New developments in fruit and vegetables consumption in the period 1999-2004 in Denmark – A quantile regression approach

Hansen A.H.

Paper prepared for presentation at the 12th EAAE Congress
'People, Food and Environments: Global Trends and European Strategies',
Gent (Belgium), 26-29 August 2008

New developments in fruit and vegetables consumption in the period 1999-2004 in Denmark - A quantile regression approach

Hansen A.H. [1]

[1] University of Copenhagen, Institute of Food and Resource Economics, Copenhagen, Denmark

Abstract - **The development in the consumption of fruit and vegetables in the period 1999-2004 in Denmark was investigated using quantile regression and two previously overlooked problems were identified. First, the change in the ten percent quantile samples decreased. This could have been caused by changes in the distribution of covariates. Therefore, the counterfactual decomposition of Machado and Mata (2005) was used and the methodology established that the change was not caused by alterations in the distribution of covariates but by changes in the coefficients and therefore a change in behaviour. The reason for this development is probably due to low income groups becoming relatively more income constrained since the gap to the high income group have grown considerably at the lower end of the distribution. The second problem was that the education inducing gap became larger in 2004 indicating that uneducated people have not responded as well to the health related information flow. These results suggest that information campaigns have not been as successful as previously thought; more importantly the results indicate that information campaigns alone will do a poor job in solving the identified problems. Other instruments targeting uneducated and low income groups more directly are needed. Instruments which make fruit and vegetables relatively cheaper would undoubtedly have an effect on low income groups and send a strong signal to the uneducated population.**

Keywords: **Quantile regression. Counterfactual decomposition. Expenditure distribution**.

I. INTRODUCTION

Governments initiate information campaigns expecting a behavioral effect in the population. In a cost benefit context, it is important to investigate the effectiveness of such campaigns. Governments are also interested in the inequality in the population because of the stability of society and moral concerns. Therefore, monitoring factors influencing health is important in a cost benefit context but also important in evaluating if government initiatives are increasing equality or inequality. There is perhaps reason to fear that information campaigns can increase inequality in health if only subgroups of the population respond to the campaign message. It could be theorized that the educated part of the population would be more likely to respond to the campaign message if decoding it needs formal schooling; if education helps develop strategies to change behavior; if persons more responsive to authorities self-select into becoming educated1[1]. Validating this is obviously an empirical question. Recently in Denmark, as well as in other comparable countries, there have been campaigns with the purpose of increasing the population's intake of fruit and vegetables(FaV) occurring at the same time as an increased intensity of the information flow of health related issues. This suggests that the gap in intake of FaV between uneducated and educated subgroups should increase if the above hypothesis is correct. A purpose of this poster is to: *Investigate if the gap in FaV consumption between educated and uneducated people has increased.* Because low consuming groups have a high return on consumption, the development in this group is especially interesting. Also, if low income groups consume less then this is a barrier that policy makers have partial control over, as opposed to if the household dislikes the taste. *Investigate the development in low consuming groups with emphasis on any links to income.*

[1] Self-selection exists because authorities say: get an education; then they says: eat more fruit and vegetables.

II. MODELLING EXPENDITURE DISTRIBUTIONS

A model is needed to perform the counterfactual decomposition. The model used is the quantile regression model which estimates quantiles conditional on covariates

$$Q_{\theta}(y \mid z) = z^{\mathrm{T}}\beta_{\theta} \quad (1)$$

As explained in Machado and Mata[1], this is the brick needed for estimating unconditional densities based on different covariates distributions.

A. Decomposing the changes in the distribution

Denote by $f(FaV(t))$ the sample estimator of the density of FaV (the log of spending on fruit and vegetables) at time t based on the observed sample and by $f^*(FaV(t))$ the model based estimator of the density based on the generated sample $\{FaV^*(t))\}$ in step 4 below. The counterfactual density will be denoted by $f^*(FaV(2004);Z(1999))$, for the density that would result in 2004 if all covariates had the distribution of 1999. The changes in the quantiles of the sample estimator, $q_{\theta}[f(FaV(2004))]-q_{\theta}[f(FaV(1999))]$, are decomposed by using the model based estimator into the coefficients effect and the covariates effect as can be seen in table I.

B. Unconditional densities implied by the conditional model

Let $FaV(t)$, $z(t)$, t=1999,2004 be spending and the covariates age, sex, income and education. Let $g(z;t)$ be the sample density of the covariates at time t. Generating a random sample from the FaV density that would prevail in t if model (1) were true and the covariates were distributed as $g(z;t)$ can be accomplish by:

1. Generating a random sample of size k from a uniform distribution: $\tau_1,\ldots,\tau_k$.
2. From the sample data at time t and each $\{\tau_i\}$ estimate $Q\tau_i(FaV|z;t)$ yielding estimates $\beta(\tau_i)$, i=1,.,k.
3. Generating a random sample of size k from $g(z;t)$, $\{z_i^*(t)\}$,i=1,.,k.
4. Finally $\{FaV^* = z_i^*(t)^{\mathrm{T}}\beta(\tau_i)\}, i = 1,\ldots,k$

A sample from $f^*(FAV(2004))$ and $f^*(FAV(1999))$ can be calculated by letting t equal 2004 and 1999 whereas a sample from the counterfactual density $f^*(FAV(2004);Z(1999))$ can be accomplished by letting t=2004 in step 1-2 and t=1999 in step 3-4.

III. EMPIRICAL ANALYSIS

A. Data

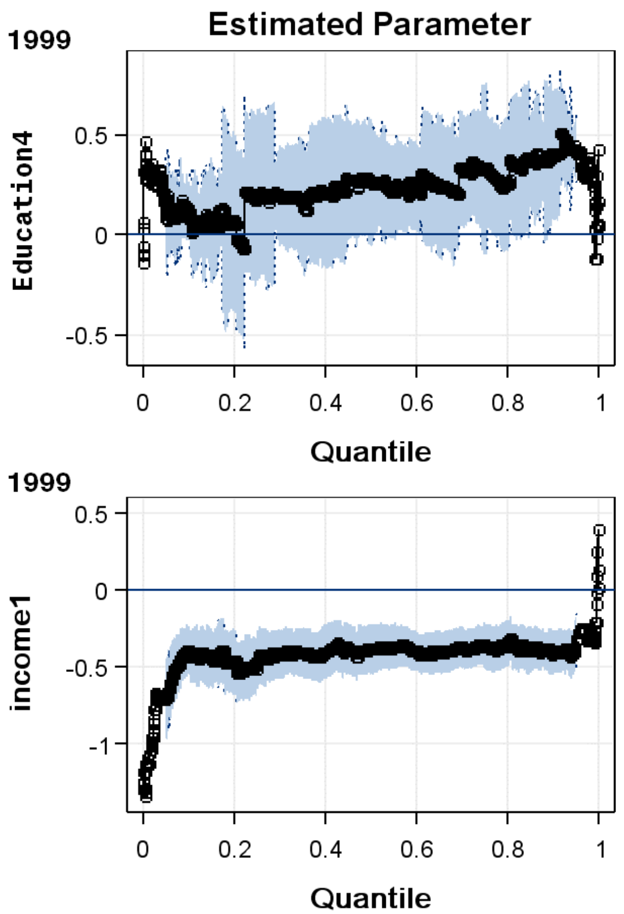
The data consists of weekly shopping trips registered by the Consumer Scan Panel of GfK(Gfk.com).
The cross-sections analyzed were obtained by summing over all of the expenditures during the years 1999 and 2004 for each household. Only households included all year were used.

B. Development in sociodemographic differences

The plotting of income1 in figure1 and 2 show that households having low income spend less than the high income group and in 2004(1999) it is visible that this income gap increases when moving from the 0.4(0.1) quantile and down through the distribution. This effect implies that the distribution is more dispersed than that for the high income group[2]. Households in the lower end of this distribution are hit twice; belonging to the low income group means that they will spend less, and, since their distribution is much more dispersed, households in the lower end of the distribution spend much less. The gap gradually becomes larger in 2004 while 1999 seems fairly constant when moving from the 0.4 quantile and down the distribution; revealing a low income group with inequality increasing more than the high income group from 1999 to 2004. This evolution is a cause for concern. The marginal return to consuming more FaV is large in the lower end of the distribution and the consumption could be increased if the household was given money moving it to the high income group. A pressing question is whether this group has become relatively more income constrained? An explanation for this development could be that households disliking FaV - belonging to the lower quantiles- seem to sacrifice FaV spending relatively more when being more income constrained. The low income group could have become more income constrained because of increased prices and new goods and services competing for the households' money. Consumption of FaV increases with education across the whole distribution. The highly educated group plotted in figure1 display high variability in the estimates as are visible in the broad confidence bounds. Households with a middle length education (not shown) have much smaller variability in the estimates and they are significant over the whole distribution in 2004 and from the

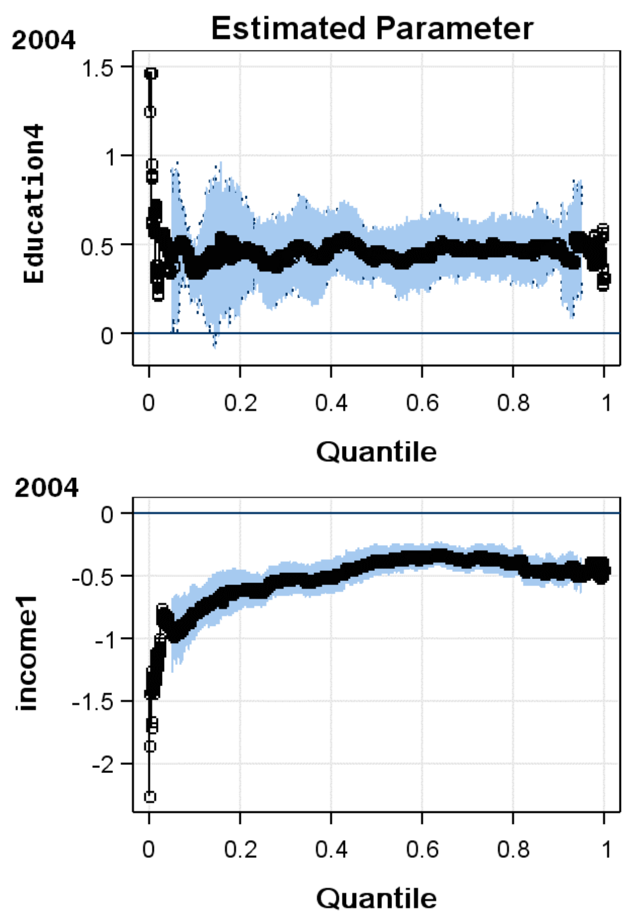
[2] A lower proportion of households belonging to the low income group will contribute towards reduced inequality.

Fig. 1 Quantile plots for 1999



Parameter estimates are for highest education group(Education4) compared to uneducated and lowest income group(income1) compared to highest income group.

Fig. 2 Quantile plots for 2004



Parameter estimates are for highest education group(Education4) compared to uneducated and lowest income group(income1) compared to highest income group.

0.5 quantile and up in 1999 even the coefficients are smaller.

In 1999, education had an inequality increasing effect as Education4 increases when moving up the distribution. This effect disappeared in 2004 which has the same constant effect over the entire distribution. From a comparison of the change in Education4 it can be concluded that the gap between the highly educated and the uneducated has increased over the period to roughly 0.5 over the entire distribution. The gaps between the other education classes and the uneducated also increased during the period (not shown). As the former discussion argued, these results could be attributed to the better capacity of the educated population to comply with campaigns and the health information flow in general.

C. Counterfactual decomposition

As Table I surprisingly shows, the 10th quantile sample change is negative.

Table 1 Counterfactual decomposition

	10thquant	25thquant	Median	75thquant	90thquant
2004 quant[1]	11,248	11,863	12,355	12,790	13,193
1999 quant[2]	11,284	11,703	12,163	12,582	12,932
Change [3]	-0,036	0,160	0,192	0,209	0,261
Residual	0,006	0,027	0,005	-0,009	0,009
Coefficients[4]	-0,05716	0,094599	0,14421	0,15128	0,19116
Covariates[5]	0,01507	0,03794	0,04260	0,066094	0,060319

[1]Sample 2004 quantiles: $q_\theta[f(FaV(2004))]$; [2] Sample 1999 quantiles: $q_\theta[f(FaV(1999))]$; [3] Change=Coefficients+Covariates+Residual ; [4]Coefficients:$q_\theta[f^*(FaV(2004);Z(1999))]-q_\theta[f^*(FaV(1999))]$; [5]$q_\theta[f^*(FaV(2004))] - q_\theta[f^*(FaV(2004);Z(1999))]$

The effect of changes in the distribution of the covariates are opposing indicating that it should be behavioral changes that drive the unexpected evolution; this is also the case as the model based changes in the coefficients are negative. An explanation for the development could be that the low income group reduced spending in the lower end of the distribution. Throughout the rest of the distribution expenditures have increased, the proportional growth being larger at the highest quantiles. Both covariates and coefficients contribute to the actual evolution of the location estimates and the behavioral effect is quantitatively more important than the effect of covariates at each of the estimated quantiles. As the residuals account for a relatively small portion of the sample change the model works fairly well.

IV. CONCLUSIONS

Countless previous publications have shown inequality in health behavior among different sociodemographic groups. Unsurprisingly, this study also reveals age, sex, income and educational differences in spending on fruit and vegetables. Inequality in education and income are spreading to the health area. Relatively cheaper FaV would benefit low income groups more because spending on FaV amounts to a much larger part of their available budget. Two surprising new results were demonstrated. The sample 10th quantile became smaller from 1999 to 2004 and a counterfactual decomposition showed that this development was due to behavioral changes. This fact has been hidden in former analyses because of approaches based on mean statistics. Nevertheless, it is an important result as households in the lower end of the distribution have higher marginal returns to consumption. This becomes more important from a government perspective if inequality is a concern. Arguably, some of this development could be attributed to low income groups becoming relatively more income constrained since the gap to the high income group have grown considerably at the lower end of the distribution. This also means higher inequality within the low income group than within the high income group. The educational gap has expanded from 1999 to 2004. This is in accordance with the hypothesis that information campaigns only have a limited effect on uneducated populations. Earlier publications showed a positive trend in the average FaV consumption during the period 1995-2001, which came to a standstill in 2004 and expressed a need for new and stronger efforts. This study has shown that an information based approach alone is unlikely to have the desired effect since it is not able to reach the uneducated part of the population and low consuming, low income groups. An information based approach alone would also be unlikely to have the desired effect in the future as most of the reachable population has already understood the message. Other methods should be undertaken directed at the two at risk groups; instruments which make FaV relatively cheaper would undoubtedly have an effect on low income groups and send a strong signal to the uneducated population.

V. REFERENCES

[1] Machado J. A. F., Mata J. (2005) *Counterfactual Decomposition of changes in wage distributions using quantile regression.* J. Appl. Econ. 20: 445-465.