Forecasting the consumption effect of taxing foods containing saturated fat

Aslak H Hansen

University of Copenhagen, Institute of Food and Resource Economics
aslak@foi.dk

Paper prepared for presentation at the 113th EAAE Seminar "A resilient European food industry and food chain in a challenging world", Chania, Crete, Greece, date as in: September 3 - 6, 2009

Forecasting the consumption effect of taxing foods containing saturated fat

Aslak H Hansen
University of Copenhagen, Institute of Food and Resource Economics

Abstract. The purpose of the current paper is to explain how one can forecast the effect of an elected tax on saturated fat on the demand for butter. The tax is to take affect from the first of January 2010 in Denmark. The tax is supposed to affect the consumption of saturated fat and especially high consuming households are of interest. Quantile regression is thus better suited than mean regression. Interest centre on at risk groups with larger consumption, but we are also interested in a simple measure that measure the total effect of the tax change, i.e. the unconditional quantile. The former can easily be obtained from the quantile regression while it is proposed to use simulations in the latter case.

In mean regression a close form formula for calculating the unconditional mean from the conditional mean exist; unfortunately this is not the case for quantile regression. Hence, simulations are needed. The principle in the proposed method is the same as the methodology used in a recent published paper for comparing labour income distributions. A refinement of this methodology is suggested.

Keywords: Quantile Regression, Simulation, Healthy Diet, Public Policy.

1. Introduction

The citizens in Denmark as well as over most of the world are becoming increasingly obese. Partly to confront this development a new tax reform has been set up increasing the prices of some food products believed to affect the obesity rate. This tax reform is special since it not only taxes certain products but introduces a new element; the taxation is based on specific nutrients contained in the food products. One of these specific nutrients is saturated fat. Butter contains a large amount of saturated fat and the reform raises the price of butter considerable in a relative sense and butter is therefore a good case for assessment of the reforms effect on food demand behaviour. It could be interesting to try and forecast what the effect of the reform will be even before the reform is taking effect; we can then compare these a priori estimates with ex post estimates that are obtained after the reform has been introduced. If the a priori and ex post estimates are in reasonable concordance then this gives some support for the accuracy of the suggested methodology and gives an indication of the accuracy of similar a priori estimates of different but similar policy experiments. In this paper we consider how the expected result of the tax reform can be calculated most appropriately.

In section 1.1 we introduce the Danish tax reform of 2010 which suit as an excellent case for the proposed methodology. In section 2.1 we introduce a conditional quantile model that can be used for the policy experiment and we then consider two ways of obtaining the unconditional quantile based on this conditional model. In section 2.2.1 we consider a previously used approximation that is inspired by the law of iterated expectations and section 2.2.2 introduces a new approximation based on simulation for obtaining the unconditional quantile. Finally we draw some conclusions in section 3.

1.1 The Danish tax reform of 2010

Their have already been some estimates of the reforms health effects. These estimates are in the form of average price elasticity's that are translated into expected increases in life expectancy The purpose of this section is to motivate why our suggest methodology in section 2 is appropriate and to set the policy experiment into a context.

In this section we will describe the motivation underlying the tax reform. The aim of the reform is to give incitements to consume less of goods which are believed to have a negative effect on health.

A relatively high consumption of tobacco, saturated fat and sugar contained goods contributes to incidence of folk illness. The most serious health effects of smoking are cancer, cardiovascular disease and chronic pneumonia. Too much sugar intake can cause diabetes and overweight, and too much saturated fat intake can increase the risk for cardiovascular disease and cancer. The consumption of goods with high sugar content, e.g. ice cream, chocolate and soda is more than double since 1975. The Danish government therefore wishes to strengthen people's health through the tax reform. The purpose of this reform is to help the population to live longer and have more healthy life years without illness.

The tax reform involves a rise of the taxes on tobacco, ice cream, chocolate, sweets and sugar contained soda. Moreover, the reform will impose a tax of 25 kr. pr. kg saturated fat in dairy products (except milk) and in vegetable oil and fat. A tax increase on saturated fat in dairy products and vegetable fat is a new type of tax.
The government has set up a commission with the object of finding cost-effective ways of increasing the Danes life expectancy which is low relatively to citizens in comparable countries under other circumstances. In the analysis made by the commission it is shown that a tax of 20 (Danish) kroner pr. Kg saturated fat can reduce the consumption of saturated fat with around 3 percent. This result is obtained by using price elasticity's estimates taken from a number of research papers. The estimates are mean estimates. We will also consider how to estimate the effect of the reform but our methodology will focus on estimating higher quantiles hereby getting a fuller picture of the subgroups which are most at risk. The analysis of the commission then continues with the estimated implications of the reduction in consumption on life expectancy. The 3 percent reduction on the total consumption of saturated fat means that an estimated 3.800 life years will be gained over a 10 year lifespan. This corresponds to a increasing of the average lifetime for man and women with a little less than one month.
These benefits arise primarily from the reduced risk of cardiovascular disease. The introduction of a tax on saturated fat will with the uncertainty that connect to the calculation, make around 0,9 to 1 billion kroner net benefit to the state.
Taxes on unhealthy goods should be used carefully; because Denmark is part of an international economy. Increasing the taxes too much will not strengthen the people's health, but merely increase the incitement to engage in cross-border shopping. The implication is that an optimal tax policy should take the possibility of cross-border shopping into account. The tax on saturated fat in dairy products and vegetable fat will be imposed in 2010. This means that the benefit by the reduced consumption of tobacco, saturated fat and sugar contained goods increases. Only the tax increase on unhealthy goods won't solve the problems of a rising incidence of folk illnesses. Every single Dane should also realize his personal responsibility for one's own health. But the tax reform can help the development of people's health in the right direction and support a healthy lifestyle.

2. Theory

In section 2.1 we introduce a conditional quantile model that can be used for the policy experiment, i.e. the linear quantile model. We then consider two ways of obtaining the unconditional quantile based on this conditional model. In section 2.2 we consider a previously used approximation that is inspired by the law of iterated expectations. Finally we introduce an approximation based on simulation for obtaining the unconditional quantile.

2.1 Quantile regression

Many empirical analyses cannot be fully accomplished by simply looking at the conditional mean of a regression function. The simple mean regression cannot satisfactorily fit the data in many cases if the distributions of the error term is non-Gaussian and asymmetrical and/or have a fat tail. Also, dependent variables may include outliers which often happen in household survey data. Under these circumstances, the conditional mean estimator can be sensitive to the presence of outliers and thus can be misleading. The quantile regression introduced by Koenker and Basset (1978) is an alternative method to provide estimates of a dependent variable corresponding to various quantile values of the explanatory variables so that a more comprehensive picture of conditional distribution of a dependent variable can be obtained. Koenker and Basset (1987) proposed a generalization to the linear quantile regression model from the location model. Quantile regression is especially well suited for analyzing food demand behaviour where we often are interested in subpopulations with consumption far from the median consumption.
Consider the following linear regression model:

$$y_i=\beta_i x_i+v_i \qquad (1)$$

$\boxed{\theta}$

where the error term is assumed identically and independently distributed. The th quantile of the conditional distribution of y given x is defined as

$$Q\theta(yx)=\mathrm{infy}F(yx)\geq\theta \qquad (2)$$

where $F(yx)$ is a conditional cumulative density function. Based on equation 1, the conditional quantile is given by:

$$Q\theta(yixi)=\beta ixi+ Q\theta(uixi) \qquad (3)$$

where βixi and $Q\theta(yixi)$ are not separately identified. Therefore, the quantile regression equation is

$$yi=Q\theta(yixi)+wi,\ i=1,\ldots,n \qquad (4)$$

where $wi=vi\text{-}Q\theta(uixi)$. Following Koenker and Bassett's (1978) estimation procedure, the conditional quantile $Q\theta(yixi)$ is obtained by minimizing the following objective function:

$$\mathrm{min}\beta i\in i{:}yi\geq\beta ixin\theta yi\text{-}\beta ixi+i\in i{:}yi<\beta ixin(1\text{-}\theta)yi\text{-}\beta ixi \qquad (5)$$

Note that when $\theta=12$, the minimization of (5) yields the Least Absolute Deviation (LAD) estimator which is an important special case for the quantile regression. The estimated conditional quantile is $Q\theta(yixi)=\beta\theta xi$ in which the estimated coefficient is a function of the specific quantile value θ. Koenker and Bassett (1978) established the asymptotic normality of $\beta\theta$

$$\sqrt{n}(\hat\beta_\theta-\beta_\theta)\xrightarrow{D}N(0,\sigma_\theta^2E[xx']^{-1}) \qquad (6)$$

Where $\sigma_\theta^2=1-\theta\theta/f_{u\theta}^2(0)$ is the density function of u_θ.

2.2 Quantile regression and a policy experiment

How large is the policy effect on individuals at most at risk?

Obviously society will benefit more from introducing the policy if subpopulations with high consumption are more sensitive to the price increase accomplished by the tax than low consuming subgroups and vice versa. This is so because the marginal benefit from reducing consumption is higher for the highest consuming individuals. This non-constant marginal benefit over the distribution of consumption is the reason why quantiles are more appropriate than the mean when evaluating the effect of the policy introduction. Ideally we would like to target individuals that could benefit from reduced consumption, e.g. we definitely do not want persons suffering from anorexia to reduce their consumption of butter. Thus, our object is to estimate quantiles in the population after the tax introduction. We therefore need a model for the policy experiment. We estimate a conditional quantile based on the sample, i.e. the consumption as it is before the policy introduction. A tax introduction will change the price level of the considered good and likely the consumption of the population. We can predict how the consumption will be in the population after the introduction for specific subgroups. To be concrete we consider a simple sample with only two types of covariates. The first one is the average price and the second the gender of the individual buying the consumption bundle. The dependent variable is grams of saturated fat in the consumption bundle. We then have the linear regression model

$$y_i=\beta_0+\beta_{sex}I_i^{sex}+\beta_p x_i^p+v_i \qquad (7)$$

here I_i^{sex} is an indicator of the gender of individual *i*, x_i^p is the price paid by individual *i* for the consumption bundle and y_i is the grams of saturated fat in this bundle. We then estimate the conditional quantile by applying a linear programming algorithm to the problem in equation (5); the estimates are $\hat\beta_{0\theta},\hat\beta_{sex\theta},\hat\beta_{p\theta}$. We can use this model to predict the effect of the reform on consumption in different subpopulations: men at the θth quantile who paid x_i^p before the reform will now be faced by a price $x_i^p+\tau$ and consume $\hat\beta_{0\theta}+\hat\beta_{sex\theta}+\hat\beta_{p\theta}(x_i^p+\tau)$ and woman $\hat\beta_{0\theta}+\hat\beta_{p\theta}(x_i^p+\tau)$.

(Note: Both men and woman are assumed to be affected the same way by prices; we could have relaxed this assumption by introducing interaction terms).

The consumption level at any specific quantile depends on whether we are talking about a woman or a man and on the price level we are considering. The conclusion is that we need to be very specific about which subgroup and at what price level we are considering when talking about the consumption at a quantile. It would be much simpler to communicate what the effect of the reform is on consumption if we had estimates of the consumption effect in the whole population, i.e. the Danes consumption at different quantiles after the reform. A second reason why a simpler measure is warranted is related to the wish of estimating the effect of the tax reform on life expectancy. In section 1.1 we mentioned some estimates of the effect of the tax reform on life expectancy. These estimates are based on mean estimates and therefore not able to take into account that the marginal benefit over the distribution from reduced consumption is non-constant. If we had estimates of the unconditional quantiles over the entire distribution before and after the reform then this information could be utilized to get better estimates of the effect of the reform on life expectancy than the estimates mentioned in section 1.1. We now consider two different approaches for finding the unconditional quantile.

2.3 Unconditional densities implied by the conditional model

In the case of mean regression it is straightforward to find the unconditional mean when a regression model (conditional model) is available. One simple sum (or integrates) out the conditioning covariates of the expectation operator, i.e. utilizes the law of the total probability (the law of iterated expectations). In the case of quantile regression, finding the unconditional value of the dependent variable is another matter; when a conditional model is available receiving the unconditional quantile is non-trivial, because no exact closed form formula similar to the law of iterated expectations exists. We first consider a closed form approximation for finding the unconditional quantile from the conditional model.

2.3.1 A closed form approximation

To estimate the unconditional quantile after a tax increase a popular choice is formula (8), e.g. Gustavsen et. al (2006) and Gustavsen et. al (2008)

qθ=1ni=1nxi-pβ-pθ+xip+τβpθ=τβpθ+βθ1ni=1nxi=τβpθ+βθx (8)

Where xi=xi-p,xip is the covariates for observation *i* in the sample, βθ=β-pθ,βpθ is the estimate of the

θth quantile in the linear quantile model and τ is the levered tax. The approximation in (8) is inspired by

the well known Law of Iterated Expectation (or Law of Total Probability) which is of course not an approximation.

Note that the estimated effect on the θth quantile of the tax reform is:

qEffectθ=τβpθ+βθx-βθx=τβpθ (9)

, we see that the tax simple shift the quantile.
To better understand when the approximation in (8) works well and the opposite we construct two examples to illustrate either case.

Example where the closed form approximation works well
To better see when the approximation (8) is a good one we consider an example of a data generating process where (8) is exact when stochastic uncertainty is set aside. We consider a regression model

$y=\beta 0+\beta 1x1+v$ (10)

where x1 is a Bernoulli variable with equal probability of the two possible events; u is a normally distributed error term, i.e. $v\sim N(0,1)$. This simple model can also be described as a mixture model, i.e.

$$y=\alpha 1N\beta 0,1+\alpha 2N\beta 0+\beta 1,1 \quad (11)$$

where $\alpha 1=\alpha 2=0.5$ and $\alpha 1$ is the mixture component representing the probability that y comes from the normal distribution with mean $\beta 0$. Now, the 10th quantile of the distribution $N(\beta 0,1)$ is $q101=u10+\beta 0$ and the 10th quantile of the distribution $N(\beta 0+\beta 1,1)$ is $q102=u10+\beta 0+\beta 1$, where v10 is the 10th quantile of a standardized normal distribution ($v\sim N(0,1)$). The tenth quantile of y is $q10=v10+\beta 0+\alpha 2\beta 1$. We know from section 2.1 that the quantile regression model produces consistent results. Assume that our estimate in fact satisfies $\beta 0=\beta 0,\beta 1=\beta 1$ and that we know the distribution of the error term. Our estimates of the 10th quantile in the two subsamples are $q101=v10+\beta 0=v10+\beta 0=q101$ and $q102=v10+\beta 0+\beta 1=v10+\beta 0+\beta 1=q102$. When applying the approximation in (???) we will use that the share of observation in the two samples are $\alpha 1$ and $\alpha 2$ which is implied by the assumption of no

statistical uncertainty: $q10=\alpha 1v10+\beta 0+\alpha 2v10+\beta 0+\beta 1=v10+\beta 0+\alpha 2\beta 1=q10$, thus the approximation

is correct.

(Note: We have added $u10$ in all the calculation above because we assumed that we knew the distribution

of the error term; in general, we do not know this of course.).

Why did the approximation work?
The approximation gives us a weighted average of the quantiles in the subpopulations, i.e.

$$q10=v10+\beta 0+\alpha 2\beta 1=\alpha 1q101+\alpha 2q102 \qquad 12$$

and this was also the case for the 10th quantile of y:

$$q10=\alpha 1q101+\alpha 2q102 \qquad 13$$

While our definition of $q10$ makes (12) true in general, (13) is due to our selection of normal

distributions for both subsamples. We note that (13) are still true no matter how we change the distribution of the mixture components. The reason why (13) hold is as mentioned due to the shape of the cumulative distribution functions in the two subsamples or rather to a property of them: Any normally distributed random variable will by a linear transformation be transformed into another normally distributed random variable. By a similar argument we can conclude that if the distributions of the subsamples belongs to the same class, e.g. the normal distributions, and this class has the property that a random variable within the class can be transformed to another random variable within the same class by a linear transformation then the approximation in (8) is a good one. Examples of classes of distribution with this property are the classes belonging to the location scale family, e.g. Normal, Cauchy etc.

Example where (8) is inaccurate
We now construct a simple example that highlights why a simulation approach can be advantages compared to using the approximation (8). We have a sample available which consist of two subsamples.

To be concrete we assume that they are of equal size and label them $yi1,yi2i=1,\ldots,100$ with

$y11<y21<\ldots<y1001$ and $y12<y22<\ldots<y1002$; the tenth quantile of the respectively subgroups are then $q101=y101$ and $q102=y102$ We assume $y52=y151$ which imply that the tenth quantile of the total sample is $q10=y52=y151$, in addition we assume that $y102=q102$ is far from $q10=y52=y151$ relatively to how close $q10$ is to $q101$. This can happen if e.g. the cumulative distribution function of subsample 2 is quite flat in the interval $q10,q102$ The approximation in (8) then estimates $q10$ as $q101+q102/2$ which is far from the true $q10$. In this simple example we could of course calculate $q10$ directly, i.e. by sorting the sample and finding $q10=y20=y52=y151$, but we assume that this is not possible. This example is only for illustrating why simulation can be better than using the approximation in (8), in general we do not have a sample available for direct calculation of $q10$, e.g. when we evaluate a policy experiment.

2.3.2 Approximation by simulation

The closed form approximation can give inaccurate results as we saw in the last example. The reason is that it does not utilize the shape of the cumulative distribution functions in the two subsamples in the interval between the two quantiles. If we do not have a special relationship between the shape of the two cdf's in the subsamples, e.g. as outlined in the first example, the approximation is likely to be inaccurate. We are therefore interested in an approximation that is able of taking the shape of the cdf's in the interval between the quantiles into account. As mentioned in the end of the last example it would be easy to calculate $q10$ if we could use the union of the subsamples, but this option is not available because we do not know how samples from the sub distributions will be after the policy experiment has taken effect. A solution suggest itself, namely to construct subsamples from the subpopulations as it will be after the reform and then calculate the quantile from the union of these subsamples. This is in fact not difficult to accomplish by simulation. For simplicity we will first consider how it is done before the policy experiment and within the model introduced in the previous section. Again we consider model (10). We still do not need to assume that the error term is normal distributed, but we do so for notational simplicity. We are given a random sample (x,y) and our object is to generate a sample (y) using the model (10) and our sample (x,y). If (y) is close to (y) then we know that the model is good at fitting the data. Before we can generate a sample (y) we need to know how we can generate random draws in the subpopulations defined by the covariates (x). To make a draw from subsample 1 (remember that we only have two subgroups because x is a Bernoulli variable) we use the inverse transformation theorem: A random draw from a stochastic variable Z0, with cdf F0(z) can be accomplished by generating a random draw u from a uniform distribution and then calculate, F0-1(u). We note that the inverse cdf for subsample 1 is F0-1u=β0u=β0u and a random draw from subgroup 1 can be accomplished by generating a random draw u from a uniform distribution and then calculate, β0u=β0u. Generating a random draw from the unconditional distribution is straightforward when using the mixture interpretation introduced in (11).

First we make a random draw from the distribution of mixture components; assume x=0 (x=1) is drawn. Then conditional on this draw, we draw a random draw from the corresponding distribution which are distribution 1 (distribution 2). The two steps gives the random draw β0u (or β0u+β1u if x=1 where drawn in the first step) from the unconditional distribution. Repeating this many times and we get the sample (y). We will now introduce a price variable making it possible to conduct a policy experiment. We have both a continuous and a discrete variable in our distribution of covariates which is called mixed data in Lee et al. (2003). For ease of exposition we assume that the price can only take two values, $x2∈pL,pH$. The regression model becomes

$$y=β0+β1x1+β2x2+v \quad (14)$$

, or as a mixture model

$$y=α1Nβ0+β2pL,1+α2Nβ0+β1+β2pL,1$$

$$+α3Nβ0+β2pH,1+α4Nβ0+β1+β2pH,1 \quad 15$$

To generate a random sample before a tax is imposed we do as before. 1. Generate a random draw from the covariate distribution. 2. Conditional on this generate a random draw from the corresponding distribution. When we introduce the policy experiment we need to change step two slightly: Assume that mixture component 2 was selected in the first step. In step two we generate a random draw u from a uniform distribution and estimate $β0u,β1u,β2u$ using the random sample (x,y) and model (2.4); a draw from the uncondional population before a tax reform is $β0u+β1u+β2upL$ and after the reform: $β0u+β1u+β2upL+τ$. Thus, we have assumed that the price level has changed as a result of the reform and that the parameters is unaffected by this. The assumption is similar to the one used in the approximation (???) when predicting the effect of the tax change. By repeating the process we end with a sample $(yτ)$. To calculate the $θ$th quantile as it is after the reform we simple calculate the $θ$th quantile of the sample ($y\tau$), i.e. $q\tau\theta$. Because the simulation approach introduces bias we calculate $q\tau\theta$ many times

and report the middle value and the 95% confidence bounds.

Resume of approximation by simulation:

Let (x,y), be the dependent variable and the covariates. Let *g*(x) be the sample density of the covariates.

Generating a random sample from the density y that would prevail if model (1) were true and the

covariates were distributed as *g*(x) can be accomplish by:

1. Generating a random sample of size *k* from a uniform distribution: $u_1,\dots,u_k$.

2. From the sample data x,y and each $u_i i=1,\dots,k$ estimate

$Qu_i(y|x)$ yielding estimates

$$\beta ui=\beta ui,\ i=1,..,k. \qquad (16)$$

3.1 Generating a random sample of size *k* from *g*(*x*), $x_i i=1,\dots,k$

4.1 A random sample from the density of y is:

$$y=xi\beta uii=1,\dots,k \qquad (17)$$

The method for generating a random sample from the unconditional population when the population is modelled as a conditional quantile model was suggested in Machado and Mata(2005). They used the method to construct counterfactual exercises in wage distributions, i.e. to evaluate how much of the wage distribution was coursed by changes in the distribution of covariates, e.g. higher education, and how much was coursed by changes in returns to these covariates. Hansen(2008) used a similar method as Machado and Mata(2005) to compare distributions of expenditures on fruit and vegetables. Here the decomposition was to a higher extent used to correct the differences in sample selection bias in the two samples. In this paper we suggest yet another purpose for random samples generated in the above manner; namely to evaluate the effect of a policy experiment on different quantiles of the distribution. We still need a couple of steps to finalize the evaluation of a tax introduction:

3.2 Transform the price level by the reform, $xii=1,\ldots,k \rightarrow xi\text{-}p,xip+\tau i=1,\ldots,k$

4.2 A random sample from the density of y after a reform:

$$y\tau=xi\text{-}p,xip+\tau\beta uii=1,\ldots,k \tag{18}$$

5.1 The before reform quantile

$$q\theta y=\inf xi\beta ui\in y:ui\geq\theta \tag{19}$$

5.2 The after reform quantile

$$q\tau\theta y\tau=\inf xi\text{-}p,xip+\tau\beta ui\in y\tau:ui\geq\theta \tag{20}$$

6. The effect of the reform on the θth quantile is

$$qEffect\theta=q\tau\theta y\tau\text{-}q\theta y \tag{21}$$

One question remains as how to calculate the distribution of the covariates, $g(x)$. Machado and Mata(2005) suggest dividing continuous variables into groups or bins. A continuous variable is split into ten bins and the bins lengths are decided by the 10^{th} quantile, 20^{th} quantile, ..., 90^{th} quantile of the continuous variable, i.e. if an observation of the continuous variable is larger than the 10^{th} quantile but smaller that the 20^{th} quantile it is put into the second bin. After doing this to all continuous variables they are left with a set of discrete variables and $g(x)$ is then simple estimated by the frequency estimator, i.e. the maximum likelihood estimator. It is then easy to generate a sample from the estimated distribution of $g(x)$. Another simple way of generating a sample from $g(x)$ is suggested in Hansen(2008). Here a sample from $g(x)$ is simply obtained by resampling with replacement from the sample data x. A more appropriate way of generating samples from $g(x)$ is to treat the data sample x as a mixed data sample. Li et al.(2003) and Ouyang et. El.(2006) takes this approach by introducing kernels into the estimation of $g(x)$. It is well known that even more important than the choice of kernel in probability and density estimation is the selection of the bandwidth. They therefore choose the bandwidth by a datadriven method, i.e. they choose the bandwidth that minimizes the mean square error of the probability/density estimator by cross-validation. The way of generating an unconditional sample implied by a conditional quantile model suggested by Machado and Mata(2005) could therefore be improved by treating the covariate distribution as a mixed data sample and then apply datadriven methods in the selection of the estimator of the probability/density function $g(x)$.

3. Conclusion

In this paper we considered the problem of constructing an unconditional quantile from a conditional quantile model. This is an important problem because in many policy relevant problems the marginal benefit of the policy effect is uneven distributed over the distribution, hence the effect of a policy can be better evaluated by using information of the effect over the entire distribution compared to using only the mean effect.

We showed that a closed form approximation used in earlier published papers can be a good one under certain circumstances. But since we can not know if these requirements are satisfied we instead considered an approximation that works under more general conditions. This approximation is based on simulations and is an adaption of a methodology introduced in a recently published paper. We also suggested a way of improving this methodology.

References

Gustavsen, G. W., and Rickertsen K. (2006). “A Censored Quantile Regression Analysis of Vegetable Demand: The Effects of Changes in Prices and Total Expenditure.” Canadian Journal of Agricultural Economics 54:631-45.

Gustavsen, G. W., Jolliffe, D. And Rickertsen K. (2008). “Censored Quantile Regression and Purchases of Ice Cream”. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, Florida, July 27-29.

Hansen A. H. (2008). ”**New developments in fruit and vegetables consumption in the period 1999-2004 in Denmark - a quantile regression approach”.** People, food and environments : global trends and european strategies. Ghent, Belgien. 26.08. - 29.08. Congress of the European Association of Agricultural Economists: EAAE Congress Nr.:12.

Machado J. A. F. and Mata J. (2005). ”Counterfactual Decomposition of changes in wage distributions using quantile regression”. J. Appl. Econ. 20: 445-465

Koenker, R. and G. Bassett Jr.(1978). ”Regression Quantiles.” Econometrica 46:33-50.

Li Q. and Racine J (2003). ”Nonparametric estimation of distributions with categorical and continuous data”. Journal of Multivariate Analysis 86, 266–292.

Ouyang D, Li Q. , Racine J. (2006). ”Cross-validation and the estimation of probability distributions with categorical data”. Journal of Nonparametric Statistics, 18:1,69 — 100.

Stewart, H., N. Blisard, and D. Jolliffe. (2003). ”Do income constraints inhibit spending on fruits and vegetables among low-income households?” Journal of Agricultural and Ressource Economics 28: 465-80.