Household expenditures in Brazil: A non-parametric analysis using panel data[1]

Alexandre N. Almeida[†]
Carlos R. Azzoni[†]
André L. S. Chagas[†]
Francisco J. S. Tabosa[††]

Food expenditures, mainly in low-income families, have a substantial weight in the household budget, deserving special attention. The estimation of the Engel curves, relating the share of food in total household expenditure to income levels, is relevant for public policy design, both for poverty reduction (e.g. cash-transfer programs) and to agribusiness decisions, (e.g. definition of retail prices of food). In this paper we apply parametric and non-parametric methods to pseudo-panel data from four Brazilian household expenditure surveys, for six product groups (food, housing, transportation, education, health and clothing). The use of panel data allows to control for unobserved characteristics. In addition, non-parametric methods allow to verify if the function is adequately specified according to the nature of the data and its functional form. We estimate alternative models, including fixed effects and non-parametric panel data estimators. The results for the mean values of non-parametric estimates are not different compared to fully parametric estimations. However, in different percentiles, the non-parametric estimates of household expenditures and household components, the independent variables, produced different impacts on the share spent by households on each consumption group. We also concluded that non-parametric methods with controls for the unobserved characteristics of household heads in panel data produced similar results to those of non-parametric methods that follow a cross-sectional data structure. For the fully parametric models, very similar coefficients of the independent variables were also found between cross-sectional (pooled) and panel data (fixed effects) for six consumption goods analyzed.

Key words: household expenditures, non-parametric models, panel data.

[1] The first author gratefully thanks the São Paulo Research Foundation (FAPESP) (grant # 2016/17566-5) for the financial support of this research.
[†]University of Sao Paulo
[††]Federal University of Ceara

Household expenditures in Brazil: A non-parametric analysis using panel data

1 Introduction

Brazil has been living a very different moment from that the previous decades. Since the implementation of the Real Plan in 1994 and the stability of the economy, new generations of individuals have experienced a new economic situation in the country. This aspect is significantly important in the life of the families given that the sustained control of inflation allows the head of the family, for a longer period of time, plans his/her household budget in function of the income obtained, improving the well-being (Almeida and Azzoni, 2016).

Moreover, the opening up of the economy during the 1990s, which also allowed access to new varieties and brands of products, with potential effect on all the goods available to consumers, cannot be neglected. In addition, there was a substantial increase in agricultural supply, due to significant productivity gains, which substantially changed relative prices, mainly food (Azzoni et al., 2009). This impact is also distinguished not only from the degree of development and urbanization, but also according to the different cultural aspects between regions of Brazil. In practice, it is also observed that there is a great regional disparity in the costs of life among the Brazilian cities, which recommends a further concern with the regional variation of consumption patterns (Azzoni et al., 2003, Almeida and Azzoni, 2016).

Another relevant point is the differentiation of consumption patterns among different family structures. Camarano and Pasinato (2004) and Camarano et al., (2014) report that the rapid aging of the population has changed the traditional family typology. In general, families who have older people demand a lot more health care than the adult and young population. This burden, due to the precariousness of the public health services, can compromise the whole domestic budget due to the higher costs of private treatment (Almeida and Freitas, 2006), influencing thus the composition of the expenditure structure with other equally important goods and service for the family.

This scenario of potentially influential factors on changes in household consumption and in its well-being happened in a relatively short period (Almeida and Azzoni, 2016). By its intensity, however, it may already have exerted an important influence on the families. Therefore, the importance of analyzing the demand and patterns of household expenditures, especially of low income, is to assist in the diagnosis of their living conditions by identifying the collective and individual consumption profile. This picture also provides subsidies for the direction of social policies such as those of minimum income, food security, financial education, etc. (Almeida and Azzoni, 2016).

To reach this scope it is fundamental to make use of household budget surveys, since such surveys depict the observed economic behavior of the agents. Although they are not very frequent, due to its high execution cost, this type of research effectively approximates the standard of living information of the individuals, which allows to investigate aspects of the well-being and the cost of living of their families (Azzoni et al., 2003, Menezes et al., 2008, Almeida and Azzoni, 2016).

In households where the income is lower, the portion spent with food corresponds to the most significant part of the total expenditures, as already proposed by Engel's Law (Deaton, 1997). According to the Brazilian Institute of Geography and Statistics (IBGE), in 2009, the share of expenses with food items corresponded to 19.8% of all consumer expenditures (IBGE, 2010). Therefore, given the crucial importance of food to ensure a minimum welfare for families as well as its important participation in the expenses of the poorest families (which are more numerous), to focus attention on this item is indispensable, which is one of the main objectives of this paper. However, considering that the consumption structure also involves other items, equally relevant to the well-being of individuals living in large cities, it is not reasonable to restrict this analysis to food alone.

In Brazil, following the international literature, the number of studies on demand is quite extensive as certified by two volumes edited by Silveira et al. (2006) and Silveira et al. (2007). More recently, the literature also includes the works of: Resende and Oliveira (2008), Coelho et al. (2009), Pintos-Payeras (2009), Coelho et al. (2010), Hoffmann (2010), Silva et al., (2012), Pereda and Alves (2012), Barbosa et al. (2014), Silva and Coelho (2014), Oliveira and Hoffman (2015), Travassos and Coelho (2015), Almeida et al. (2016) and Ferreira e Coelho (2016).[2]

In general, most of the aforementioned works: 1) depart from an extensive use of IBGE family budget surveys; and 2) are based on the seminal works of Working (1943) and Leser (1963). That is, the specifications of the functional form of the demand function are based on linear regressions that associate the expenditure of a good (or group of goods) by the family with its logarithm of the total expenditure. Essentially, these models follow the structure:

$$w_{ji} = \alpha + \beta \ln(x_j) + e_{ji} \qquad (1)$$

[2] The works of Rossi (1982) and Medeiros (1978) were pioneers in Brazil in estimating Engel curves for the city of Rio de Janeiro.

where w_{ij} is share spent with good *i* by family *j*, ln x_j is the log of total expenditure of family j, while the unobserved component e_{ij} satisfies the following orthogonality condition $E(e_{ji}|x_j) = 0$ (Blundell and Duncan, 1998).

The most common form of estimation of a demand model was developed by Deaton and Muellbauer (1980), and became known as "Price Independent Generalized Logarithmic-PIGLOG" (Santana and Menezes, 2009).

Nonetheless, equation (1), by adding a price vector, can also be estimated simultaneously through a demand system known as the Almost Ideal Demand System (AIDS), which is currently very popular within the empirical analysis. A demand system with *n* equations (1) has the advantage of capturing consistently the mutual interdependence of a large number of goods, according to consumer choices, as well as budget constraints (De Janvry and Saudolet, 1995).

In addition, the AIDS model, not only it satisfies all the assumptions of consumer theory, but also presents as main advantage a linear format. This format facilitates the inclusion of constraints inherent to the demand models, such as additivity, homogeneity and symmetry (Barbosa et al., 2014).

However, over the years it was observed that the consumption behavior between poorer individuals (necessary items) and richer (luxury items) was not linear, and that the equation (1) suggested above would be no longer appropriate (Banks et al., 1997).

Thus, Banks et al., (1997) developed a demand system derived from the AIDS model, called the Quadratic Almost Ideal Demand System (QUAIDS). It adds in Eq. 1 a quadratic term of the logarithm of the total expenditure (income) x_j allowing to measure which goods consumed by the households would be driven by a path of expansion that is non-linear in income following the Engel curve (Pereda and Alves, 2012).

Similar to the AIDS model, QUAIDS has become quite popular because it is simple to estimate and also to check statistically the theoretical constraints embedded in consumer theory (Coelho and Aguiar, 2007). However, there is still no consensus on the true shape of the Engel curve that allows to capture the well-behaved consumer preferences according to the nature of the data (Blundell et al., 2003).

In general, economic theory does not provide much guidance on the choice of functional form for the estimation of any econometric model (Schmalensee and Stoker, 1999; Blundell and Duncan, 1998, Yatchew 2003, Henderson and Parmeter, 2015). It is also noted that few works go further by looking for alternative methods that can fit a functional form of demand, rather than simply based on the piglog specification, as suggested by Blundell and Duncan (1998), Bhalotra and Attfield (1998), Blundell et al. (1998) and Gong et al. (2005).

According to Blundell and Duncan (1998), there is an obsession with linearity coming from empirical economists. The use of non-parametric or semi-parametric models is an important advance, since it allows to verify if the specification of the function is adequate, or if there is a better way to represent the preferences of the empirical consumer theory (Blundell and Duncan, 1998; Blundell et al., 2003; Santana and Menezes, 2009).

To date, in Brazil, Santana and Menezes (2009) is the only work that estimates patterns of spending on education among races, doing it through semiparametric methods and taking into consideration the flexibilization of functional forms of the estimated econometric model.

Therefore, the objective of this study is to analyze how Brazilian families allocate their expenditures using the IBGE Family Budgets Surveys (POFs) of 1987/88, 1995/1996, 2002/2003, 2008/2009. Specifically, this research estimates the relationship between income and food expenditure, and also non-food items. We depart from the use of non-parametric econometric estimation techniques developed exclusively for panel data. Six groups of goods are considered in the analysis: food, housing, transportation, education, health and clothing. We start from a function in which the relation of interest between the share spent (w_{ij}) with good *i* by household *j* and the log of total family expenditure (or income) *j* (ln x_j) is given by:

$$w_{ji} = \varphi(\ln(x_j)) + e_{ji} \qquad (2)$$

where $\varphi(\cdot)$ is a regression function of the conditional mean of w_{ij} given x_j estimated non-parametrically.

The use of this condition as an empirical strategy is justified because non-parametric regression method has the great advantage of "let the data speak for themselves regarding the observed form of the regression curve" (p.10, Eubank, 1999). The use of classical econometrics imposes linearity on the parameters to be estimated (as is the case with the AIDS model and QUAIDS), and constitutes a limitation that may be quite restrictive, but it can be relaxed depending on the nature of the data through the use of non-parametric methods as expressed in Eq. (2) (Li and Racine, 2008, Henderson and Parmeter, 2015).

It is expected, through the non-parametric estimates obtained by the present study, to establish a comparison between these and the results of traditional econometric methods, which are widely used to subsidize most economic and policy recommendations. The major hypothesis that guides this research lies mainly in diagnosing the shape of the Engel curve non-parametrically for several consumption groups, aiming at providing more specific adjustments in public policies and new consumer markets. The analysis will take place from panel datasets, where individuals or "identical" households are tracked over time.

The remaining of this article is organized as follows: the next section briefly reviews the literature related to the scope of work, following by section 3 that describes the data used. Section 4 presents the econometric and non-parametric approaches used. Section 5 presents the results, followed by the concluding remarks.

2 Literature

The literature on studies related to the demand for goods is quite rich both in Brazil and abroad. However, analyzes that focus more specifically on methods of non-parametric econometrics aimed at determining the true shape of the Engel curve are still scarce.

Bhalotra and Attfield (1998) estimated semiparametrically the Engel curve for the rural areas in Pakistan, and found that the shape of the curve assumes a quadratic logarithmic specification, therefore, nonlinear. In this case, the use of the traditional Almost Ideal Demand System, Log-Translog and Linear Expenditure System would be inappropriate for an analysis of food demand.

A similar conclusion was found by Blundell and Duncan (1998) when analyzing UK food and beverages expenditures using household surveys from 1980 to 1982. Using nonparametric regression methodologies, and controlling for demographic components, authors have shown that the shape of the Engel curve for these two groups differ substantially between one and the other. While the share of food expenditures falls when total household spending increases, in the case of alcoholic beverages the Engel curve is shown as an inverted U (in relation to the portion spent with this item and total household expenditure and the logarithm of total expenditure).

Blundell et al. (1998), also using household surveys from the United Kingdom between 1980 and 1982, estimated non-parametrically Engel curves for the groups: food, household fuel, clothing, clothing, transport and other goods for households where one or two children coexist with other adult members. Among the most important results the linear specification piglog was rejected in favor of a non-parametrically estimated function for several consumption groups, among them: alcoholic beverages, clothing, transport and other goods.

Gong et al. (2005) used a partially linear model to identify differences in spending on food, education, and alcoholic beverages in China's rural households. It was found that the structure of consumption differs according to the gender of the children. The hypothesis of a linear (inverse) relationship between the food item and the total household expenditure was rejected, indicating that the use of linear models would be inappropriate as mentioned above.

Santana and Menezes (2009) estimated the Engel curve for Brazil, focusing on education spending. They adjusted a semi-parametric model to the POF 2002-2003 dataset, using as control variables the characteristics of the head of the family, race, household composition and location. However, the authors ignored any analysis concerning the non-parametric (total expenditure) part of the variable dependent on interest, focusing only on the parametrized variables.

The few mentioned studies provide important subsidies to conclude that the flexibilization of the function for estimation of the Engel curve must always be revisited. Another limitation present in all the mentioned studies is that their analyzes are based on cross-sectional data. This obstacle implies that the analyst is allowed to make statistical inferences only on the observed determinants that explain the expenditure, while other equally important characteristics concerning the decision-making process by families, but which are not observed (e.g., emotional intelligence of the heads), are completely ignored (Cameron and Trivedi, 2005; Baltagi, 2008).

In this context, the use of panel data overlaps the use of cross-sectional data with success, since panel data have as main analytical advantage reduce the bias of omitted variables, by allowing other characteristics that are invariant in time and that are not observed (referring to the unit of analysis) can be controlled (Cameron and Trivedi, 2005, Baltagi, 2008).

3 Data

The database used in this study is the Household Budget Surveys (POF), carried out by the Brazilian Institute of Geography and Statistics (IBGE) between March 1987 and February 1988, between October 1995 and September 1996, between July 2002 to June 2003 and again between May 2008 and May 2009. The POF diagnoses the quality of life of Brazilian families based mainly on their household budget (IBGE, 2004).[3]

Given the high cost of its execution, there are only five POF's available until now. The first, at the national level, was the National Family Expenditure Survey (ENDEF) 1974-1975.

[3] There is a great expectation of the new POF 2017/2018 by society and academic community. According to a press release published by IBGE, the field collection began in June 2017 and expected to finish in May 2018. The results are expected to be out in 2019.

The two later ones were carried out in 1987-1988 and in 1995-1996 (named as POFs), and covered nine metropolitan regions, Goiânia and the Federal District. The last two POF editions (2002/03 and 2008/09) available, in addition to being carried out throughout Brazil, also investigated anthropometric measurements of individuals.

The POFs of 87/88, 95/96, 02/03 and 08/09 interviewed 13,611, 16,013, 48,470 and 55,970 household, respectively. For their execution, sample expansion factors were constructed from the results of the Demographic Censuses (IBGE, 2004, IBGE 2010). In order to maintain consistency between the first two POFs and the last two, only observations of the residents in the 9 metropolitan regions (Belém, Recife, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba and Porto Alegre) plus those of the Federal District and of Goiânia are used in the present analysis. In addition, all items related to expenses and income are deflated by the IPCA/IBGE with the base for June 2017.

The data of the POFs are very important to evaluate the cost of living of the people, however, its execution is costly by the degree of detail of the information related to the expenses and the income of the families investigated. In addition, the data have a major limitation for not being structured in a panel form, that is, same individuals and households are not followed over time. This feature decreases the accuracy of econometric estimates (Cameron and Trivedi, 2005). A good alternative, then, would be to make use of data from repeated cross-sections as an approximation of longitudinal data, a technique that became known as "pseudo-panel" (Deaton, 1997).

Browning et al. (1985) and Deaton (1997) were the first to use pseudo-panels to analyze labor supply in the United Kingdom and Taiwan, respectively (Cameron and Trivedi, 2005).

The technique to construct a pseudo-panel is quite simple, and consists of converting a cross-sectional database from household or individual units in a determining period of time *t* to the level of cohorts *c*. A cohort can be interpreted as a group of individuals who share - at a given time - a common experience, defined from their similar average characteristics observed over time, such as: age, sex, income, geographic location, parents' schooling etc. (Deaton, 1997). Generally, for assembling a cohort *c*, if "t" is the moment of occurrence, for example the year of the survey, and "a" is the age of the individual at this time (time *t*), then one can say that this applies to the cohort born in year "t-a", as well as observations for ages a-1 at time t-1 and so on (Ryder, 1965).

In Appendix A we illustrate all the steps of assembling the cohorts using the STATA software and POF datasets. Only the resident heads are selected in the nine metropolitan regions, Goiânia and DF of the four POFs available, totaling 21,321 individuals between 1987 and 2009. The cohorts for these individuals are then constructed based on their age, sex, and year of the survey, with an interval of five years (arbitrarily defined) between the cohorts. A final unbalanced pseudo-panel dataset covering the period from 1988 and 2009 were then assembled totaling 3,961 cohorts (observations) in total.

During any econometric estimation that makes use of household surveys, one obstacle must be carefully investigated. It refers to the number of zeros (corner solution) in the households that declared that they did not consume the good during the week of the survey. Although it does not mean that the good has not been consumed ever, rather it may exist in storage in the household. Or, one would say that the good may simply not be part of the consumer basket of the household. Different families may have the most distinct consumption habits as well as income, and etc.

Two-stage Tobit or Heckman procedures are most commonly used in these cases and are easily implemented when data are censored or truncated (Coelho et al., 2009). Another solution would be to aggregate the families ranked by income classes of each cohort, reducing the number of observations with zero values. This practice is quite common in demand studies (Menezes et al., 2008).

However, both procedures described above were not necessary because during the construction of the cohorts, followed by the assembly of the pseudo-panel, average values were generated for: number of household components, household income and expenditure of each cohort. As a result, few observations had zero values and were therefore discarded.

4 Econometric Strategy

The application of non-parametric models is still limited, but has been increasing in many fields of economic research (Henderson and Parmeter, 2015). Henderson and Parmeter (2015) also suggest that, for policy analysis purposes, a fully parametric model, as put forward by Eq. (1) during the introduction, is always preferable to a non-parametric model (Eq. 2), since it is easy to interpret.

In the case of the present study, the main interest is to investigate, through non-parametric methods, how robust is the impact of family expenditure on the portion spent (in percentage terms) on various goods and services relative to the total family budget. In addition, assuming that there is an unobserved heterogeneity among the different heads' cohorts, which affect their consumption decisions, the use of panel data allows us to control characteristics that are typically considered as unobserved and time invariant (e.g., emotional intelligence).

This heterogeneity that exists between the units of analysis, when incorporated into the model, besides resulting in unbiased estimates, brings to the researcher more information, more variability, less collinearity, more degrees of freedom and, consequently, more efficient results (Baltagi, 2008). However, one of the major limitations in working with panel data is the attrition of the samples, which indicates that it is not always easy to follow the same individuals over time. The use of the pseudo-panel is able to fill this gap, since it allows to follow pseudo-individuals during the assembly of the different cohorts (Deaton, 1997).

The main reason for using the non-parametric method is to avoid constraints or specification errors, which may result in erroneous specific functional forms, producing biased estimates (Yatchew, 2003; Li and Racine, 2007; Henderson and Parmeter, 2015). In this case, consider first the following fully parametric specification of the cohort *c* for the period t = 1988, 1996, 2003 and 2009:

$$\bar{Y}_{c,t,i} = \bar{L}_{c,t}\beta + \bar{X}'_{c,t}\theta + \bar{C}_c + \bar{\varepsilon}_{c,t} \qquad (3)$$

where $\bar{Y}_{c,t}$ represents the average of the portion of expenses with a given good i for the cohort c; $\bar{L}_{c,t}$ is the average monthly total expenditure of each cohort; $\bar{X}_{c,t}$ is a vector with mean values belonging to each cohort from household and individual characteristics; and the vector $\bar{C}_c$ represents the mean of unobserved individual components for each cohort that are assumed to be constant over time (therefore, without the subscript *t*), but possibly correlated with $\bar{L}_{c,t}$ e $\bar{X}_{c,t}$[4]. The term error is represented by $\bar{\varepsilon}_{c,t}$, while β e θ are parameters to be estimated. The model proposed in (3) is known as the fixed-effects model (Cameron and Trivedi, 2005; Baltagi, 2008).

Using the technique known as "within transformation" is to eliminate the invariant component in time $\bar{C}_c$ (fixed effects) and produces consistent and unbiased estimates. Therefore, Eq. (3) becomes:

$$\ddot{Y}_{c,t} = \ddot{L}_{c,t}\beta + \ddot{X}'_{c,t}\theta + \ddot{\varepsilon}_{c,t} \qquad (4)$$

where (4) is estimated using the traditional OLS method.

[4] At first, no distribution of the C_c component (fixed effects) is assumed. However, a Hausman type test in order to verify the presence of random effects versus fixed effects must be performed. Another point is that when working with cohorts, measurement errors are incorporated because the cohort means are estimates of the population mean values, however, if the number of observations per cohort is high this error can be ignored (Verbeek and Nijman, 1992; and Trivedi, 2005).

Next, consider the same specification of Eq. (4), but now without the presence of linear parameters (β and θ) in the model:

$$\ddot{Y}_{c,t} = \varphi(\ddot{L}_{c,t}, \ddot{X}_{c,t}) + \ddot{\varepsilon}_{c,t} \qquad (5)$$

where $\varphi(\cdot)$ is an unknown function of total expenditure $\ddot{L}_{c,t}$, and the vector $\ddot{X}_{c,t}$ which follows the same specification of Eq. (4).

One of the main limitations of estimating Eq. (5) using nonparametric approaches is the "curse of dimensionality" (Li and Racine, 2007). In other words, if the number of independent variables is high, the convergence rate of the nonparametric estimator to its true value ($\hat{\varphi} \underset{p}{\rightarrow} \varphi$) decreases, thus producing inconsistent estimates (Li and Racine, 2007; Henderson and Parmeter, 2015). It is then recommended for applied works to use at most three variables in the non-parametric part of Eq. 5 (Yatchew, 2003). Moreover, non-parametric estimates based mainly on kernel functions, as is in the present case, are quite demanding from the computational point of view, depending in turn on the number of observations and the number of interactions during the non-parametric component optimization (Härdle, 1990).

Thus, to avoid the curse of dimensionality, facilitates estimation, and be still able to control observed and not observed individual characteristics on determinants that influence the dependent variable of interest $\ddot{Y}_{c,t}$, one can choose to:

1) A semi-parametric version as follows:

$$\ddot{Y}_{c,t} = \varphi(\ddot{L}_{c,t}) + \ddot{X}'_{c,t}\theta + \ddot{\varepsilon}_{c,t}; \qquad (6)$$

2) Or by a more parsimonious specification of Eq. (5), that is, with a reduced number of variables in the vector $\ddot{X}_{c,t}$.

Thus, for the present work we depart from a totally non-parametric and a more parsimonious version of eq. (5), which can be expressed as follows:

$$\ddot{Y}_{c,t} = \varphi(\ddot{L}_{c,t}, \ddot{M}_{c,t}) + \ddot{\varepsilon}_{c,t} \qquad (7)$$

where: $\ddot{L}_{c,t}$ is the average monthly total expenditure of each cohort; and $\ddot{M}_{c,t}$ is the average number of people in the household within each cohort. The goal is to estimate $\varphi(\cdot)$ and a slope (gradients) of $\varphi(\cdot)$ which can be derived as $\beta(\cdot) = \partial\varphi(\cdot)/\partial(\cdot)$. Henderson et al. (2008) developed an iterative non-parametric kernel estimator to estimate $\varphi(\cdot)$.

In order to verify the performance of non-parametric methods, developed specifically for panel data, it is interesting to compare these estimates following the same functional form of the traditional fully parametric methods used by most empirical studies. In addition, it is also common to report estimates of panel data as if they were a single cross-section (i.e., following a pooled data structure). However, by doing that, the possibility of controlling for fixed effects, that measure the unobserved heterogeneity of the cohorts, is therefore excluded (Cameron and Trivedi, 2005; Baltagi, 2008), although there are indications that this distinction between cross-sectional and panel data structures has little effect on the final results when the number of observations is high (Baltagi, 2008). Figure 1 shows the various parametric and non-parametric specifications proposed by this paper.

Parametric models:	*Non-parametric models:*
Pooled – I $Y_c = L_c\beta + M_c\ \gamma + \varepsilon_c$ (8) Pooled – II $Y_c = L_c\beta + M_c'\gamma + D_c\ region + D_c\ year + \varepsilon_c$ (9) Panel – I $\ddot{Y}_{c,t} = \ddot{L}_{c,t}\beta + \ddot{X}'_{c,t}\gamma + \ddot{\varepsilon}_{c,t}$ (10)	Pooled – I $Y_c = \varphi(L_c, M_c\ ) + \varepsilon_c$ (11) Pooled – II $Y_c = \varphi(L_c, M_c\ , region_c, year_c) + \varepsilon_c$ (12) Panel – I $\ddot{Y}_{c,t} = \varphi(\ddot{L}_{c,t}, \ddot{M}_{c,t}) + \ddot{\varepsilon}_{c,t}$ (13) Panel – II $Y_{c,t} = \varphi(L_{c,t}, M_{c,t}, idch_{c,t}) + \ddot{\varepsilon}_{c,t}$ (14)

where for each cohort we have:
Y_c = Ln of share spent on the good;
L_c = Ln of the total monthly expenditure used by total income as instrument;
M_c = Ln of household size;
$region$ = dummies variables where households are located;
$year$ = dummies variable for the year of survey;
idch = id of cohort;
$\beta, \gamma, D_c^R, D_c^A$ are parameters to be estimated parametrically; and
$\varphi\ (\cdot)$ is a function to be estimated non-parametrically.

Figure 1 - parametric and non-parametric models adopted.

The specifications ranging from 8 to 10 are estimated using the traditional method of ordinary least squares (OLS) with instrumental variable, aiming at correcting biases of estimates from omitted variables between error term, total expenditure and dependent variable (Cameron and Trivedi, 2005). This has been addressed by Blundell and Duncan (1998), Bhalotra and Artfield (1998) and Gong, van Soest and Zhang (2005) who suggest using total household income as instrument for total household expenditure. Anticipating part of the results, in fact, the Wu-Durbin-Hausman test, which is easy to implement in STATA, and applied to the models (8) and (9), with the pooled data, the hypothesis that total expenditure is exogenous is rejected at the 1% level of significance for all consumption groups with the exception of the clothing group.

Therefore, in both parametric and non-parametric specifications, the total monthly expenditure, our main variable of interest, is included in the models using the monthly income as an instrument following the steps of the IV-procedure during the estimation (Cameron and Trivedi, 2009).

In summary, the rationale behind all these different models from 8 to 14 is to allow the analyst to check the sensitivity of the impact of the main variable of interest at work (i.e., total expenditure) by comparing parametric and non-parametric approaches. In other words, the results observed by equation (8) should be compared with the results of equation (11), the results of equation (9) with equation (12), and, finally, equation (10) with those results coming from equations (13) and (14).

All the theoretical properties for each of these four estimators are presented in Henderson and Parmeter (2015), and are omitted here. Henderson and Parmeter (2015) also provide programming codes for software R to replicate all practical examples of their book, being these codes easily adaptable by users for their respective research problems.

Before entering into the analysis of results, it should be noted that the starting point for all these non-parametric estimators was developed by Nadaraya (1964) and Watson (1964). The Nadaraya-Watson estimator or "Local Constant Linear Square-LCLS" based on the Eq. (2) of the introduction – $w_{ji} = \varphi(\ln(x_j)) + e_{ji}$ – can be expressed by:

$$\hat{\varphi}\left(x_j\right) = \frac{\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x_n - x_j}{h}\right) w_{ji}}{\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x_n - x_j}{h}\right)} \tag{15}$$

where $k(\cdot)$ is a kernel function that satisfies all the statistical properties of: symmetry, integration for 1 and continuously differentiable (Blundell and Ducan 1998, Henderson and Parmeter, 2016). The specification of the most popular kernel functions - uniform, Gaussian and epanechinikov - are listed below.

Uniform $$k(u) = \frac{1}{2}\mathbb{I}(|u| \leq 1) \tag{16}$$

Gaussian $$k(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{17}$$

Epanechinikov $$k(u) = \frac{3}{4}(1 - u^2)\mathbb{I}(|u| \leq 1) \tag{18}$$

For any non-parametric estimation of the Kernel type using the expression (15), the choice of the bandwidth parameter h is more important than the choice of the kernel function $k(\cdot)$, since h regulates the trade-off between variance and bias of estimates (Li and Racine, 2007; Henderson and Parmeter, 2016). There are several methods that take into account the nature of the data to calculate this bandwidth, among them are: the Silverman rule-of-thumb (1986), least squares cross-validation, and the maximum likelihood of cross-validation (Henderson and Parmeter, 2015).

5 Results

As mentioned earlier, 3,961 cohorts are used to generate an unbalanced pseudo-panel between 1988 and 2009 in total. Table 1 shows the mean and standard deviation of each variable of interest. In summary, it can be observed that, in the 9 metropolitan areas (Belém, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba and Porto Alegre), Federal District and Goiânia, that food consumed within and out of household represent on average about 30% of total monthly expenditure, followed by expenditures on housing (21%), other goods (15%), transportation (12%), clothing (9%), health (8%) and education (4%). The goods and services that make up each group are presented in the Appendix B. The group of other goods and services (e.g., hygiene, leisure, personal expenses, other household expenses, other real estate, other expenses, charges, communication and smoking) is not considered in the analysis.

The average number of members per household within each cohort is around 3.69, while the average total expenditure is around R$ 2,665. The average total income is around R$ 4,130. About 60% of families lived between 1998 and 2009 in the three major capitals of the northeast and southeast regions, and 29% of the sample collected came from the 1995/1996 family budget survey.

Table 1 – Mean and standard deviation of variables.

Variables	Mean	S.D.
Monthly household share spent on food (%)	0,31	0,15
Monthly household share spent on housing (%)	0,21	0,09
Monthly household share spent on transportation (%)	0,12	0,06
Monthly household share spent on clothing (%)	0,09	0,05
Monthly household share spent on health (%)	0,08	0,06
Monthly household share spent on education (%)	0,04	0,03
Monthly household share spent on other goods (%)	0,15	0,10
Number of household components	3,69	1.09
Monthly household expenditures in R$*	2.665,4	2.939,6
Monthly household income in R$*	4.130,3	5.553,4
=1 if household is located in the North region	0,10	0,30
=1 if household is located in the Northeast region	0,31	0,46
=1 if household is located in the Southwest region	0,28	0,45
=1 if household is located in the South region	0,16	0,37
=1 if household is located in the Center region	0,15	0,35
= 1 if observation is from 1988	0,23	0,42
= 1 if observation is from 1996	0,29	0,45
= 1 if observation is from 2003	0,24	0,43
= 1 if observation is from 2009	0,23	0,42

Note: * (CPI 2017/Jun = 100)

It is important to point out that within the literature applied to non-parametric models, it is also very common to show the distribution of interest estimates, in the form of tables, at the percentiles level, or through two-dimensional graphs (Henderson and Parmeter, 2015).

As an illustration, the non-parametric (Gaussian kernel) regressions are plotted in Figure 2 from bandwidths estimated by the least squares cross-validation method (LSCV) (Henderson and Parmeter, 2015). These are based on the Local Constant Linear Squares-LCLS method of Eq. (11) of the log of the share of each item spent against the logarithm of total expenditure with the logarithm of the number of components per household fixed to its median value, therefore, constant.

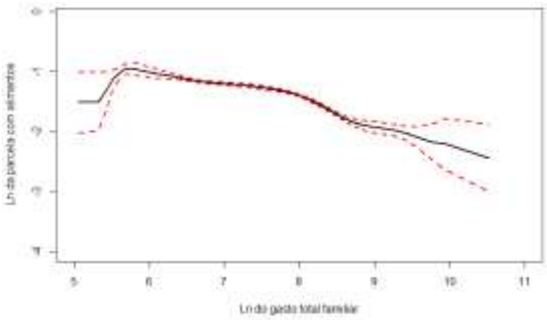
For each figure, the 95% confidence intervals, with standards errors estimated by bootstrap methods are also presented, following the outlines proposed by Henderson and Parmeter (2015, p.139). The use of bootstrap has become very popular in recent years by producing estimates of robust standard errors for heteroscedasticity (Cameron and Trivedi, 2005).

In summary, the non-parametric estimation figures appear to converge to a good approximation of the linear piglog specification. A linear behavior is observed for the curve estimated mainly for the items: food, housing and transportation. Together these corresponds on average to 64% of the total family budget.

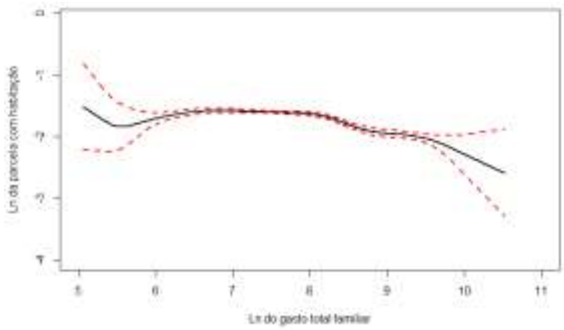
It is also interesting to note the format of an inverted N for the education group. For example, in taking the antilog of ln from total family expenditure (not shown), the share of the expenditures on education is reduced up to R$ 403 of total monthly expenditures, followed by an increase up to R$ 13,500 of total monthly expenditures, and followed by a slight decrease.

It should be also noted that health spending appears to decline, even at rates lower than the increase in total spending. In addition, from this level of income, the "health" good seems to assume a behavior of "luxury goods".

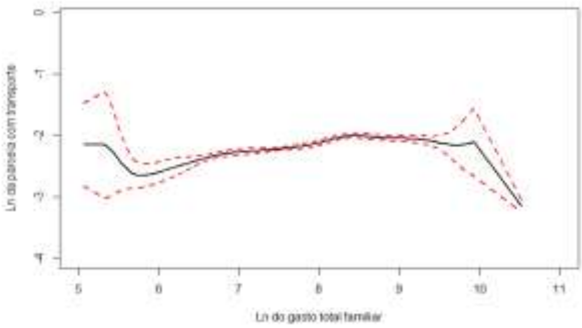
The Engel curve, which represents expenditure on clothing, which includes clothing and footwear, presented a similar format to that of the education group, that is, the N-inverted format, but with similarities also to the group food and housing.



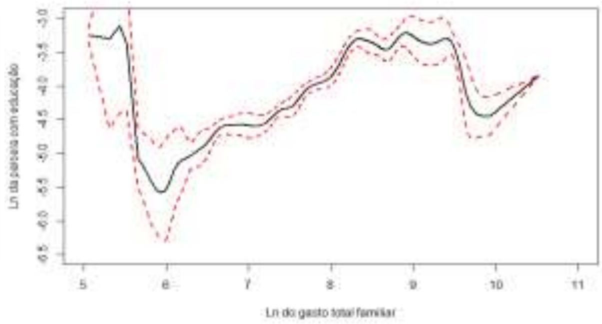
Engel Curve - Food



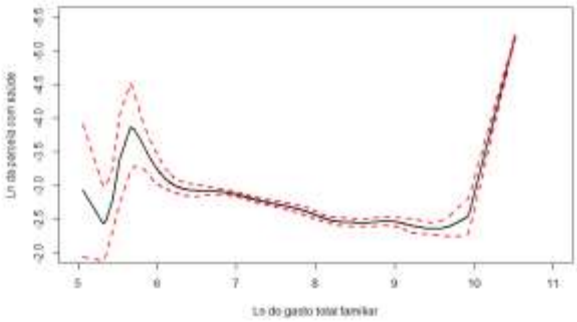
Engel curve - Housing



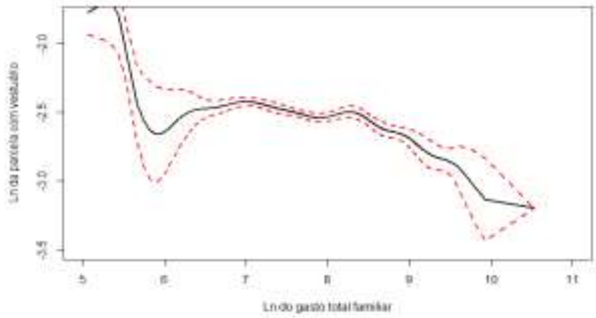
Engel Curve - Transportation



Engel Curve - Education



Engel Curve - Health



Engel Curve - Clothing

Figure 2 - Non-parametric (kernel) regressions for selected consumption groups. Dashed lines represent 95% of the confidence interval with standard errors estimated using bootstrap methods from 200 replications.

Tables 2 to 13 report the parametric and non-parametric results for the six consumption groups in the form of elasticities. More specifically, in each even-numbered table (Tables 2 to

10) the results of the three fully parametric models are shown, while their non-parametric counterfactual results are shown in odd-numbered tables, Table 3 to 13.

As already discussed, parametric and non-parametric estimates are considered with the data pooled, and the remaining in their panel data format given the objective of the latter is to increase the accuracy of the estimates by controlling unobserved determinants of each cohort (Cameron and Trivedi, 2005).

In addition to Henderson et al. (2008) non-parametric estimator as discussed in the methodology, results from three more non-parametric estimators are also shown to check the robustness of non-parametric results. The first is the Local Constant Linear Square-LCLS I specifically designed for cross-sectional data, followed by an adapted Local Linear Square-LCLS II estimator, when part of the non-parametric function is composed of continuous variables (total expenditure and number of people) and discrete variables (region and year). Finally, we introduce a non-parametric fixed-effects estimator developed by Li and Racine (2004) in which the identification of each cohort is repeated over the years plus the year of the survey entering as discrete variables in the model. This last one is an extension of the LCLS estimator called Local Linear Least Squares-LLLS (Henderson and Parmeter, 2015).

In the non-parametric results, the standard errors are estimated by the bootstrap method using 200 replications. Partial effects of each regressor estimated non-parametrically in the P25, P50 and P75 percentiles and in the percentile mean are also reported.

In general, it can be observed that the results of the mean results of the non-parametric estimates did not appear to be substantially different from the results obtained from the fully parametric piglog models. In summary, it can be suggested that: 1) based on the main assumption of the non-parametric regression of imposing no constraint on the estimated function relative to the error term and linearity of the parameters, traditional econometrics provided a good approximation regarding the impact of total expenditure and the number of members on the expenditure shares of the consumption groups analyzed; and 2) therefore, the use of more advanced econometric such as nonparametric regression techniques would not be in fact needed.

However, by taking a look closer at the different percentiles of the non-parametric results, interesting and significant differences between the two methods are observed, to be described below.

Starting with Table 2 which shows the results of the elasticities for the food group coming from the fully parametric results, while Table 7 show the elasticities estimated non-parametrically by the different methods for the same consumption group.

When comparing the two methods, considering the pooled data, it can be observed that elasticity of the total expenditure in the first percentile (P25) by the non-parametric method LCLS I produced a partial effect (-0.34), very similar to the result (-0.34) of the fully parametric model estimated by OLS I. However, it is observed that this partial effect is increasing (less and less negative) from the P50 percentiles (elasticity of -0.19) and P75 (elasticity of -0.12), as proposed by Engel's law when the total expenditure increases over the portion spent with food. This same tendency of the estimates is observed in the percentiles of the other non-parametric estimators.

As already mentioned, the use of fixed effects estimators is to allow the control of unobserved effects among the cohorts, increasing the precision of the estimates. As they were constructed from the heads, it is assumed, for example, that there is a congenital component of them that could be significantly influencing the consumption decision of the whole family's food. Econometrically, if this hypothesis is true, the estimates with pooled data are expected to be not similar to the panel data estimates. More specifically, in the case of the OLS method (OLS FE), the panel fixed effects model produced an estimate of the total expenditure on the portion spent with food slightly lower (-0.37) than that of its counterfactuals with pooled data, with elasticities of -0.34 (OLS I) and -0.35 (OLS II). Therefore, these effects are too small and it is not possible to conclude that by controlling the in-born characteristics of heads, the impact of total expenditure on the portion spent on food is statistically different than not controlling them. By the same token mixed and inconclusive results are also found in the non-parametric estimates when a panel data structure is taken into account compared to pooled dataset.

Similar to the parametric results, the non-parametric estimates also corroborate the hypothesis that increasing the number of household members increases the amount spent with food, as expected.

Housing expenditures also showed similar consumption characteristics as those observed for food items in relation to total expenditures. This is true for all specifications and methods used. Among the parametric results described in Table 4, the elasticities observed range from -0.09 to -0.13 and are relatively lower than the mean of the partial effects of the non-parametric methods in Table 5 with values ranging from -0.03 to -0.09; although not all are statistically significant at the levels of 1%, 5% or 10%.

In all the results of this group, both parametric and non-parametric, the negative values of the elasticities for the number of people at home calls some attention to typical characteristics of economy of scale of household consumption behavior hypothesis. Clearly, among non-parametric methods, it is observed that these elasticities are more expressive in the lower percentiles (25º) (between -0.51 and -0.77) than in the medians (50º) and superior (75º) (between -0.05 and -0.17). The impact of this variable is the parametric methods is -0.18 (OLS I) and -0.14 (OLS-FE). This result seems to be relevant in suggesting that parametric models - with imposed linearity constraints - are overestimating the impact of this variable on the consumption of housing goods and services.

Household expenditure on transportation accounts for 12% of the budget in the metropolitan regions. The impact of total expenditure on the portion spent with this group is between 0.16 and 0.19 in the parametric estimates (Table 6), and between 0.16 and 0.31 in the non-parametric estimates (Table 7).

It is also highlighted in this consumption group that the elasticities of the parametric models orbit between 0.017 and 0.083 for the number of people in the household. However, negative and statistically significant partial effects are observed in the lower percentiles of P25 (elasticities between -0.35 and -0.16) but not in the upper percentiles of P75 (between 0.28 and 0.31) of non-parametric methods. This important result, similar to the housing group, also brings evidence that any suggestion in terms of mobility policies based only on the results of the parametric models could be poorly dimensioned in relation to households with less residents.

Another example is to inspect the impact of total family expenditure on the portion of the family budget spent on education. If the elasticities of expenditure between the parametric models (Table 8) vary little, from 0.61 to 0.66, in the non-parametric models (Table 9) few results are inconclusive, based on the level of significance. The elasticities of 1.03, 1.09 and 1.12 in the highest income percentile (P75), while of 0.59, 0.69 and 0.80 in the median percentile (P50) are the only ones statistically significant. For the variable number of household members, larger and statistically significant elasticities are observed only in the upper percentile P75 (values ranging from 1.77 to 1.82) and P50 (0.46), while the parametric counterparts, the results are: 0.42; 0.28; 0.36, for OLS I, OLS II and OLS-FE, respectively.

In the parametric results (Table 10) of the health group, the pooled and panel data elasticities of total expenditure produced very close estimates to each other orbiting around the value of 0.27. In non-parametric models (Table 11), only the estimators that consider the nature of the data as pooled produced statistically significant estimates, with elasticities varying in the mean of the percentiles between 0.22 (LCLS I) and 0.33 (LCLS II). In relation to the household members for the health expenditure, the elasticities of the parametric models ranged from -0.15 to -0.27; while in non-parametric methods, a range with negative values between -0.19 and -0.35 was found at P25 percentiles and a range with positive values between 0.32 and 0.48 was found at the P75 percentiles.

In the clothing group, consumption behaviors similar to those of the food and housing group are also observed in relation to total expenditure, following Engel's law. In the parametric methods, the negative elasticities are around -0.08 and -0.04 (Table 12). In non-parametric methods (Table 13), negative and statistically significant elasticities of -0.17 and 0.08 are observed in the P25 lower percentiles of the LCLS I and HCL methods, respectively; and positive and statistically significant only in the LCLS II method with a value of 0.27 in P75 is also observed.

Table 2 - Parametric regressions for food. Dependent variable: Ln of the family budget share spent with food.

	Pooled OLS I	Pooled OLS II	Fixed Effects OLS-FE
Ln (monthly expenditure)	-0,348***	-0,356***	-0,375***
	(0,012)	(0,010)	(0,008)
Ln (household size)	0,464***	0,459***	0,503***
	(0,032)	(0,028)	(0,019)
North region		0,086***	
		(0,017)	
Southeast region		-0,092***	
		(0,014)	
South region		-0,062***	
		(0,018)	
Midwest region		-0,144***	
		(0,018)	
1995		-0,640***	
		(0,012)	
2003		-0,819***	
		(0,015)	
2009		-0,847***	
		(0,015)	
Constant	0,701***	1,405***	0,857***
	(0,084)	(0,065)	(0,056)
R2	0,25	0,62	0,25
Durbin-Wu-Hausman	2,67*	9,09***	
Hausman			12.54***

Notes: 1) Standard errors are below the coefficients in parentheses.
2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3 - Summary of elasticities for several non-parametric estimators with estimated bandwidths via Least-Squares Cross Validation (LSCV) and Gaussian kernel function for food. Dependent variable: Ln of the family budget share spent with food.

Pooled					
LCLS I	Mean	Q1	Q2	Q3	R2
Ln (monthly expenditure)	-0,244	-0,340***	-0,193***	-0,124***	0,25
	(0,204)	(0,025)	(0,056)	(0,035)	
Ln (household size)	0,258***	0,086	0,321***	0,436***	
	(0,062)	(0,064)	(0,060)	(0,046)	
LCLS II*					
Ln (monthly expenditure)	-0,388***	-0,475***	-0,332***	-0,251***	0,62
	(0,058)	(0,132)	(0,034)	(0,024)	
Ln (household size)	0,216***	0,122	(0,204)	0,303***	
	(0,069)	(0,092)	0,365	(0,034)	
Fixed Effects HCL (2008)					
Ln (monthly expenditure)	-0,286***	-0,387***	-0,259***	-0,187***	0,24
	(0,018)	(0,020)	(0,017)	(0,017)	
Ln (household size)	0,347***	0,176***	0,364***	0,537***	
	(0,027)	(0,039)	(0,044)	(0,019)	
Li e Racine (2004)**					
Ln (monthly expenditure)	-0,372***	-0,462***	-0,314*	-0,203***	0,70
	(0,132)	(0,065)	(0,034)	(0,106)	
Ln (household size)	0,294***	0,052	0,277	0,520**	
	(0,076)	(0,354)	(0,190)	(0,255)	

Notes: 1) Table shows the elasticities for 25º(Q1), 50º(Q2), e 75º(Q3) percentiles along with standard errors estimated by the bootstrap method from 200 replications.
2) * Regressions contain dummies for regions and years.
3) ** Regressions contain dummy for cohorts.
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 - Parametric regressions for housing. Dependent variable: Ln of the family budget share spent with housing.

	Pooled OLS I	Pooled OLS II	Fixed Effects OLS-FE
Ln (monthly expenditure)	$-0{,}092^{***}$	$-0{,}132^{***}$	$-0{,}092^{***}$
	(0,012)	(0,012)	(0,010)
Ln (household size)	$-0{,}180^{***}$	-0,033	$-0{,}141^{***}$
	(0,030)	(0,030)	(0,024)
North region		$-0{,}078^{***}$	
		(0,024)	
Southeast region		$0{,}293^{***}$	
		(0,016)	
South region		$0{,}112^{***}$	
		(0,020)	
Midwest region		$0{,}293^{***}$	
		(0,021)	
1995		$-0{,}131^{***}$	
		(0,018)	
2003		$0{,}035^{**}$	
		(0,018)	
2009		$0{,}083^{***}$	
		(0,018)	
Constant	$-0{,}760^{***}$	$-0{,}768^{***}$	$-0{,}809^{***}$
	(0,085)	(0,083)	(0,068)
R2	0,089	0,228	0,09
Durbin-Wu-Hausman	$42{,}90^{***}$	$27{,}41^{***}$	
Hausman			$11{,}79^{***}$

Notes: 1) Standard errors are below the coefficients in parentheses.
2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 - Summary of elasticities for several non-parametric estimators with estimated bandwidths via Least-Squares Cross Validation (LSCV) and Gaussian kernel function for housing. Dependent variable: Ln of the family budget share spent with housing.

Pooled					
LCLS I	Mean	Q1	Q2	Q3	R2
Ln (monthly expenditure)	$-0{,}044^{*}$	-0,127	-0,029	0,024	0,14
	(0,026)	(0,126)	(0,020)	(0,027)	
Ln (household size)	$-0{,}274^{**}$	-0,531	$-0{,}321^{**}$	-0,047	
	(0,128)	(0,328)	(0,132)	(0,139)	
LCLS II *					
Ln (monthly expenditure)	-0,094	-0,296	-0,078	0,136	0,34
	(0,217)	(0,163)	(0,179)	(0,086)	
Ln (household size)	$-0{,}318^{**}$	$-0{,}537^{***}$	$-0{,}378^{***}$	-0,092	
	(0,150)	(0,089)	(0,085)	(0,169)	
Fixed Effects HCL (2008)					
Ln (monthly expenditure)	$-0{,}054^{***}$	$-0{,}127^{***}$	$-0{,}050^{***}$	0,011	0,13
	(0,017)	(0,023)	(0,013)	(0,015)	
Ln (household size)	$-0{,}316^{*}$	$-0{,}514^{***}$	$-0{,}293^{***}$	$-0{,}171^{**}$	
	(0,170)	(0,067)	(0,087)	(0,074)	
Li e Racine (2004)**					
Ln (monthly expenditure)	-0,038	-0,221	-0,033	0,135	0,40
	(0,256)	(0,301)	(0,136)	(0,113)	
Ln (household size)	$-0{,}356^{*}$	-0,775	-0,443	-0,055	
	(0,207)	(0,563)	(0,404)	(0,312)	

Notes: 1) Table shows the elasticities for 25º(Q1), 50º(Q2), e 75º(Q3) percentiles along with standard errors estimated by the bootstrap method from 200 replications.
2) * Regressions contain dummies for regions and years.
3) ** Regressions contain dummy for cohorts.
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 - Parametric regressions for transportation. Dependent variable: Ln of the family budget share spent with transportation.

	Pooled OLS I	Pooled OLS II	Fixed Effects OLS-FE
Ln (monthly expenditure)	0,183***	0,168***	0,195***
	(0,016)	(0,016)	(0,014)
Ln (household size)	0,063	0,083**	0,017
	(0,041)	(0,042)	(0,034)
North region		-0,044	
		(0,034)	
Southeast region		0,181***	
		(0,025)	
South region		0,058*	
		(0,032)	
Midwest region		0,223***	
		(0,029)	
1995		0,592***	
		(0,025)	
2003		0,450***	
		(0,030)	
2009		0,658***	
		(0,028)	
Constant	-3,730***	-4,172***	-3,761***
	(0,109)	(0,108)	(0,096)
R2	0,028	0,199	0,031
Durbin-Wu-Hausman	41,56***	28,57***	
Hausman			11,55***

Notes: 1) Standard errors are below the coefficients in parentheses.
2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7 - Summary of the elasticities for several non-parametric estimators with estimated bandwidths via Least-Squares Cross Validation (LSCV) and Gaussian kernel function for transportation. Dependent variable: Ln of the family budget share spent with transportation.

Pooled					
LCLS I	Mean	Q1	Q2	Q3	R2
Ln (monthly expenditure)	0,166***	0,086	0,172***	0,271***	
	(0,052)	(0,060)	(0,059)	(0,067)	0,09
Ln (household size)	0,059	-0,203**	-0,026	0,309**	
	(0,094)	(0,083)	(0,163)	(0,137)	
LCLS II *					
Ln (monthly expenditure)	0,312**	0,044	0,260	0,571**	
	(0,130)	(0,152)	(0,419)	(0,218)	0,33
Ln (household size)	0,096	-0,164	0,076	0,286***	
	(0,082)	(0,175)	(0,218)	(0,065)	
Fixed Effects HCL (2008)					
Ln (monthly expenditure)	0,182***	0,135***	0,195***	0,247***	
	(0,025)	(0,039)	(0,026)	(0,041)	0,09
Ln (household size)	0,043	-0,235***	-0,011	0,310**	
	(0,071)	(0,073)	(0,057)	(0,135)	
Li e Racine (2004)**					
Ln (monthly expenditure)	0,279*	0,008	0,222	0,485***	
	(0,164)	(0,122)	(0,195)	(0,128)	0,43
Ln (household size)	0,079	-0,356	0,049	0,406	
	(0,369)	(0,569)	(0,356)	(0,367)	

Notes: 1) Table shows the elasticities for 25º(Q1), 50º(Q2), e 75º(Q3) percentiles along with standard errors (in parentheses) estimated by the bootstrap method from 200 replications.
2) * Regressions contain dummies for regions and years.
3) ** Regressions contain dummy for cohorts.
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8 - Parametric regressions for education. Dependent variable: Ln of the family budget share spent with education.

	Pooled OLS I	Pooled OLS II	Fixed Effects OLS-FE
Ln (monthly expenditure)	$0{,}612^{***}$	$0{,}661^{***}$	$0{,}632^{***}$
	(0,030)	(0,031)	(0,029)
Ln (household size)	$0{,}427^{***}$	$0{,}281^{***}$	$0{,}360^{***}$
	(0,083)	(0,088)	(0,071)
North region		-0,049	
		(0,067)	
Southeast region		$-0{,}109^{**}$	
		(0,055)	
South region		$-0{,}360^{***}$	
		(0,063)	
Midwest region		$-0{,}159^{**}$	
		(0,068)	
1995		$0{,}608^{***}$	
		(0,054)	
2003		$0{,}588^{***}$	
		(0,062)	
2009		$0{,}555^{***}$	
		(0,064)	
Constant	$-9{,}461^{***}$	$-9{,}982^{***}$	$-9{,}533^{***}$
	(0,214)	(0,225)	(0,203)
R2	0,127	0,161	0,132
Durbin-Wu-Hausman	$38{,}49^{***}$	$50{,}32^{***}$	
Hausman			2,13

Notes: 1) Standard errors are below the coefficients in parentheses.
2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9 - Summary of elasticities for several non-parametric estimators with estimated bandwidths via Least-Squares Cross Validation (LSCV) and Gaussian kernel function for education. Dependent variable: Ln of the family budget share spent with education.

	Mean	Q1	Q2	Q3	R2
Pooled					
LCLS I					
Ln (monthly expenditure)	0,526	-0,022	0,500	$1{,}092^{**}$	0,20
	(0,481)	(0,249)	(0,610)	(0,472)	
Ln (household size)	$0{,}520^{**}$	0,128	$0{,}460^{***}$	0,840	
	(0,223)	(0,230)	(0,139)	(0,541)	
LCLS II *					
Ln (monthly expenditure)	$0{,}712^{***}$	0,388	$0{,}801^{**}$	$1{,}128^{***}$	0,28
	(0,200)	(0,237)	(0,342)	(0,090)	
Ln (household size)	$1{,}038^{**}$	0,204	0,930	$1{,}829^{**}$	
	(0,505)	(0,739)	(0,927)	(0,719)	
Fixed effects					
HCL (2008)					
Ln (monthly expenditure)	0,498	0,177	$0{,}592^{***}$	0,925	0,23
	(0,304)	(0,812)	(0,183)	(0,235)	
Ln (household size)	0,704	-0,518	0,390	1,806	
	(1,203)	(1,880)	(1,289)	(1,416)	
Li e Racine (2004)**					
Ln (monthly expenditure)	0,577	0,187	$0{,}699^{*}$	$1{,}034^{***}$	0,38
	(0,576)	(0,520)	(0,347)	(0,228)	
Ln (household size)	0,827	-0,212	0,780	$1{,}776^{**}$	
	(1,018)	(1,094)	(1,191)	(0,891)	

Notes: 1) Table shows the elasticities for 25º(Q1), 50º(Q2), e 75º(Q3) percentiles along with standard errors (in parentheses) estimated by the bootstrap method from 200 replications.
2) * Regressions contain dummies for regions and years.
3) ** Regressions contain dummy for cohorts.
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10 - Parametric regressions for health. Dependent variable: Ln of the family budget share spent with health.

	Pooled OLS I	Pooled OLS II	Fixed Effects OLS-FE
Ln (monthly expenditure)	$0{,}256^{***}$	$0{,}280^{***}$	$0{,}282^{***}$
	(0,019)	(0,018)	(0,016)
Ln (household size)	$-0{,}159^{***}$	$-0{,}259^{***}$	$-0{,}275^{***}$
	(0,055)	(0,053)	(0,039)
North region		$0{,}102^{***}$	
		(0,038)	
Southeast region		$0{,}086^{***}$	
		(0,029)	
South region		0,004	
		(0,035)	
Midwest region		0,001	
		(0,039)	
1995		$0{,}982^{***}$	
		(0,031)	
2003		$0{,}778^{***}$	
		(0,034)	
2009		$0{,}836^{***}$	
		(0,033)	
Constant	$-4{,}470^{***}$	$-5{,}230^{***}$	$-4{,}521^{***}$
	(0,138)	(0,129)	(0,110)
R2	0,028	0,251	0,033
Durbin-Wu-Hausman	$43{,}52^{***}$	$45{,}28^{***}$	
Hausman			$4{,}84^{*}$

Notes: 1) Standard errors are below the coefficients in parentheses.
2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11 - Summary of elasticities for several non-parametric estimators with estimated bandwidths via Least-Squares Cross Validation (LSCV) and Gaussian kernel function for health. Dependent variable: Ln of the family budget share spent with health.

Pooled					
LCLS I	Mean	Q1	Q2	Q3	R2
Ln (monthly expenditure)	$0{,}223^{***}$	0,101	$0{,}235^{***}$	$0{,}328^{***}$	
	(0,075)	(0,118)	(0,065)	(0,100)	0,08
Ln (household size)	-0,058	$-0{,}197^{***}$	-0,056	0,066	
	(0,080)	(0,063)	(0,069)	(0,046)	
LCLS II *					
Ln (monthly expenditure)	$0{,}339^{***}$	$0{,}230^{**}$	$0{,}362^{***}$	$0{,}486^{***}$	
	(0,082)	(0,117)	(0,099)	(0,061)	0,33
Ln (household size)	-0,104	$-0{,}308^{*}$	-0,065	0,069	
	(0,083)	(0,160)	(0,132)	(0,165)	
Fixed Effects HCL (2008)					
Ln (monthly expenditure)	0,231	0,159	$0{,}262^{***}$	$0{,}324^{***}$	
	(0,155)	(0,095)	(0,097)	(0,046)	0,09
Ln (household size)	-0,130	$-0{,}351^{**}$	-0,153	0,076	
	(0,098)	(0,138)	(0,204)	(0,333)	
Li e Racine (2004) **					
Ln (monthly expenditure)	0,194	0,146	0,325	$0{,}522^{**}$	
	(0,262)	(0,198)	(0,218)	(0,243)	0,47
Ln (household size)	-0,095	-0,576	-0,106	0,328	
	(0,702)	(1,846)	(0,510)	(0,976)	

Notes: 1) Table shows the elasticities for 25º(Q1), 50º(Q2), e 75º(Q3) percentiles along with standard errors (in parentheses) estimated by the bootstrap method from 200 replications.
2) * Regressions contain dummies for regions and years.
3) ** Regressions contain dummy for cohorts.
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12 - Parametric regressions for the clothing group. Dependent variable: Ln of the family budget share spent with clothing.

	Pooled OLS I	Pooled OLS II	Fixed Effects OLS-FE
Ln (monthly expenditure)	-0,084***	-0,040***	-0,062***
	(0,016)	(0,015)	(0,013)
Ln (household size)	0,210***	0,103**	0,161***
	(0,042)	(0,042)	(0,033)
North region		0,116***	
		(0,032)	
Southeast region		-0,185***	
		(0,025)	
South region		0,033	
		(0,027)	
Midwest region		-0,152***	
		(0,030)	
1995		0,541***	
		(0,029)	
2003		0,627***	
		(0,030)	
2009		0,583***	
		(0,031)	
Constant	-2,179***	-2,764***	-2,285***
	(0,113)	(0,113)	(0,094)
R2	0,014	0,187	0,014
Durbin-Wu-Hausman	0,73	1,58	
Hausman			1,43*

Notes: 1) Standard errors are below the coefficients in parentheses.
2) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13 - Summary of elasticities for several non-parametric estimators with estimated bandwidths via Least-Squares Cross Validation (LSCV) and Gaussian kernel function. Dependent variable: Ln of the family budget share spent with clothing.

Pooled					
LCLS I	Mean	Q1	Q2	Q3	R2
Ln (monthly expenditure)	-0,056	-0,171***	-0,066	0,072	0,05
	(0,088)	(0,056)	(0,070)	(0,077)	
Ln (household size)	0,057	-0,104	-0,010	0,181***	
	(0,053)	(0,088)	(0,093)	(0,050)	
LCLS II *					
Ln (monthly expenditure)	0,025	-0,296	-0,013	0,271***	0,33
	(0,178)	(0,278)	(0,253)	(0,120)	
Ln (household size)	0,079	-0,141	0,156	0,348***	
	(0,104)	(0,257)	(0,182)	(0,121)	
Fixed effects HCL (2008)					
Ln (monthly expenditure)	0,060**	-0,089***	-0,070***	-0,021	0,04
	(0,023)	(0,016)	(0,018)	(0,033)	
Ln (household size)	0,070*	-0,071**	0,044	0,194***	
	(0,036)	(0,032)	(0,045)	(0,045)	
Li e Racine (2004)**					
Ln (monthly expenditure)	0,032	-0,227	-0,023	0,223	0,43
	(0,243)	(0,279)	(0,155)	(0,294)	
Ln (household size)	0,125	-0,280	0,206	0,569**	
	(0,312)	(0,296)	(0,749)	(0,269)	

Notes: 1) Table shows the elasticities for 25º(Q1), 50º(Q2), e 75º(Q3) percentiles along with standard errors (in parentheses) estimated by the bootstrap method from 200 replications.
2) * Regressions contain dummies for regions and years.
3) ** Regressions contain dummy for cohorts.
4) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 Concluding remarks

The objective of this work was to carry out an econometric analysis based on the Engel curve. Of major interest we are interested in comparing parametric estimates of the piglog model, widely adopted by the empirical studies in Brazil, with the nonparametric models, being the latter still little used. As mentioned by Henderson and Parmeter (2015), the recent availability of large data sets in addition to advances in computational capacity have promoted the diffusion of more applied works in the field of non-parametric econometrics.

As also mentioned by these authors, from the economic point of view related to the design of public policies, the results of a fully parametric model would be more favorable than the results of non-parametric methods because they are easier to interpret. In the case of the present study, it was exactly what was observed for the elasticities of total expenditure and the number of household components of all consumption groups when comparing the average elasticity of the percentiles generated by different non-parametric estimators with their fully parametric counterpart models. However, when opening to the different percentiles from the former, one can see a different impact of the variables mentioned above in relation to the share spent of each good, deserving therefore more attention of the future works of new procedures that are able to fit with the nature of the data accordingly.

In addition, we did not corroborate the hypothesis that the use of parametric or non-parametric methods, specifically developed to control the unobserved characteristics of heads of households from a panel data structure, would not converge to similar results to those that follow a structure of cross-sectional data as stated by the literature. In almost all groups of goods, the results of the former were very close to those of the latter in relation to total expenditure and the number of components in the family.

An important limitation of the present work is that it must be taken into account that all decisions to acquire the good are made based on the budget constraint of families, as postulated by the theory of the consumer. In this case, as exposed during the introduction, the use of AIDS or QUAIDS models of simultaneous estimation along with the piglog equations for each consumption group would be more appropriate. However, the use of this tool within the field of non-parametric econometrics that relaxes the piglog function of the linear models, to date, has not yet been developed, and represents a promising avenue in the research of quantitative methods in economics.

References

Almeida, A. N; Freitas, R. E. (2006). 'Renda e Despesa Familiar no Brasil segundo a Pesquisa de Orçamentos Familiares (POF) 2002-2003'. *Texto para Discussão 1235.* Brasília: IPEA.

Almeida, A. N. & Azzoni, R. E. (2016), 'Custo de vida comparativo das regiões metropolitanas brasileiras: 1996-2014', Estudos Econômicos, 46(1), 253-276, 2016.

Almeida, A. T. C., Mesquita, S. P. & Silva, M. V. B. (2016), 'Impactos do programa bolsa Família sobre a diversificação do consumo de alimentos no Brasil', Pesquisa e Planejamento Econômico, 46(1), 7-39.

Azzoni, C. A., Carmo, H. E. & Menezes, T. (2000), 'Índice de custo de vida comparativo para as principais regiões metropolitanas brasileiras: 1981-1999', Estudos Econômicos, 30(1), 165-186.

Azzoni, C. A., Carmo, H. E. & Menezes, T. A. (2003) 'Comparação da paridade do poder de compra entre cidades: aspectos metodológicos e aplicação ao caso brasileiro', Pesquisa e Planejamento Econômico, 33(1), 91-126.

Azzoni, C. R., Guilhoto, J., Silveira, F. G., Menezes, T., Haddad, E. A., Hazegawa, M. (2009). 'Commodity price changes and their impacts on poverty in Developing Countries: the Brazilian case.'. *Studies in Regional Science* / Chiikigaku kenky, v.39, p.131 – 147.

Baltagi, B. H. (2008), 'Econometric Analysis of Panel Data', John Wiley & Sons Ltd. 4ed, West Sussex.

Barbosa, A. L. N. H., Menezes, T. A. & Andrade, B. C. (2014), 'Demanda por produtos alimentares nas áreas rurais e urbanas do Brasil', Pesquisa e Planejamento Econômico, 44(3), 507-543.

Banks, J., Blundell, R. & Lewbel, A. (1997) 'Quadratic Engel curves and consumder demand', The Review of Economics Statistics, 79(4), 527-539.

Bhalotra, S. & Attfield, C. (1998) 'Intrahousehold resource allocation in rural Pakistan: a semiparametric analysis', Journal of Applied Econometrics, 13(5), 462-480.

Blundell, R. W., Browning, M. & Crawford, I. A. (2003) 'Nonparametric engel curves and Revealed Preference', Econometrica, 71(1), 205-240.

Blundell, R. & Duncan, A. (1998), 'Kernel regression in empirical economics', Journal of Human Resources, 33(1), 62-87.

Blundell, R., Duncan, A. & Pendakur, K. (1998) ‘Semiparametric estimation and consumer demand’, Journal of Applied Econometrics, 13, 435-461.

Browning, M., Deaton, A. & Irish, M. (1985), ‘A profitable approach to labor supply and commodity demands over the life-cycle’ Econometrica, 53, 503-543.

Camarano, A. A.; Pasinato, M. T. (2004). ‘Introdução. In: Camarano, A. A. (org). *Os Novos Idosos Brasileiros: Muito Além dos 60?’* IPEA: Rio de Janeiro.

Camarano, A. A., Kanso, S., & Fernandes, D. (2014). Menos jovens e mais idosos no mercado de trabalho? In A. A. Camarano (Ed.), Novo Regime Demográfico uma Nova Relação entre População e Desenvolvimento? (pp. 377–406). Rio de Janeiro: IPEA.

Cameron, A. & Trivedi, P. (2005) ‘Microeconometrics: Methods and Applications’, Cambridge University Press, Cambridge.

Coelho, A. B. & Aguiar, D. R. D. (2007), ‘O modelo Quadrático Almost Ideal Demand Sytem (Quaids): uma aplicação para o Brasil’, in ‘Silveira, F. G. et al. (org). Gasto e Consumo das Famílias Brasileiras Contemporâneas’, 2ed. IPEA, Brasília.

Coelho, A. B., Aguiar, D. R. D. & Fernandes, E. A. (2009), ‘Padrão de consumo de alimentos no Brasil’, Revista de Economia e Sociologia Rural, 47(2), 335-362.

Coelho, A. B., Aguiar, D. R. D. & Eales, J. S. (2010) ‘Food demand in Brazil: an application of Shonkwiler & Yen two-step estimation method’, Estudos Econômicos, 40(1), 185-211.

De Janvry, A. & Saudolet, E. (1995), ‘Quantitative development policy analysis’, The John Hopkins University Press, Baltimore.

Deaton, A. (1997), ‘The Analysis of Household Surveys: A Microeconometric Approach to Development Policy’, The Johns Hopkins University Press, Baltimore.

Deaton, A. & Muellbauer, J. (1980), ‘Economics and consumer behavior’, Cambridge University Press, New York.

Delgado, M. S., Henderson, D. J. & Parmeter, C. F. (2014), ‘Does education matter for economic growth?’, Oxford Bulletin of Economics and Statistics, 76 (3), 334-359.

Eubank, R. L. (1999), ‘Nonparametric regression and spline smoothing.’ 2ed. Series Statistics: A Series of Textbooks and Monographs, CRC Press.

Ferreira, A. S. & Coelho, A. B. (2016), ‘Substituição entre alimentos orgânicos e convencionais no Brasil: o papel dos preços e do dispêndio’, in: Encontro Nacional de Economia 44, ANPEC, Foz do Iguaçu.

Gong, X., Van Soest, A. & Zhang, P. (2005), ‘The effects of the gender of children on expenditure patterns in rural China: a semiparametric analysis’, Journal of Applied Econometrics, 20, 509-527.

Härdle, W. (1990), ‘Applied Nonparametric Regression’, Cambridge University Press, New York.

Henderson, D. J., Carroll, R. J. & Li, Q. (2008), ‘Nonparametric estimation and testing of fixed effects panel data models’, Journal of Econometrics, 144(1), 257–275.

Henderson, D J. & Parmeter, C. F. (2015), ‘Applied nonparametric econometrics’, Cambridge University Press, New York.

Hoffmann, R. (2010), ‘Estimação das elasticidades-renda de várias categorias de despesa e de consumo, especialmente alimentos, no Brasil, com base na POF de 2008-2009’, Revista de Economia Agrícola, 57(2), 49-62.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2010), ‘Pesquisa de Orçamentos Familiares 2008-2009: Despesas, Rendimentos e Condições de Vida’, IBGE, pp. 222, Rio de Janeiro.

Leser, C. E. V. (1963), ‘Forms of Engel Functions’, Econometrica, 31(4), 694-703.

Li, Q. & Racine, J. S. (2004), ‘Cross-validated local linear nonparametric regression’, Statistica Sinica, 14, 485-512.

Li, Q. & Racine, J. (2007), ‘Nonparametrics Econometrics: Theory and Practice’, Princeton University Press, Princeton.

Medeiros, J. A. S. (1978), ‘Curvas de Engel e transformação de Box-Cox: Uma aplicação aos dispêndios em alimentação e educação na cidade do Rio de Janeiro’, Pesquisa e Planejamento Econômico, 8(3), 795-828.

Menezes, T. A., Azzoni, C. R. & Silveira, F. G. (2008), ‘Demand elasticities for food products in Brazil: A Two-Stage Budgeting System’, Applied Economics, 40, 2557-2572.

Nadaraya, E. A. (1964), ‘On estimating regression’, Teoriya Veroyatnostei i ee primeneniya, 9, 157-159.

Oliveira, F. C. R. & Hoffman, R. (2015), ‘Consumo de alimentos orgânicos e de produtos light ou diet no Brasil: fatores condicionantes e elasticidade-renda’, Revista Segurança Alimentar e Nutricional, 22(1), 541-557.

Pereda, P. C. & Alves, D. C. O. (2012), ‘Qualidade alimentar dos brasileiros: teoria e evidência usando demanda por nutrientes’, Pesquisa e Planejamento Econômico, 42(2), 239-260.

Pintos-Payeras, J. A. (2009), ‘Estimação do sistema quase ideal de demanda para uma cesta ampliada de produtos empregando dados da POF de 2002-2003’, Economia Aplicada, 13(2), 231-255.

Resende, A. C. C. & Oliveira, A. M. H. C. (2008), ‘Avaliando resultados de um programa de transferência de renda: o impacto do Bolsa-Escola sobre os gastos das famílias brasileiras’, Estudos Econômicos, 38(2), 235-265.

Rossi, J. W. (1982), ‘Elasticidades de Engel para dispêndios Familiares na cidade do Rio de Janeiro.’, Pesquisa e Planejamento Econômico, 12(2), 579-606.

Ryder, N. B. (1965), ‘The cohort as a concept in the study of social change’, American Sociological Review, 30(6), 843-861.

Santana, P. J. & Menezes, T. A. (2009), ‘Diferenças raciais no padrão de gastos com educação: Uma Abordagem Semiparamétrica’, Nova Economia, 19(3), 383-405.

Schmalensee, R. & Stoker, T. M. (1999), ‘Household gasoline demand in the United States’, Econometrica, 67(3), 645–662.

Silva, A. S. P., Scherer, C. E. M. & Porsse, A. A. (2012), ‘A nova classe média: alterações de consumo e seus efeitos regionais, Revista Brasileira de Estudos Regionais e Urbano, 6(1), 1-15.

Silva, M. M. C. & Coelho, A. B. (2014), ‘Demanda por frutas e hortaliças no Brasil: uma análise da influência dos hábitos de vida, localização e composição familiar’, Pesquisa e Planejamento Econômico, 44(3), 545-578.

Silveira, F. G., Servo, L. M. S., Menezes, T. & Piola, S. F. (org.). (2006), ‘Gasto e Consumo das Famílias Brasileiras Contemporâneas’, 1ed. IPEA, Brasília.

Silveira, F. G., Servo, L. M. S., Menezes, T. & Piola, S. F. (org.). (2007), ‘Gasto e Consumo das Famílias Brasileiras Contemporâneas’, 2ed. IPEA, Brasília.

Travassos, G. F. & Coelho, A. B. (2014), ‘A questão da separabilidade fraca na estimação de sistemas de demanda: uma aplicação para a demanda de carnes no Brasil’, Pesquisa e Planejamento Econômico, 44(3), 545-578.

Verbeek, M. & Nijman, T. (1992), ‘Can cohort data be treated as genuine panel data?’ Empirical Economics, 17, 9-23.

Watson, G. S. (1964), ‘Smooth regression analysis’, Sankhay, 26, pp. 175-184.

Working, H. (1943), ‘Statistical laws of family expenditure’, Journal of the American Statistical Association, 38, 43-56.

Yatchew, A. (2003), ‘Semiparametric Regression for the Applied Econometrician’, in P.C.B. Phillips (Ed.) Themes in Modern Econometrics, Cambridge University Press, Cambridge.

Appendix A

To illustrate an example of cohort, we consider using a sample without distinguishing men and women as household heads. Of the 21,321 observations of the total sample, 23.36% corresponds to the heads of POF 08/09. Thus, Table 1A shows the number of occurrences (heads) of each cohort for the POF 2008/2009.

It can be seen that 21 cohorts are constructed, with the fourth cohort corresponding to a 103-year-old head individual, while in the younger cohort (last one) there are 49 individuals with a mean age ranged between 16 and 25 years. The largest number of individuals was enrolled in the cohort 16, with approximately 521 male and female heads, and a mean age of 43 years. The next step is to apply this same procedure to POFs 87/88, 95/96 and 02/03.

Table 1A- Number of male and female heads ranked according to their age in 2009 for each cohort.

Cohort	Frequency	Mean-Age	Min	Max
1	-	-	-	-
2	-	-	-	-
3	-	-	-	-
4	1	103	103	103
5	2	99	98	100
6	7	92	91	95
7	31	88	86	90
8	75	83	81	85
9	162	78	76	80
10	285	73	71	75
11	371	68	66	70
12	389	63	61	65
13	440	58	56	60
14	477	53	51	55
15	450	48	46	50
16	**521**	**43**	**41**	**45**
17	454	38	36	40
18	480	33	31	35
19	519	28	26	30
20	273	23	21	25
21	49	19	16	20

Fonte: Own's calculation.

Next, it is necessary to verify if the same cohort - in our example the cohort 16 compound from individuals with average age of 43 years in POF 08/09 - will effectively represent younger individuals, according to the previous POFs. Following the same previous example, Table 2A reports the cohorts for the year 1988 according to the number of individuals and ages at their mean values. It can be observed that the cohort 16 now consists of 99 individuals with a mean age of 23 years.

Table 2A - Number of male and female heads ranked according to their age in 1988 for each cohort.

Cohort	Frequency	Mean-Age	Min	Max
1	3	96	95	98
2	3	91	90	93
3	15	87	85	89
4	40	81	80	84
5	61	77	75	79
6	85	72	70	74
7	94	67	65	69
8	105	62	60	64
9	101	57	55	59
10	101	52	50	54
11	105	47	45	49
12	99	42	40	44
13	105	37	35	39
14	106	32	30	34
15	108	27	25	29
16	**99**	**23**	**20**	**24**
17	30	18	15	19
18	1	13	13	13

Fonte: Own's calculation.

Finally, following the example above, Table 3A shows the evolution of the mean age and average birth year of the cohort 16. In summary, all the head individuals who were born between 1964 and 1968 and had a mean age of 23 years in 1988, 30 years in 1996, 37 years in 2003, and 43 years in 2009 will be assumed from now on as if they were a group of same individuals, and monitored over time.

This logic of accompanying different individuals at different ages, according to the years of socioeconomic surveys, is what characterizes a pseudo-panel. Consequently, cohorts should constitute as a good approximation of panel data in econometric analyzes (Deaton, 1997, Cameron and Trivedi, 2005).

Table 3A - Evolution of the mean age and birth year of the cohort 16 based on the year of each POF.

	Year of implementation of POFs			
	1988	1996	2003	2009
Year of Birth (mean)	1966	1966	1966	1966
Age (mean)	23	30	37	43
Cohort	16	16	16	16

Fonte: Own's calculation.

Appendix B

Composition of Consumption Groups

Habitation	Transportation	Clothing	Hygiene
Electricity	Public transportation	Men´s Clothing	Perfume
Water and Sewage	Gasoline – Own Vehicle	Women´s Clothing	Hair Products
Cooking gas	Ethanol - Own Vehicle	Children's Clothes	Soap
Condominium fees	Vehicle accessories and Maintenance	Shoes and similar	Products for Personal Use
Home Phone	Occasional Travel	Jewelry	
Rental Income			

Health	Education	Leisure	Personal Expenses
Medicines	Paid Courses	Toys and Games	Hair
Health Insurance	Other Courses and Activities	Periodicals, Non-textbooks and Magazines	Pedicure
Dental treatment	Textbooks and Technical Journals	Recreation and Sports	Repair of Personal Items
Medical Appointments	School Supplies	Games and Bets	
Surgery services			
Hospitalization			
Physical Examinations			

Other Home Expenses	Other Property	Other Expenses	Charges
Cleaning Supplies	Occasional-use Properties (rented)	Ceremonies and Parties	Labour Contributions
Home Furnishings	Properties (owned)	Professional Services	Private Pension
Home Maintenance			Taxes
Home Renovation	**Communication**	**Smoke**	
Other Related Products	Post Office Fees	Cigarettes, cigars, etc.	

Food

Rice	Tomato	Fresh Fish	Ground Coffee
Beans	Onion	Chicken	Soft Drinks
Prepared Foods	Lettuce	Eggs	Beer and Chips
Pasta	Banana	Cow's Milk	Other Alcoholic Beverages
Wheat Flour	Orange	Powdered Milk	Canned Food
Cassava Flour	Apple	Cheeses	Tomato Paste
Potato	Beef (high quality)	Bread (French)	Mayonnaise
Carrot	Beef (inferior quality)	Cookies	Refined Salt
Cassava	Pork	Soy Oil	Eating Away from Home
Refined Sugar		Olive Oil	
Crystal Sugar			