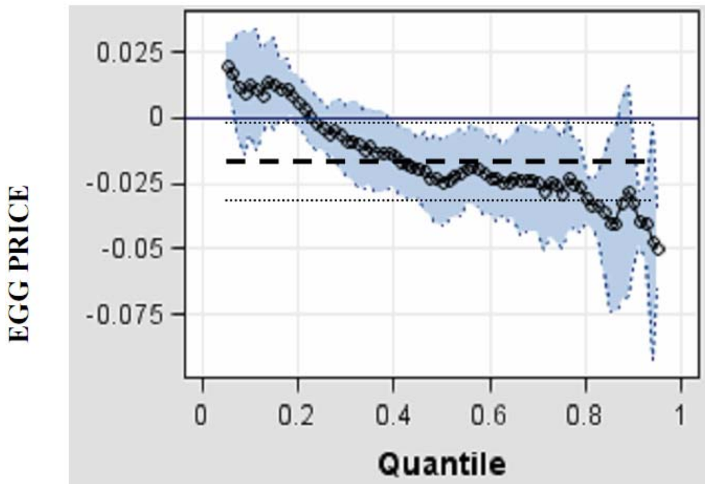


Figure 5. OLS and Quantile Regression Estimates: Rice price

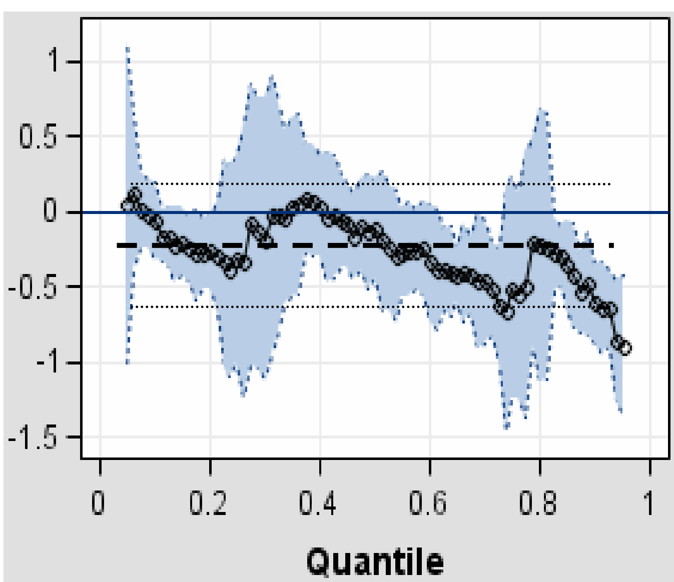
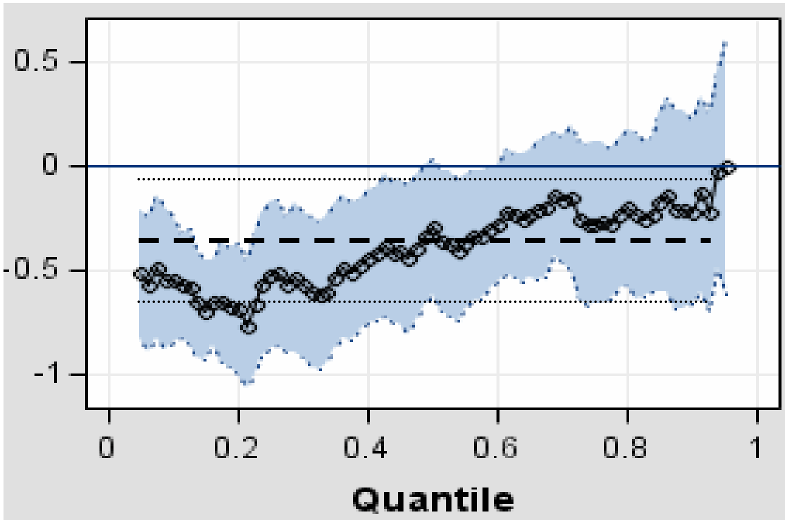


Figure 5. OLS and Quantile Regression Estimates: Soybean Oil Price



Conclusion

Quantile Regression methods have much to offer the investigation of the determinants of dietary intake. Dietary inadequacy or excess occurs at the tails of nutrient and food intakes, and it seems intuitive that intake responses in these areas will differ from elsewhere along the intake distribution. Our application of QR methods to examine the drivers of a key aspect of dietary health, fat density of energy intake in China, has revealed the following key insights: (i) Fat density increases with income, but worryingly the income effect is more pronounced at the upper conditional tail of fat intake (ii) On the other hand, while it is confirmed that an urban location contributes to higher fat density for the most part, this effect disappears at the upper conditional quantiles, suggesting that, at the most unhealthy levels of fat density conditional on covariates, the problem is as much a rural one as it is urban.

References

Abdulai, A., and D. Aubert (2004) “Nonparametric and Parametric Analysis of Calorie Consumption in Tanzania”. *Food Policy*, **29**, 113-129.

Behrman, J. and B. Wolfe (1984) “More Evidence on Nutrition Demand: Income Seems Overrated and Women’s Schooling Underemphasized”. *Journal of Development Economics*, **14**, 105-128.

Behrman, J. and A.B. Deolalikar (1987) “Will developing country nutrition improve with income? A case study of rural south India”. *The Journal of Political Economy*, **95**, 492-507.

Bray, GA, Popkin, BM “Dietary fat intake does affect obesity!” Am J Clin Nutr 1998 68: 1157-1173

Cade BS, Noon, B.R. (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and Environment* **1**: 412-420

Chen CM (1995). Eating patterns - a prognosis for China. *Asia Pacific Journal of Clinical Nutrition* 4 S1:24-28.

Du, S., Mroz, T., Zhai, F. and B.M. Popkin (2004) “Rapid income growth adversely affects diet quality in China, particularly for the poor!” *Social Science and Medicine*, **59**, 1505-1515.

Fang, C. and J.C. Beghin (2002) “Urban demand for edible oils and fats in China: evidence from household survey data”. *Journal of Comparative Economics*, **30**, 732-753.

Gibson, J. and S.D. Rozelle (2002) “How elastic is calorie demand? Parametric, nonparametric and semiparametric results for urban Papua New Guinea”. *The Journal of Development Studies*, **38**, 23-46.

Guo, X., Popkin, B.M., Mroz, T. and F. Zhai (1999) “Food price policy can favourably alter macronutrient intake in China”. *Journal of Nutrition*, **129**, 994–1001.

Koenker R, Bassett, G.S. (1978) Regression quantiles. *Econometrica* **46**: 33-50

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

Ng, Zhai, Popkin, Impacts of China's edible oil pricing policy on nutrition, Social Science & Medicine, Volume 66, Issue 2, January 2008, Pages 414-426,

Popkin, B.M., Paeratakul, S. and K. Ge (1995) “A review of dietary and environmental correlates of obesity with emphasis on developing countries”. *Obesity Research*, **3**, 145s-153s.

Yen, S.T., Fang, C. and S. Su (2004) “Household food demand in urban China: a censored system approach”. *Journal of Comparative Economics*, **32**, 564-585.

Variyam JN, J. Blaylock, and D. Smallwood. (2002) Characterizing the distribution of macronutrient intake among U.S. adults: A quantile regression approach. *American Journal of Agricultural Economics*, **84**: 454-466