the review of income and wealth

Review of Income and Wealth
Series 62, Supplement 1, August 2016
DOI: 10.1111/roiw.12234

CONSTRUCTION OF A CONSUMPTION AGGREGATE BASED ON INFORMATION FROM POF 2008-2009 AND ITS USE IN THE MEASUREMENT OF WELFARE, POVERTY, INEQUALITY AND VULNERABILITY OF FAMILIES

by Leonardo S. Oliveira*, Debora F. De Souza, Luciana A. Dos Santos, Marta Antunes, Nícia C. H. Brendolin and Viviane C. C. Quintaes

Brazilian Institute of Geography and Statistics (IBGE), República do Chile Avenue, 500, Rio de Janeiro – RJ, Brazil

Given the complexity of family and individual welfare, this study aims to explain the construction of a family consumption aggregate, using data from Brazilian Family Expenditure Survey (2008–2009), and also to measure and analyze welfare, poverty, inequality and vulnerability. Following the literature, some aspects were considered: the selection of expenditures, the analysis of extreme values, the imputation of food consumption, the user cost of durable goods and a spatial price deflator. After the definition of the family consumption aggregate, we analyzed the Generalized Lorenz Curves, social welfare functions and inequality measures. Then we presented the sensitivity of the identification exercise to different poverty lines and analyzed the severity of poverty. Finally, based on Chaudhuri *et al*. (2002) and Elbers *et al*. (2002), the vulnerability to poverty was studied taking account of area (clusters) effects. In this last exercise, the poverty line was based on half the minimum wage in 2008.

JEL Codes: C21, D63, I32

Keywords: social welfare functions, inequality, poverty measurement, error component models, heteroscedastic

1. Introduction

The Brazilian Family Expenditure Survey (POF) provides information about the household budget composition, consumer habits, expenditure and income distribution, household and people characteristics. The data collection focus on expenditure and acquisition of goods and services. To make consumption aggregate, it is necessary to identify the components of current expenses associated with consumption as well as the value of consumption associated with the ownership of assets, which guarantee a flow of services for the consumption unit.[1] Therefore, for assembling the consumption aggregate through the POF 2008–

Note: IBGE is exempt from any responsibility related to the opinions, information, data and concepts stated in this article that are of exclusive responsibility of the authors.

The authors would like to thank Paulo Roberto Coutinho Pinto for his collaboration and Elisa L. Caillaux and Marina Aguas for their comments.

*Correspondence to: Leonardo S. Oliveira, Brazilian Institute of Geography and Statistics (IBGE), República do Chile Avenue, 500, Rio de Janeiro – RJ, Brazil.

[1]The Brazilian Family Expenditure Survey worked with the concept of Consumption Unit, which can be approximated to the idea of household units or family. For further details, see IBGE (2008). The Brazilian Family Expenditure Survey has also investigated the self-perception of life quality (POF questionnaire 6) and the characteristics of the nutritional profile in the Brazilian population (POF questionnaire 7). However, in the present stage of the aggregate consumption construction these data will not be used.

2009, a number of decisions had to be taken based on theoretical hypothesis and empirical results, as it is presented in Section 1.

The choice of using consumption for measuring welfare, poverty, inequality and vulnerability of consumption units is justified by two issues.[2] First, both income and consumption present a variation over time but consumption tends to be less variable than income and to reflect the average long term welfare more accurately (Deaton, 1997; Deaton and Zaidi, 2002; Haughton and Khandker, 2009). Income fluctuations do not replicate directly into consumption fluctuations because people might adapt or use credit, get donations or sell their assets in order to keep their consumption standard. Note that consumption units with restricted access to credit will face more difficulties to boost their consumption. Even considering some consumption seasonal fluctuations, associated with holidays or festivities, these are smoother when compared to the income fluctuations of consumption units, especially when their members are self-employed workers or employees of the informal sector. Those who work in agriculture sectors are subject to higher seasonal fluctuations in income than in consumption because a bigger portion of their consumption comes from their own production and not from the market.[3] The action of asking informants to estimate the value of goods acquired outside the market (donations, production for their own consumption or withdrawal from their own businesses) in the survey allows the measurement of non-monetary consumption.[4] If this non-monetary consumption was not considered, there would be an underestimation of welfare and a over-estimation of poverty.

Secondly, consumption reflects the value that people attribute to different goods and services at their disposal: food consumption; durable goods; housing; healthcare, education and transportation and other non-food items. In this process, the consumption units will base their choices on market prices and the possibility of substituting goods and services. As a result, the consumption aggregate weights the different goods and dimensions by market prices. Furthermore, Ravallion (2011) emphasized the role of prices in the definition of opportunity costs and marginal rates of substitution as one of the major advantages of using consumption aggregates as welfare indicators.

[2]Haughton and Khandker (2009) clearly emphasized that both consumption and income are imperfect proxies of utility once they exclude important contributions to welfare such as publicly provided services and goods. Atkinson *et al.* (2002) highlighted that surveys on living conditions measure expenditure but not consumption, that is to say, the amount spent by a consumption unit during the specific period of time of the survey expenditure collection may differ from the effective consumption in the same period of time. This difference can be due, for example, to the use of stock holdings. The same argument applies for durable goods (see Section 1.2). Limitations, criticism and alternatives to the use of both expenditure (consumption) and income as welfare measures can be found in Sen (2004, 2008 and 2009), Kakwani and Silber (2007 and 2008), Oliveira (2010) and in the Journal of Economic Inequality (2011).

[3]Haughton and Khandker (2009) compared the welfare measurement through income, which they called "potential", and through consumption, which they called "result". They stated that income tends to be more seasonal and underreported than consumption, reaching the same opinion of Atkinson (1998), Deaton and Zaidi (2002) and OECD (2013).

[4]The POF team evaluated and selected part of these data for imputation in order to assure consistency. Nevertheless, if informants were not asked to estimate the value attributed to these goods, 100% of imputation would be needed.

An analysis of individual welfare based on a consumption aggregate has an implicit money metric utility function[5] that returns the necessary amount for keeping the consumption unit welfare level and requires consumption to be adjusted by a price index. In order to do that, it is necessary to construct price deflators that indicate life costs differences among distinct Brazilian geographical contexts, as described in Section 1. In Section 2, an analysis of social welfare is performed using the consumption aggregate constructed, in which the Generalized Lorenz Curve and (abbreviated) social welfare functions based on Sen and Atkinson's works are considered. To understand the weight of inequality in the reduction of social welfare, two breakdowns are done: i) through the Gini index, consumption inequality is broken down by component; and ii) through the mean logarithmic deviation, inequality is broken down by population subgroups considering the years of education, sex, color and race of the person responsible for the consumption unit. In Section 3, poverty and vulnerability analyses are presented based on poverty curves, square poverty gap index (severity of poverty) and the probability of a consumption unit becoming poor.

To make it easier to understand and/or reproduce the structure of the consumption aggregate, we provide all the codes of POF items (Consumption Aggregate codes.xlsx, Supporting information 1) used for its construction as supporting information. You can also find a database containing the variables to form the households' consumption unit's key used in POF 2008–2009 in addition to the consumption aggregate and the spacial price deflator (database.csv, Supporting information 1). We provide, in the end of the paper in Appendix 1 (Supporting information 2), some descriptive statistics and graphics comparing the consumption aggregate constructed in this paper with the total expenditure identified by the survey (that includes all the monetary and non-monetary acquisitions registered by POF) and the general consumption expenditure registered in POF (that results from the total expenditure minus the items that represent patrimonial variation and other recurrent expenses: taxes, labor deductions, bank services, pensions, allowances, donations and private social security). An analysis of this material permits the identification of differences in the distribution of these three welfare indicators as well as significant impact on the measuring of inequality and poverty. Such differences behave as expected considering the way information was handled to assemble the consumption aggregate, as described in the following sections.

1. Consumption Aggregate

Assembling the consumption aggregate is a complex exercise that requires fine discrimination between which expenses items might be included or not, so as to allow the comparison of the consumption units' welfare levels with its correct ranking. This discrimination is guided by applied literature and theoretical hypotheses about welfare contribution of different goods and services, as well as by the necessary adaptations to the culture of the country under study.

Hentschel and Lanjow (1996) presented general principles to guide the construction of the consumption aggregate and applied them to the poverty analysis

[5]See Varian (1992, p.108–110) and Deaton and Zaidi (2002, p.4–13) to learn about the usefulness of the money metric utility and its relation with other forms of measuring welfare.

in Ecuador. Slesnick (2001) studied poverty and social welfare in the U.S. and gave special attention to consumption and its components. Lanjow and Lanjow (2001) analyzed the effect of the measurement error in consumption and suggested the exclusion of some components in the welfare analysis. Deaton and Zaidi (2002) advanced in the discussion about the consumption aggregate using family expenditure survey data of eight countries. Haughton and Khandker (2009) also helped to disseminate steps used in the construction and analyses of the consumption aggregate. Lanjow (2009) examined the Brazilian case and recommended procedures to the inclusion of consumption items in the welfare analysis. These authors suggested methodological ways to theoretical and practical problems faced in the construction of such aggregates.

Considering the POF methodology, which measures expenses made per consumption unit by type and in differentiated periods of time (7, 30 and 90 days and 12 months), some criteria needed to be determined to deal with the information collected by the survey, adapting them to the consumption aggregate construction. First, it was necessary to group the expenses of the different POF blocs into consumption groups in order to select the ones that would compose the aggregate and the ones that would be excluded. The following consumption groups were defined: food; durable goods; housing; education, health and transportation; and other non-food items. Subsequently, each item of these groups was analyzed as to verify if they complied with the following criteria:

(a) The item acquisition is not sporadic, i.e., it is frequently acquired in such a way that the collection period of the survey is sufficient and does not distort the welfare analysis among the consumption units. The durable goods whose acquisition tends not to occur annually received a differentiated treatment (see Section 1.2).
(b) The item is acquired for the consumption unit own consumption, i.e., the acquisition of such good will increase the welfare of the consumption unit under analysis and not that of another unit.
(c) The item contributes to welfare comparability between different units and its corrected ranking.

1.1. *Food Expenditure*

Food expenditures are obtained in the questionnaire "Collective Acquisition Booklet" (POF 3) and in the bloc meals out-of-home (bloc 24) from the questionnaire "Individual Acquisition" (POF 4). The expenses with food were totally included, considering this is an important group when it comes to measure the consumption units' welfare (the share of food expenditures and other items in the aggregate consumption are presented in section 2.3). It was found that 5.4 percent of the consumption units presented null food expenditure. This can be explained by the reference time period of information collection (7 days) that reported zero for consumption units that did not acquire food in that week. In order to correct possible distortions in social welfare, inequality, poverty and vulnerability due to these null food expenditures, an imputation was made following the Propensity Score method. In this method, we estimate the probability of each consumption unit to have food expenditure and we compare the consumption units with and

without food expenditures using those probabilities. In the end, the consumption units with food expenditure are used as donors of food expenditure for the consumption units without food expenditures. For further details, see Rosenbaum and Rubin (1983).

Appendix 2 (Supporting information 3) shows the explanatory variables of the logit model applied and density functions of the per capita consumption with and without food expenditure imputation.

1.2. *Durable Goods*

The possession of durable goods has positive impacts over consumption unit's welfare. However, while the acquisition of durable goods occurs in a particular point in time, its consumption may occur during several years, as Haughton and Khandker (2009) and Atkinson (1998) pointed out. There is difficulty in defining which goods might be considered durable, once, as Atkinson (1998) explained, there is a durable element in several goods. In the construction of the consumption aggregate, the following were considered as durable goods: "Main household durable goods inventory" (bloc 14); "Machines, equipment and household utilities acquisition" (bloc 15); "Tools, pets, musical instruments and camping gear acquisition" (bloc 16); "Furniture acquisition" (bloc 17) and "Vehicles acquisition" (bloc 51). The exception were veterinarian services and expenses with pets in bloc 16 that were included in the aggregate as other non-food expenses.

However, these durable goods had to be selected and treated before being included in the aggregate. Following the suggestion made by Hentschel and Lanjow (1996), Slesnick (2001), Deaton and Zaide (2002), ILO (ICLS 17- (2003)), Haughton and Khandker (2009) and the OCDE (2013) guide for household income, consumption and wealth statistics, the durable goods were included in the consumption aggregate taking account of the value of services (not their value of acquisition), which may be measured by the "user cost" or "rental equivalent" that the consumption unit "receives" for all durable goods in its possession during a period of time of one year. This user cost value (UC) can be approximated by:

$$UC_{tij} = S_{tij} P_{ti} (r_t - \pi_t + \sigma_i) \tag{1}$$

where S_{tij} is the stock of the durable good i (quantity listed in the inventory) in the consumption unit j during the survey period t; P_{ti} is the current value of durable good i during the survey period t; r_t is the nominal interest rate; π_t is the inflation in the survey period t; and σ_i is the depreciation rate of the durable good i.

The depreciation rate is given by the following formula:

$$\sigma_i - \pi_t = 1 - \left(P_{ti} / P_{i(t-T_i)}\right)^{1/T_i} \tag{2}$$

where T_i is the lifetime of durable good i in years and $P_{i(t\text{-}Ti)}$ is its price in the year it was acquired.

POF bloc 14 ("Main household durable goods inventory") informs which goods are owned by the consumption unit during POF collection period t and its

stock S_t, as listed in the inventory. Moreover, in relation to the last acquisition of these goods, it is possible to know the way it was acquired, the year it was acquired and its condition, whether firsthand or secondhand. Thus, we have the stock S_t and the lifetime of the item in years (T). The exception is for the secondhand goods, for which there is only the last acquisition date. Therefore, in order to obtain the user cost for each good, it is necessary to calculate the current prices of each good, the average nominal interest rate for the POF time period and the regional deflators. These steps will be detailed below.

a) *Median price calculation by Federative Unit* ($P_{ti\ UF}$)

Considering that through POF collected data it is not possible to calculate the current price of each durable good, the solution was to estimate this through the calculation of the median price for every durable good, by Federative Unit, using the information about the price of these goods in POF bloc 15 ("Machines, equipment and household utilities acquisition") and POF bloc 51 ("Vehicles acquisition"). This calculation used only firsthand products prompt paid per Federative Unit. The use of the median price aimed to minimize the outliers' impact in each estimated price.

It must be highlighted that the only goods that had their prices studied were the ones that appeared in POF bloc 14 ("Main household durable goods inventory") and POF bloc 15 or POF bloc 14 and POF bloc 51, since it was necessary to match the information in stock and price. Therefore, the goods that were not in the inventory but were present in bloc 15 were excluded from the consumption aggregate. Their insertion would generate a distortion between the consumption units that acquired durable goods during the period of the survey (from May 2008 to May 2009), which would have a higher consumption aggregate, and the ones that acquired the same goods in another time period not covered by POF, which would have a smaller aggregate.

For the calculation of durable good i current price (P_{tiUF}), the median price of the Federative Unit where the consumption unit locates was chosen. However, in some Federative Units there was no acquisition of some specific goods according to previously established standards: goods acquired firsthand and prompt payment. For these cases where it was not possible to calculate the median price of the durable good i per Federative Unit, the median price of the good i of the corresponding Major Region was used for the calculation of the user cost.

In a second exercise, we calculated the mean prices instead of median prices, reaching values slightly larger due to the asymmetric distribution and possible outliers. Thus, we continued using the median prices.

b) *Mean nominal rate of interest calculation* (r_t)

Deaton and Zaidi (2002) suggested the use of one real interest rate only, based on an average of several years for all durable goods. SELIC (Special System of Settlement and Custody) daily rate information provided by the Brazilian Central Bank was used to calculate the average nominal interest rate (1.1261), opting for the POF period – from May 2008 to May 2009.

c) *Calculation of the average regionalized real interest rate of the period*

For the calculation of the real interest rate, besides the average nominal interest rate the inflation rate of the period was needed. Even though there is not available price index (IPCA) information for all Federative Units, a deflator of a geographical area of influence was used where such information was unavailable.

d) *Depreciation rate (σ)*

POF bloc 14 provided no information on the prices of goods when they were acquired ($P_{i(t-Ti)}$). For this reason, an estimate of the depreciation rate had to be made, following the approach suggested by Deaton and Zaidi (2002), learning from other countries experiences, such as: Vietnam, Nepal, Ecuador and Panama.

The calculation of the mean usage time ($\bar{T}_i$) of durable goods acquired firsthand and prompt paid is made by using the data of the year of acquisition registered in the inventory (POF bloc 14). By usage time (T) we mean the difference of years between 2009 (top limit of the survey period) and the year of acquisition (A) of the durable good i reported in the inventory of consumption unit j:

$$\bar{T}_i = mean(2009 - A_{ij}) \tag{3}$$

Also according to Deaton and Zaidi (2002) and Hentschel and Lanjow (1996) suggestions, the mean useful lifetime of each durable good $\bar{L}_i$ was considered as twice the mean usage time (T) for each durable good, considering the sampling design of the survey:[6]

$$\bar{L}_i = 2\bar{T}_i \tag{4}$$

We applied this for all durable goods listed in the inventory. We highlight that, only items in good condition, functioning, or waiting for repairs for future use were listed in the POF's inventory.

The depreciation rate is then calculated by the following formula:

$$\sigma_i = 1/\bar{L}_i \tag{5}$$

For the durable goods that were not in the inventory and that were acquired by the consumption unit during the 12 month period of the

[6]We checked if the estimated depreciation rates were in line with the Brazilian context. Then, these results were compared with the Regulatory Instruction SRF number 162 from December 31, 1998, which establishes the useful lifetime and depreciation rate of goods related to the Mercosur Common Nomenclature (MCN) and other goods. By these norms, we have an estimated useful lifetime of about $1.7\bar{T}$. So using $2\bar{T}$ looks reasonable for the Brazilian case since it points to a longer (but not so different) lifetime than the one in the Regulatory Instruction. Note that the norm considers durable goods for commercial purposes and not durable goods acquired by a household consumption unit. Futhermore, the use of $2\bar{T}$ could be motivated by the hypothesis that the acquisitions are distributed in a uniform way over time and none of the inventory items were recently introduced in the market. It must be noted that the mean time was calculated only for goods acquired new, once there is no information about the real usage time for secondhand goods.

survey (POF blocs 16 and 17), there is no information about the date of acquisition. As a result, the option was to exclude them, since they were considered occasional expenses of the consumption units and their inclusion would introduce a distortion in the welfare analysis.

A critical data review was made in order to verify if machines, equipment and household utilities acquired during the survey period (POF bloc 15) were already part of the inventory (POF bloc 14). It concluded that the number of goods in stock was higher than the number of durable goods acquired for all consumption units. Therefore, the goods of POF bloc 15 could be excluded without losing information on durable goods' stock.

e) *User Cost*

After gathering all variables, the user cost of each durable good was calculated, according to the formula below:

$$UC_{tij} = S_{tij} P_{itUF_j} \left(r_t - \pi_{tj} + \frac{1}{\bar{L}_i} \right) \tag{6}$$

where $S_{tij}P_{itUF_j}$ is the quantity of durable good i multiplied by its median price in the Federative Unit[7] where the consumption unit j is located; $r_t - \pi_{tj}$ is the regional real interest rate; and $\bar{L}_i = 2\bar{T}_i$. The results originated for the durable goods user cost per Federative Unit are available in Appendix 3 (Supporting information 4).

1.3. *Housing*

The housing group had the biggest participation in the expenditure of Brazilian consumption units across all income classes. For this reason, this group has relevance for the welfare analysis. Items related to housing of the main household were classified in seven types of expenditure, which are: rent, public services, household refurbishments, furniture and household goods, electrical appliances, electrical appliances repairs and cleaning material.

Rental expenses were totally included. The inclusion of paid rent does not distort the comparability between the consumption units, because POF investigated, for residence-owned households, the estimated value of the amount that they would have to pay in case they were renting it. Thus, families that own their estates were not measured with lower welfare in comparison with the ones that pay rent.

Deaton and Zaidi (2002) also recommended including public services expenses (water, sewage treatment, electricity, etc.) in the consumption aggregate. These services add welfare to the consumption units. The inclusion or not of the items related to household refurbishment relates to the possibility of finding if these expenses aggregate value to the household or not. In POF 2008–2009, household maintenance expenses were considered for a period of 90 days and construction expenses for a period of 12 months, but the later ended up being

[7]Remember that for some durable goods it was not possible to calculate the median price by Federative Unit for lack of information about the acquisition of the referred good. In these cases, the median price of the corresponding Major Region was used.

excluded from the consumption aggregate since it aggregates value to the household. All expenditures with cleaning material were included because they are recurrent expenses and increase the consumption units' welfare.

1.4. *Education, Health and Transportation*

Healthcare expenses do not allow adequate measurement of welfare loss and gain associated to them, once the healthcare expenses do not necessarily generate welfare gains—they can be mere ways of minimizing welfare losses. For example, high healthcare expenditure on terminally ill patients cannot be compared to surgery or treatment expenditures that contribute to recovering a patient, or even to an aesthetic-cosmetic procedure.

Education expenditure may cause distortion due to the consumption unit age structure, because it could be seen as an investment that usually occurs at the beginning of a person's life cycle.

According to Lanjow (2009) and Deaton and Zaidi (2002), the decision to include healthcare and education expenses should be considered in cases where these expenses' elasticity (in relation to the total expenditure) is above one. As pointed out in Deaton and Zaide (2002), this procedure is similar to the suggestion presented in Lanjouw and Lanjouw (2001) to deal with measurement errors in expenditures (or welfare ranks). Thus, in order to decide about the inclusion or exclusion of these items, an analysis of these expenditure elasticities was done.

As it can be observed in Table 1, education expenditure elasticity is above one, justifying the total inclusion of these expenses in the aggregate (POF bloc 49). However, the healthcare elasticity is 0.92, requiring a more detailed analysis of elasticity to decide on its inclusion or exclusion. Those results are in line with Lanjow (2009), who estimated elasticities using data from POF 2002–2003.

Considering the low values of elasticity of healthcare expenditure in all income classes (Table 2), the decision was for including solely the healthcare and dental insurance contracts (POF bloc 42) due to their characteristic of providing welfare to the consumption units that access these services. Furthermore, these expenses are responsible for a significant proportion of the consumption units' current expenses.

Nordhaus and Tobin (1972) suggested the exclusion of commute expenses alleging that such expenses do not directly contribute to welfare (utility), being "regrettably necessary inputs" to other activities that generate welfare. This point is also mentioned in Stiglitz, Sen and Fitoussi (2009). Deaton and Zaidi (2002)

TABLE 1

Elasticity of Health and Education Expenditure

Variable	Elasticity	Standard Error	t Value	P-value
Education * Expenses	1.20	0.0200	59.96	<0.0001
Education * Income	1.05	0.0203	52.04	<0.0001
Health * Expenses	0.92	0.0122	75.51	<0.0001
Health * Income	0.83	0.0110	75.48	<0.0001

Source: IBGE, Research Directory, Brazilian Family Expenditure Survey POF -2008-2009.
Note: For the calculation, the sampling design of the survey was considered.

TABLE 2
ELASTICITY OF HEALTH EXPENDITURE VERSUS TOTAL EXPENSES BY DECILES OF THE *PER CAPITA* INCOME DISTRIBUTION

Deciles of income	Elasticity	Standard Error	t Value	P-value
1°	0.75	0.0430	17.37	<0.0001
2°	0.75	0.0428	16.71	<0.0001
3°	0.72	0.0381	18.85	<0.0001
4°	0.75	0.0369	20.4	<0.0001
5°	0.68	0.0339	19.92	<0.0001
6°	0.63	0.0411	15.23	<0.0001
7°	0.67	0.0445	15.01	<0.0001
8°	0.83	0.0505	16.38	<0.0001
9°	0.77	0.0416	18.46	<0.0001
10°	0.73	0.0388	18.84	<0.0001

Source: IBGE, Research Directory, Brazilian Family Expenditure Survey POF 2008–2009.
Note: For the calculation, the sampling design of the survey was considered.

argued that the choice of an expenditure (or a high level of expenditure) eliminates the "regrettable necessity" characteristic and reveals individual preferences. Deaton and Zaidi (2002) suggested the inclusion of such expenses to the extent that it is represented consumption and welfare for those who, by choice, live in nice places away from work ("pleasant suburb"). However, this example does not fit the Brazilian reality in most cases, since the remote locations are often in need of many services.

Based on the arguments of those authors and the information in POF bloc 23, we chose to exclude costs of mass transportation (bus, subway, train, ferryboat, alternative transportation and their connection), often used to go and return from work and often higher for those who live far away. The low income elasticity of expenditure in transportation (0.61) also suggests that not all transportation expenses must be considered in the calculation of consumption. We included expenditure on one's own vehicle (fuel, parking, toll, and carwash), taxi, airplane and car rental because, to some extent, those expenditures reflect individual choices and preferences.

Travel expenses (POF bloc 41) that are not motivated by business and professional reasons or health treatment were included in the aggregate. This kind of information allows us to consider the transportation expenses on leisure and to perform a differentiation of consumption units through luxury goods expenses.

1.5. *Other Non-food Goods*

This group aggregates expenses related to clothing (except for the item wedding dress), culture and leisure, personal services (manicurist, pedicurist, barber, hairdresser etc.), hygiene and personal care, smoking habits and other miscellaneous expenses. Among the miscellaneous expenses, we considered expenses with other properties, parties, communication and professional services, such as registry office, lawyer and forwarding agents. According to Haughton and Khandker (2009), wedding and funeral expenses should not be considered in the consumption aggregate, as well as infrequent and expensive acquisitions. Deaton and Zaidi

(2002) have the same reading on the exclusion of these items. From these, expenses related to ceremonies and parties were excluded due to their occasional character and high values, and expenses with tickets for parties or social events were included. The same was done with expenses related to games and professional services.

Frequent expenses with utilities (such as light, water, sewage, condominium fees, parking spaces fees, etc.) related to other properties of the consumption unit and used for their own benefit (summer house, as an example) were included, while taxes, social contributions, pensions, allowances, donations to other households and private social security taxes were excluded. The banking expenses were included in the consumption aggregate except for the overdraft banking services and credit card expenses.

1.6. *Deflator*

Aiming to ensure the comparability of the consumption aggregate among different geographical spaces and price patterns, in the same period of time, a deflator was calculated using data from the consumption units. In this calculation, we used the consumption units with incomes between the 2nd and the 5th deciles as suggested by Ferreira, Lanjouw and Neri (2000). Excluding those consumption units from the range made consumption baskets more homogeneous, preventing the luxury goods, with low frequency, or goods with excessive quantities interfering in the analysis.

The rationale was to create typical and comparable consumption baskets for each analyzed geographical areas. To do so, we considered a subset of the consumption aggregate items found in all areas, where only the essential expenses for the consumption units were selected for the deflator calculation: electric power, water and sewage, gas and communication (landline phone, mobile phone, paid TV and internet); housing expenses (rent and condominium); food expenses; personal hygiene; cleaning material; and home maintenance.[8]

For the spatial price analysis the choice was to use geographical contexts instead of Federative Units. Studying prices behavior through geographical contexts minimizes distortions caused by regional characteristics. As a result, according to POF sampling design particularities, it was possible to have results for the following geographical strata: Metropolitan Areas (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba and Porto Alegre) and Federal District; non-metropolitan Urban Area and Rural Areas of each Major Region.

The information of the consumption baskets and prices allowed us to calculate spatial index for each geographical context in many ways. For example, one could use a Laspeyres-like price index, as in Ferreira, Lanjouw and

[8]There is no data available for communication services quantity. Thus, the ratio between the total number of people in consumption units having expenses on communication services and the consumption unit total, by geographical area, was used to calculate an average quantity, since it is common in Brazil to contract a package of those services for all the family members. The decision not to include information on estimated rent in the housing category in the referred consumption basket is due to the fact that further study is needed in order to use it in the deflator.

Neri (2000), or a Paasche index, as suggested by Deaton and Zaidi (2002). Others could follow the purching power parities (PPP) literature that uses, for example, a combination of Fisher index (in the Èltetö-Köves-Szulc method) or a Paasche-like index (in the Geary-Khamis method) to account for spatial prices differences.[9]

We calculated a Paasche spatial price index and used it in our analysis. Appendix 4 (Supporting information 5) shows the Paasche index for each Geographical Context. We also calculated and provided (Deflator codes.xlsx, Supporting information 1) a Laspeyres index similar to Ferreira, Lanjouw and Neri (2000) and a Fisher index taken as the geometric mean of those two indexes. Those indexes had a similar behavior.

2. Analysis of Social Welfare and Inequality Based on Aggregated Consumption

The social welfare functions are usually defined in terms of utilities or in terms of the value of consumption (or income). The social welfare functions that become the sum or the average of individual utilities are called utilitarian. In this section, we work, at first, with the Generalized Lorenz Curve, which allows, in some cases, ranking social welfare of an extensive class of functions (in this case, strictly S-concaves and increasing functions).[10] That is to say, one assumes that the social welfare ascends due to the growth of consumption and Pigou-Dalton progressive transfers (when consumption (income) is transferred from a richer to a poor person without turning the poor richer than the one who transferred the consumption). Thus, the Generalized Lorenz Curve (GLC) will point the social welfare in three Geographic Areas (Metropolitan area and Federal District, Urban Area and Rural Area) and the Major Regions without the need to define a specific social welfare function.

In the second step of the analysis, one also assumes that the social welfare function is homogenous at level 1 (or that there is a monotonous transformation that makes it homogeneous at level 1). For this reason, it was possible to obtain (abbreviated) functions that show the effects of inequality towards social welfare. This analysis was based on the Sen mean and the geometric mean and also on their relations with Gini and Atkinson indexes for inequality.

Once the loss of welfare due to inequality is described, the following constitutes a study of inequality by components of the consumption aggregate, using the Gini index, and by subgroup of the population, through mean logarithmic deviation.

[9]A first presentation on life cost indexes can be found in Barbosa (1985). As Deaton and Zaidi (2002) stated, deflating consumption expenditures by Laypeyres index yields an approximation of the welfare ratio whereas defacting by Paasche index yields an approximation of the money metric utility. Blackorby and Donaldson (1987) and Ravallion (1998) suggested the use of the welfare ratio whereas Deaton and Zaidi suggested the use of the money metric utility in welfare analysis. The Geary-Khamis and Èltetö-Köves-Szulc methods can be found in Ackland, Dowric and Freynes (2007) or in the consumer price index manual (ILO 2004/2010).

[10]The $W(X_n)$ function is strictly S-concave when $W(X_n.A_{nxn})>W(X)$ for any X_n that belongs to its domain and any matrix (A_{nxn}) with non-negative elements, having 1 as each line total and 1 as each column total. See Chakravarty (2009).

2.1. *Generalized Lorenz Curve*

The GLC shows the population share (ordered from poorest to richest) on the horizontal axis and shows the consumption partial mean times the population share on the vertical axis. When the curve of an area is always above the other—as it occurs to the Metropolitan Area in Figure 1—it is noticed that there is Generalized Lorenz dominance (Shorrocks, 1983; Chakravarty, 2009; Foster *et al.*, 2013,). The Metropolitan Area dominates the Urban Area and the latter dominates the Rural. The conclusion is that any social welfare function that respects the criteria defined above will maintain the social welfare hierarchy: higher welfare in the Metropolitan area, then in the Urban Area, lastly in the Rural Area.

In the analysis by Major Regions (Figure 2), the following welfare hierarchy is seen: South, Southeast, Midwest, North and Northeast. It is noticeable that the GLC of Midwest is more similar to the GLC of Brazil, which reflects a similar distribution in terms of consumption.

The GLC is important to establish the welfare hierarchy among the Geographical Areas and the Major Regions. However, this analysis does not aim to provide a numerical value to social welfare associated with each Geographical Area or Major Region or to measure the loss of welfare due to inequality. To fill this gap, the following subsections will present two measures that permit the measuring of welfare in terms of inequality and in terms of average consumption, respecting the hierarchy found through the GLC.

2.2. *Welfare and Inequality*

In this section, one assumes that the function of social welfare is homogeneous at level 1 (or that there is a monotonous transformation that makes it homogeneous at level 1). Thus, a proportional increase in the consumption enhanced social welfare equivalently. Consequently, it is possible to obtain (abbreviated) functions that show the effects of inequality on social welfare. This study is based on the Sen mean and on the geometrical mean. More specifically, the Sen mean can be described as the (abbreviated) Sen welfare function that depends on the average of *per capita* consumption and on the Gini index (equation 7). Similarly, the geometric mean can be seen as a (abbreviated) welfare function that depends on the average of *per capita* consumption and on the Atkinson inequality index (equation 8).[11]

$$W_S(c) = \sum_i \sum_j \frac{\min\{c_i, c_j\}}{N^2} = \mu(1 - I_G) \tag{7}$$

[11] W_S can have different motivations. In general, one assumes that the contribution of consumption of one person (family) in social welfare depends on his/her position (or ranking) in the consumption distribution. In some cases, the original welfare function value is identical to the abbreviated function and to the equivalent consumption (Duclos and Abdelkrim, 2006). On this matter, see also Sen and Foster (1997) and Lambert (2001). W_G can be motivated by a logarithmic utility function and a social welfare function that consider the average of the utilities. A monotonous transformation (the exponential of this function) generated the geometric average that assures the needed level 1 homogeneity. One needs to highlight that the logarithmic utility function adopted is a particular case of utility function with constant elasticity, as presented in Atkinson (1970). On this matter, see also Lambert (2001) and Duclos and Abdelkrim (2006).

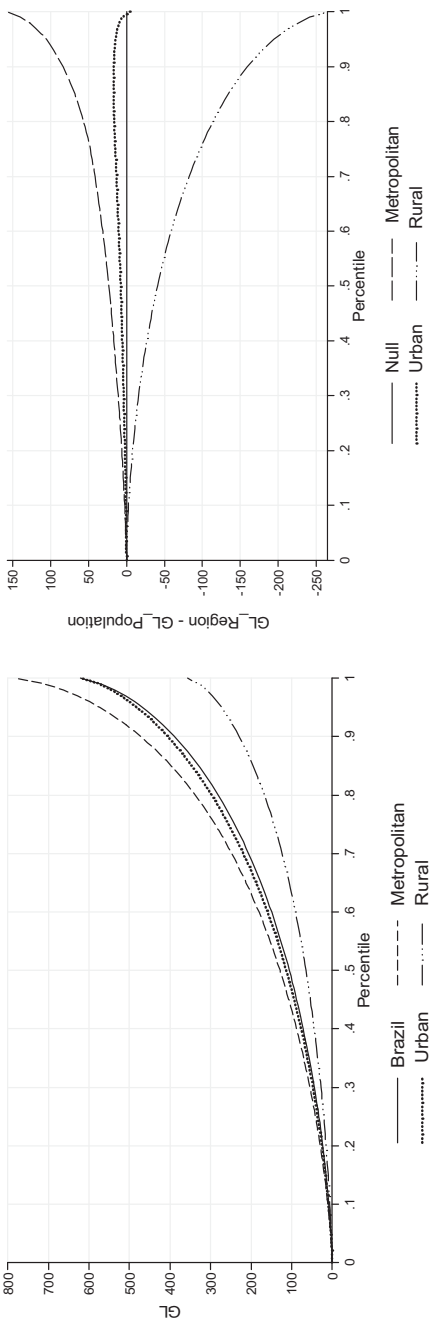


Figure 1. Generalized Lorenz Curve and Generalized Lorenz Curve Differences (Area - Brazil) by Geographical Areas

Source: IBGE, Research Directory, Brazilian Family Expenditure Survey POF -2008–2009.

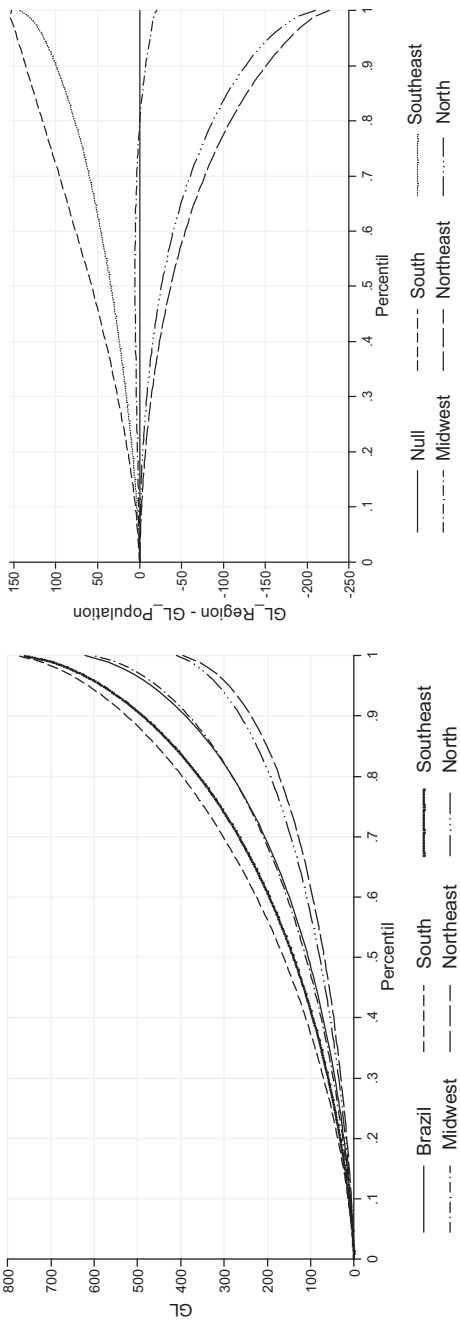


Figure 2. Generalized Lorenz Curve and Generalized Lorenz Curve Differences (Region – Brazil) by Major Regions

Source: IBGE, Research Directory, Brazilian Family Expenditure Survey POF -2008-2009.

TABLE 3

MEAN *PER CAPITA* CONSUMPTION, WELFARE FUNCTIONS AND INEQUALITY INDEXES, BY GEOGRAPHICAL AREAS AND MAJOR REGIONS

Geographical Areas	Mean (μ)	IG	IA	WS(c)	WG(c)
Metropolitan	777.45	0.5149	0.3752	377.14	485.75
	(21.45)	(0.0074)	(0.0091)	(5.78)	(7.09)
Urban	616.72	0.4714	0.3274	326.00	414.81
	(8.72)	(0.0038)	(0.0044)	(2.21)	(2.63)
Rural	356.02	0.4802	0.3329	185.06	237.5
	(6.69)	(0.0054)	(0.0064)	(2.05)	(2.31)
North	410.59	0.4696	0.3142	217.78	281.58
	(11.83)	(0.0078)	(0.009)	(3.21)	(3.71)
Northeast	394.66	0.5075	0.3609	194.37	252.23
	(8.57)	(0.0066)	(0.0079)	(2.63)	(3.13)
Southeast	763.6	0.4808	0.3361	396.46	506.95
	(17.2)	(0.0065)	(0.0077)	(4.94)	(5.89)
South	773.75	0.4333	0.2816	439.49	555.86
	(15.3)	(0.0057)	(0.0064)	(4.44)	(4.99)
Midwest	600.52	0.4805	0.3319	311.397	401.21
	(17.4)	(0.0087)	(0.0101)	(5.30)	(6.11)
Brazil	620.83	0.5010	0.3626	309.79	395.72
	(8.12)	(0.0037)	(0.0045)	(2.28)	(2.73)

Source: IBGE, Research Directory, Brazilian Family Expenditure Survey POF -2008-2009.

$$W_G(c) = \left(\prod_i c_i\right)^{1/N} = \mu(1-I_A) \tag{8}$$

where c_i is the consumption of the individual i; c_j is the consumption of individual j; N is the total population; I_G is the Gini index; I_A is the Atkinson index for inequality; and μ is the average of the *per capita* consumption.

Table 3 shows the values of μ, I_G, I_A, W_S and W_G. As we can see, both the average of consumption (μ) and the welfare measures (W_S and W_G) rank the geographical areas equally. Moreover, as expected, the values of W_S and W_G are lower than the μ in all these areas. This difference represents the loss of social welfare attributed to the inequality in the consumption. For Brazil as a whole, the I_G and the Sen measuring (W_S) both indicate that half of welfare is lost due to inequality of consumption. The Atkinson measure (I_A) and the geometrical measure (W_G) indicate a loss of 36.0 percent. Another way of saying this is that the social welfare would be unchanged if the consumption of families was reduced in 36.0 percent as long as it was distributed equally.

The other areas in Table 3 show similar results. Welfare losses between 43.0 percent and 51.0 percent by the W_S function and between 28.0 percent and 38.0 percent by the W_G function.

Given the impact of social welfare inequalities, the following subsections will present two decompositions: the first one by consumption aggregate components; and the second one by subgroups of the population.

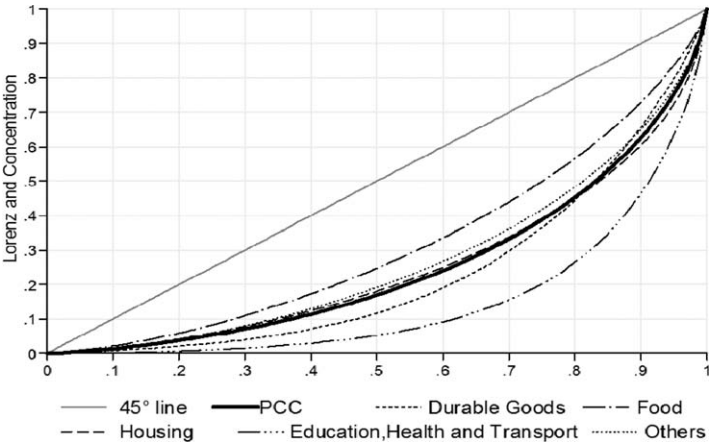


Figure 3. Concentration and Lorenz Curves, by component of consumption, Brazil

Source: IBGE, Research Directory, Brazilian Family Expenditure Survey POF -2008-2009.
Note: PCC = *Per capita* consumption.

2.3. *Decomposition of Inequality by Component of Consumption*

The decomposition of inequality by component of consumption is based on the fact that the Gini index is the result of the concentration of each component of consumption and of the participation of these components in the total consumption. As a result, it is possible to find out which are the factors with higher contribution to the level of inequality.

Figure 3 shows the concentration curves of the five components used in the construction of the consumption aggregate. The further the curve is from the 45° line, the more concentrated the component in analysis will be. Therefore, the biggest concentrations were in the consumption of the groups "Education, health and transportation" and "Durable goods". Remember that POF has no information about the value of public education services or public health services. If those services were taken into account, the group "Education, health and transportation" would be less concentrated. The food group presented the lowest concentration and that is a coherent result since food consumption is vital to living conditions.

In Table 4, we see the results of the consumption aggregate decomposition with data for Brazil. The product of the "consumption group participation" in the consumption aggregate and its corresponding "concentration index" indicate the contribution of each component to inequality.[12] It is noticeable that the housing group had the highest participation in the total consumption, 32.2 percent. It also had a high concentration (0.50), which makes this group the main responsible for inequality with a relative contribution of 32 percent. The group "Education, health and transportation" was the most concentrated of all the

[12]A concentration index could also be broken in two parts: a "gini correlation index" and a "gini of the consumption component" so the product of the "consumption group share" and those two parts yields the contribution of the consumption group to the total inequality. This other decomposition can be found in Lerman and Yitzhaki (1985).

TABLE 4

INDEX-DECOMPOSITION BY COMPONENT OF CONSUMPTION, BRAZIL

Consumption Group	Consumption Share	Gini Correlation	Gini (Component)	Concentration	Contribution	Relative Contribution
Durable Goods	0.1359	0.831	0.6479	0.5384	0.0732	0.1461
	(0.0011)	(0.003)	(0.0024)	(0.0033)	(0.0007)	(0.0019)
Housing	0.325	0.902	0.5523	0.4982	0.1619	0.3232
	(0.0025)	(0.0028)	(0.0053)	(0.0062)	(0.003)	(0.0046)
Education Health and Transport	0.1468	0.9044	0.7782	0.7037	0.1033	0.2062
	(0.0021)	(0.0026)	(0.0032)	(0.0044)	(0.0019)	(0.0031)
Food	0.2178	0.7664	0.4841	0.0371	0.0808	0.1612
	(0.0019)	(0.0043)	(0.0027)	(0.0037)	(0.0011)	(0.0023)
Others	0.1745	0.8422	0.5566	0.4688	0.0818	0.1633
	(0.0013)	(0.0034)	(0.0037)	(0.0046)	(0.0012)	(0.0022)
Total	1.0000	-	-	-	0.5010	1.0000
	(0.0000)	-	-	-	(0.0037)	(0.0000)

Source: IBGE, Research Directory, Brazilian Family Expenditure Survey POF 2008–2009.

components and its concentration index reached 0.7. However, its relative share in total consumption was small, 14.7 percent, making its relative contribution to inequality not the greatest.

2.4. *Decomposition of Inequality by Population Subgroup*

This subsection advances on the study of the decomposition of inequality by geographical area and by characteristics of the person responsible for the consumption unit: years of education, sex, and color or race. However, for the analysis of population subgroups, the Gini index was not used, since it is not decomposable by subgroups in a way that one gets only the sum of the within inequality and between inequality of the subgroups studied.

Thus, the decomposition by subgroups was made based on the mean logarithmic deviation ($ln(\mu/W_G)$). This index belongs to the class of Generalized Entropy, closely associated with the Atkinson measure of inequality. In the case of the mean logarithmic deviation index, this can be described as the sum of inequality within each subgroup of the population, weighted by the share of each subgroup, plus the existing inequality between the subgroups, see Lambert (2001) and Cowell (2000) for further details on this index.

As we can see in Appendix 5 (Supporting information 6), the inequality calculated by the mean logarithmic deviation for Geographical Areas presented results similar to the level of Brazil (0.45), being of 0.47 in the Metropolitan Region, 0.40 in Rural Areas, and 0.40 in Urban Areas. However, by having a greater number of inhabitants (53.0 percent), the Urban Area, even with a lower level of inequality among the Geographical Areas, had a greater relative contribution (46.7 percent). Concerning the Major Regions, the Southeast had the highest share of population and also a high level of inequality (0.41). It may be noted in this subgroup the cases of the Midwest and North regions, which have the smallest population rates

(7.3 percent and 8.0 percent, respectively), but a level of inequality rather high (0.40 and 0.38, respectively).

Regarding the subgroup color or race, whites are responsible for the higher relative incidence of inequality, 41.4 percent. Regarding the sex subgroup, the mean logarithmic deviation is very similar for men (0.45) and women (0.45), i.e. the sex of the person responsible for the consumption unit did not affect the consumption inequality. Therefore, the higher relative contribution of men to total inequality is determined by their greater participation (72.4 percent) in the total number of persons responsible for consumption units.

Through investigation of the results of the subgroup's years of education, it is clear that there was an inverse association between the number of years of education and the level of inequality, and for the grades 8–10 years and 11–14 years the mean logarithmic deviation was stable and fell back to people with 15 years or more of study.

Another way to examine the results of the mean logarithmic deviation was to look at the interaction of inequality of subgroups with their weight in the population and the total inequality in the country. As shown in Appendix 5, the main contribution to inequality in Brazil came from differences within each subgroup. That is, although there was a Generalized Lorenz dominance of the Metropolitan Area over the Urban and Rural Areas (see Subsection 2.1), 93.3 percent of Brazil's inequality was explained by inequality within these subgroups. The same could be observed for subgroups of the Major Regions (90.1 percent), and sex (99.9 percent). As it occurs in other studies, for example Ferreira *et al*. (2006), most inequality comes from the differences within groups and not from the differences between subgroups.

In the case of years of education and color or race, despite the importance of the within groups inequality, the inequality between groups is, respectively, 31.5 percent and 12.4 percent. This makes years of education extremely relevant to explain Brazil's inequality.

The population could be divided into five subgroups, defined by the number of services to which the consumption units (and its members) have access. More specifically, we considered access to the following services: "water supply system", "sewage system", "electric supply system" and "direct refuse collection". It was observed that a fraction of the population (1.8 percent) did not have any of these services. A greater proportion (10.2 percent of the population) had only one of four services, usually electricity services (97.8 percent of the cases); 11.9 percent of the population had only two services, the most frequent pairs were electricity and water (49.9 percent) and electricity and refuse collection (48.3 percent). The subgroup with three services corresponded to 30.4 percent of the population and the most frequent trio of services were electricity, water and garbage collection (86.6 percent of cases). Only 45.6 percent of the population had all four specified services.

For this reason, within each of those five subgroups, there was little variability in access to the four services selected. This picture contrasted with the existing consumption inequalities within the same subgroups that ranged from 0.35 to 0.41 and contributed with 88.1 percent of the total consumption inequality.

3. Analysis of Poverty and Vulnerability Based on Aggregated Consumption

In this section, the consumption units are analyzed by Geographical Areas and subgroups of population regarding poverty. In this sense, poverty measures that belong to the FGT (Foster, Greer and Thorbecke, 1984) class were used with focus on poverty severity (FGT(2)). In order to estimate the probability of a consumption unit becoming poor in a future period of time, a vulnerability analysis based on Chaudhuri *et al.* (2002) methodology adapted to incorporate cluster effects was done.

3.1. *Poverty Severity*

The consumption aggregate can also be used in studies of poverty from a monetary perspective. Following this perspective, different dimensions (food, housing, education, health, transportation, leisure etc.) were combined considering available prices and expenditure types as described in previous sections. After calculating the consumption, two extra exercises were necessary in order to evaluate poverty, the Identification and Aggregation exercises, emphasized by Sen (1976, 1982). Besides the emphasis on these exercises, Sen's studies comment the limitation concerning the most used poverty measures at the time (that did not take inequality among the poor into account) and stimulates an axiomatic approach in which poverty indexes are constructed to attend some properties and evaluated by those properties. The Identification itemizes the poor and the non-poor, while the Aggregation enables the combination of information about poverty in an index.

In general, the poor identification was based on some poverty line (z) that marked a limit to the welfare indicator (in this case, consumption). The poor were identified by the welfare indicator (consumption) that is below the line. The non-poor were identified by the indicator (consumption) that is higher or equal regarding the poverty line.[13] In this work, two absolute lines were adopted based on minimum wage. Consumption units with *per capita* income next to half of minimum wage (between R$202.50 and R$212.50) and a quarter of a minimum wage (between R$101.25 and R$106.25) were adopted. Then, the median *per capita* consumption of these two groups was calculated resulting in two poverty lines based in consumption: R$185.00 and R$117.00.

In Figure 4, the proportions of the poor in Brazil are shown for the Geographical Areas and Major Regions according to different poverty lines (R$1.00 $\leq z \leq$ R$200.00). This helps one visualize how sensitive the Identification exercise (of the poor) was towards the chosen lines. The inclination of these curves around the lines R$185.00 and R$117.00 indicates this sensitivity. As it can be seen, the sensitivity was higher in the Rural Area and in the North and Northeast regions. Even so, around these two lines, we can see a clear hierarchy within the Geographical Areas and within the Major Regions. That is to say that for an extensive set of lines next to R$185.00 and R$117.00 the proportion of the poor was higher

[13]Further details about different methodologies, definitions and interpretations regarding absolute, relative and subjective poverty lines can be seen in Ravallion (2001), Atkinson *et al.* (2002) and Soares (2009).

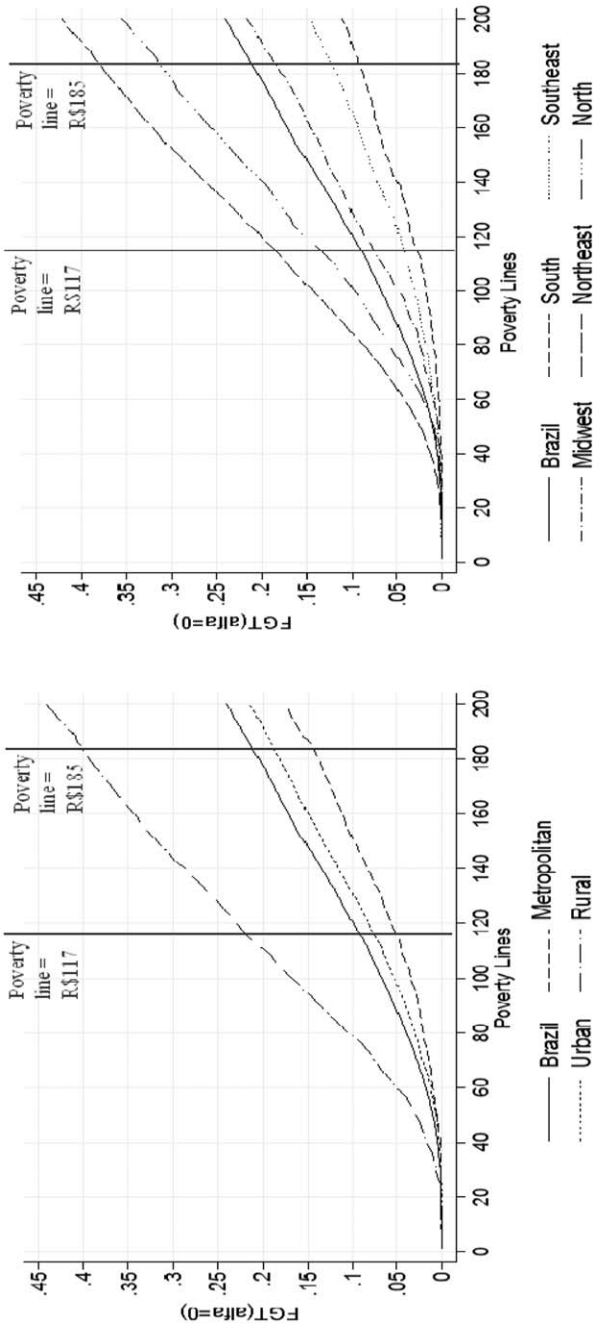


Figure 4. Poverty Curves by Geographical Area and Major Regions

Source: Brazilian Family Expenditure Survey POF 2008–2009, IBGE.

in the Rural Area followed by the Urban Area. Similarly, for an extensive set of poverty lines next to R$185.00 and R$117.00 a bigger proportion of the population was classified as poor in the North and Northeast and a smaller proportion in the South and Southeast.

Once the lines are selected, we move to the exercise of Aggregation, in which the information about the poor is combined to analyze poverty in society. Three measures of the FGT family (Foster, Geer and Thorbecke, 1984) were used to study poverty: the proportion or incidence of the poor [FGT ($\alpha=0$)], poverty intensity [FGT ($\alpha=1$)] and poverty severity [FGT ($\alpha=2$)], as defined in the expression below:

$$FGT(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{z-c_i}{z}\right]^{\alpha} S_i \tag{9}$$

where z is the poverty line value and S_i is a dummy variable that is equal to 1 if the individual is below poverty line and 0 otherwise. The bigger the coefficient α is, the bigger the poverty gap is. The values of these measures for Brazil and Geographical Areas are presented in Appendix 6 (Supporting information 7).

Among the three measures presented, only the one concerning poverty severity takes account of consumption inequality among the poor. That is to say, keeping the total amount consumed by the poor, the more heterogeneous the poor population is, the bigger the value of the indicator FGT(2). That being said, this is the most appropriate poverty measurement, and that will be analyzed in Appendix 7 (Supporting information 8).

Taking poverty severity by Geographical Area into consideration for the line R$185.00, it was noticeable that poverty is more severe in the Rural Area (0.084). However, the Urban Area presented the biggest relative contribution because of the weight of its population. When the same evaluation was done with the line R$117.00, it was noticeable that the biggest contribution in terms of severity comes from the Rural Area (45.2 percent), while the Urban Area represented 39.1 percent.

When poverty severity is analyzed by Major Regions taking the two poverty lines into account, the Northeast presented the biggest poverty level, followed by the North. When one observes the relative contribution, the Northeast still had the biggest participation.

In the subgroup related to color or race, it was recognizable that the black or mixed population subgroups were the ones that most contribute to the severity of poverty, followed by the white subgroup.

The results in Appendix 7 also show that poverty is more severe for people with no school education or with only a few years of education (0–7 years), having a relative contribution of 87.2 percent (referring to line R$185.00) and 88.7 percent (referring to line R$117.00) concerning poverty severity. In relation to sex, there were no observed differences concerning poverty severity: referring to line R$185.00, women had the index of 0.037 while men had 0.036; referring to line R$117.00, the values were 0.013 and 0.012, in the same order.

Considering again the services “water supply system”, “sewage system”, “electric supply system” and “direct refuse collection”, and separating the population into five subgroups (defined by the number of services to which the CUs

and their members had access) there was a clear association between the severity of poverty and the number of services accessed. In this case, the subgroup without access to any of the services was also the one that suffered more severely from poverty. The lowest poverty rates were reported precisely by those with all four services. Together subgroups with access to only one or none of the services were 12 percent of the population but accounted for about 33 percent or 40 percent of poverty depending on the poverty line considered (R$185.00 or R$117.00, respectively).

3.2. *Poverty Vulnerability*

Recent studies have emphasized the analysis of poverty vulnerability, understood as the chance of the welfare indicator to present a value below the poverty line. Examples of this analysis can be found in Lopes-Calvas and Ortiz-Juarez (2011), Ferreira *et al*. (2013), SAE's Report (used it for the definition of middle class) and Ribas (2007). A different approach can be found in Calvo and Dercon (2008), which suggested a class of vulnerability index. It is important to highlight that the data set that has a panel form is the most appropriate way to study vulnerability. However, Chaudhuri *et al*. (2002) and Haughton and Khandker (2009) recommended the evaluation of vulnerability even in the absence of data panels on consumption. In those cases, they suggested the use of regressions and generalized least squares estimators to model the consumption distribution and the chance of a consumption unit falling into poverty. In this section, the procedures of Chaudhuri *et al*. (2002) were adapted to include area effects (clusters). Thus, a methodology similar to the one used in the Poverty Map was applied (Elbers *et al*., 2002; IBGE, 2008).[14]

More specifically, the vulnerability of a consumption unit j to poverty in time t is defined as the probability of the *per capita* consumption of that unit in time $t+1$ being below poverty line z:

$$v_{jt} = P\left(c_{j,t+1} < z\right) \tag{10}$$

where $c_{j,t+1}$ is the *per capita* consumption of the consumption unit j in time $t+1$ and z is the poverty line calculated from the consumption aggregate.

One defines y_{dj} as a welfare variable function, being the logarithm of the *per capita* consumption aggregate, of consumption unit j in the enumeration area d. The model can be written as follows:

$$\gamma_{dj} = x_{dj}\beta + \eta_{dj}, \eta_{dj} \sim F(0, \Sigma) \tag{11}$$

meaning F is a distribution with a vector of average 0 and variance-covariance matrix Σ and x_{dj} is the vector of explanatory variables of the sample survey, regarding the consumption unit j of enumeration area d, $j=1,\ldots,N_d$ and $d=1,\ldots,D$. It is possible to introduce indicators on geographical levels that are

[14]Another possibility that might be explored in the future is the use of pseudo-panels as suggested by Bourguignon *et al.* (2006) and Dang and Lanjouw (2013).

more aggregated in order to control the area effect whenever it is not entirely explained through regressors. Such indicators can be obtained in other databases.

The model error may have two components: (i) unit effect associated with consumption unit; and (ii) area effect associated with the enumeration area where this unit is placed. As a result, η_{dj} can be written as:

$$\eta_{dj}=u_d+e_{dj} \tag{12}$$

where u_d and e_{dj} are independent from $u_d \sim N\left(0,\sigma_u^2\right)$ and $e_{dj} \sim N(0,\sigma_{e_{dj}}^2)$.

Assuming that the errors of the unit level e_{dj} are heterocedastics, Elbers *et al.* (2002) suggested estimating the logistic regression:

$$\ln\left(\frac{e_{dj}^2}{A-e_{dj}^2}\right)=z'_{dj}\alpha+r_{dj} \tag{13}$$

and they estimated the variance on consumption unit level according to the formula:

$$\hat{\sigma}_{e_{dj}}^2 \approx \left[\frac{AB}{1+B}\right]+\frac{1}{2}var(r)\left[\frac{AB(1-B)}{(1+B)^3}\right] \tag{14}$$

where $A=1.05\max(e_{dj}^2)$; $B=\exp(z'_{dj}\hat{\alpha})$, $var(r)$ is the quadratic error of the estimated logistic regression's residual; and z_{dj} is a vector of explanatory variables.

Assuming that u_d and e_{dj} have a normal distribution, Elbert *et al.* (2002) derived an estimate of the area effect variance u_d:

$$var\left(\hat{\sigma}_u^2\right) \approx \sum\nolimits_d 2\left\{a_d^2\left[\left(\hat{\sigma}_u^2\right)^2+\left(\hat{\tau}_d^2\right)^2+2\hat{\sigma}_u^2\hat{\tau}_d^2\right]+b_d^2\frac{\left(\hat{\tau}_d^2\right)^2}{n_d-1}\right\} \tag{15}$$

where $\hat{\tau}_d^2=\frac{1}{n_d(n_d-1)}\sum_j\left(e_{dj}-e_{d.}\right)^2$, $e_{d.}=\frac{1}{n_d}\sum_j e_{dj}$, $a_d=\frac{w_d}{\sum_j w_j\left(1-w_j\right)}$, $b_d=\frac{w_d(1-w_d)}{\sum_j w_j\left(1-w_j\right)}$, $\hat{\sigma}_u^2=\sum_d a_d\eta_{d.}^2-\sum_d b_d\hat{\tau}_d^2$, $\eta_{d.}^2=u_d+e_{d.}$, $w_d=\sum_j\frac{w_{dj}}{n_d}$ and n_d it is the total number of people associated with the enumeration area d and w_j is the survey expansion factor of the consumption unit j.

From the estimates found for the modeling procedure parameters, it is estimated the welfare variable and the vulnerability of each consumption unit according to the equation:

$$\hat{\upsilon}_{dj}=\Phi\left(\frac{\ln z-\hat{y}_{dj}}{\sqrt{\hat{\sigma}_{u_d}^2+\hat{\sigma}_{e_{dj}}^2}}\right) \tag{16}$$

Once the methodology is defined, the following step is the estimation of the consumption units (and their components) vulnerability. As explanatory

variables, data regarding consumption units from POF were used, such as logarithm of available *per capita* monetary income, number of residents per room, logarithm of total number of residents, indicator of bathroom, indicator of the geographical context, some indicators about of the person responsible for the consumption unit such as literacy, occupation, education, some enumeration area data such as the proportion of consumption units whose heads had healthcare insurance and data on a municipal level from other sources, for instance, the logarithm of *per capita* GDP in 2010 and the proportion of people who received *Bolsa Família* in 2010. The significance level concerning the choice of the variables was 0.05. See Appendices 8 and 9 (Supporting information 9 and 10) for a complete list of the explanatory variables of the model and the estimated coefficient vectors α and β.

It is important to mention that the variables selection aim was to select a model with good predictive power without the pretension of presenting causal effects. The model adjustment (equation 11) resulted in a R^2 of 0.75.

According to Chaudhuri *et al*. (2002), the consumption units with vulnerability index higher than 0.5 were considered highly vulnerable. The expected consumption of these consumption units and their components were below the poverty line. It is possible that these consumption units end up suffering from chronic poverty. The consumption units (and their components) with probabilities between 0.2 and 0.5 were classified as vulnerable due to the fact that they presented estimated consumption above the poverty line, but they still have big chances of falling into poverty. It is possible that these consumption units may suffer from transitory poverty.[15]

Figure 5 indicates the relation between income and consumption vulnerability in a synthetic way. More specifically, it indicates the average vulnerability (of *per capita* consumption) by percentiles of *per capita* income and their respective confidence intervals of 95.0 percent. A strict relation is observable between these incomes and the consumption units' vulnerability. Around 16.0 percent of the population presented average vulnerability higher than 0.5 and *per capita* income below R$181.70. Thus, a rather low *per capita* income (less than R$181.70) also indicates high vulnerability (of consumption). Similarly, *per capita* incomes between R$181.70 and R$327.24 can be considered a sign of vulnerability (between 0.2 and 0.5). People who presented estimated vulnerability between 0.2 and 0.5 may face a situation of transitory poverty; they represent 19 percent of the Brazilian population.

Figure 6 indicates the estimated proportion of people vulnerable to poverty according to different vulnerability threshold levels established between zero and one by (a) Metropolitan, Urban and Rural Areas and (b) Major Regions. Among all levels of vulnerability thresholds, the North and the Northeast regions presented bigger estimated proportions of vulnerable people, and the same situation was observed in the Rural Area. The South region line decay is more accentuated than in the other regions.

In conclusion to the regional analysis, Figure 7 indicates the estimated proportion of people vulnerable to poverty versus the estimated proportions of the

[15]A more appropriate explanation concerning chronic poverty, as well as transitory, would also need data panels. See, for example, Ravallion and Jalan (2000), Addison *et al*. (2009) and the Journal of Economic Inequality 10 (2012).

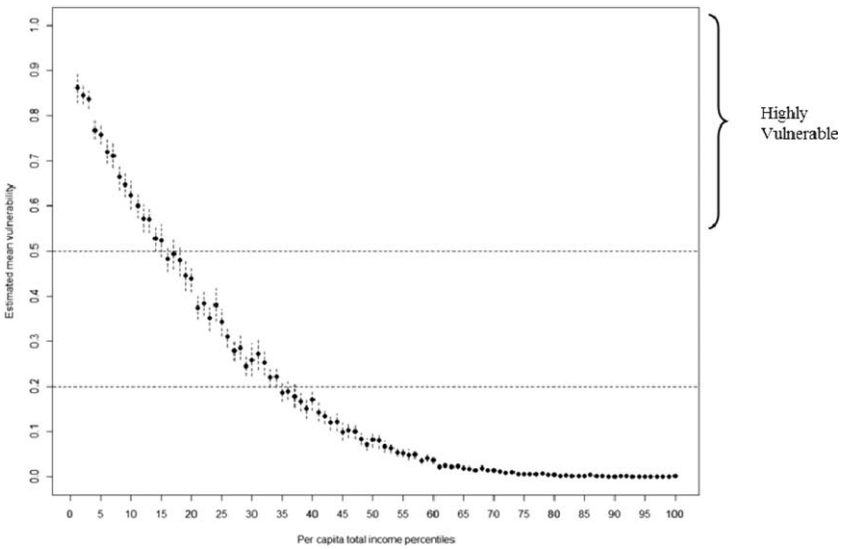


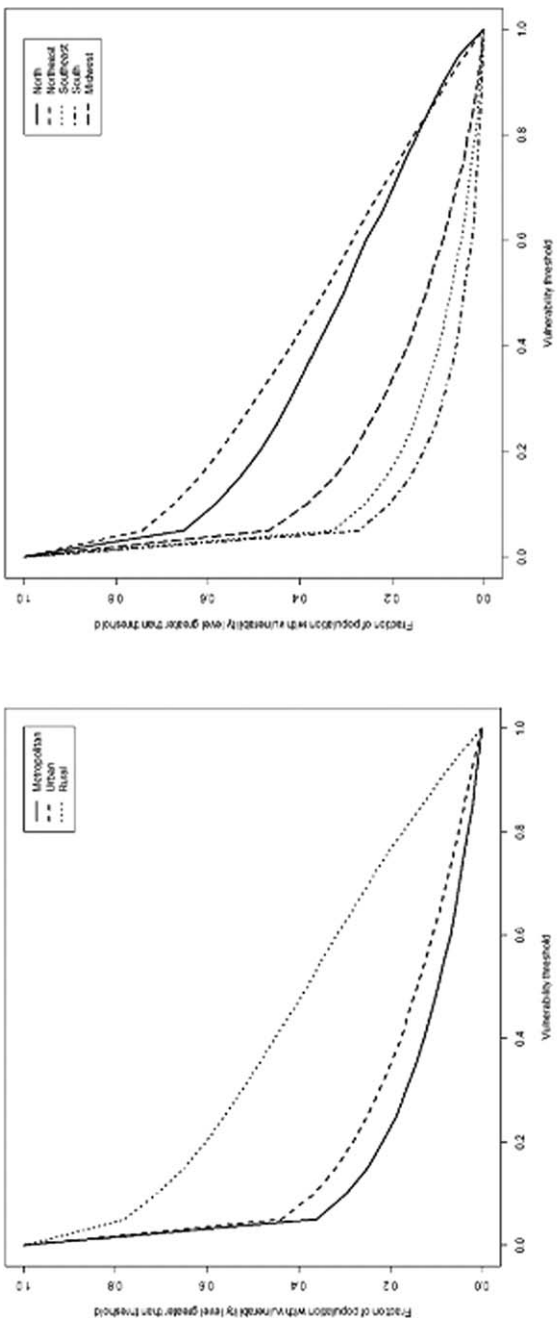
Figure 5. Estimated Average Vulnerability by Percentiles of *Per Capita* Income and their Respective Confidence Intervals of 95%

Source: Brazilian Family Expenditure Survey POF 2008–2009, IBGE.

poor estimated directly in POF for the 20 geographical contexts that compose the Geographical Areas and the Major Regions. Figure 7-a focuses on highly vulnerable people (vulnerability above 0.5), while Figure 7-b focuses on vulnerable people (vulnerability between 0.2 and 0.5). In Figure 7-a, we can observe that the contexts in the North and Northeast regions present proportions of highly vulnerable people, higher than the proportions identified for Brazil (horizontal line)—except for the Belem metropolitan region (represented in one of the Urban Area contexts of the North region). The same occurred with the estimated proportion of poor people in Brazil (vertical line). It is highlighted that, among Rural Areas, the South was the only region that had a proportion of highly vulnerable people, lower than the proportion identified for Brazil. The same tendency could be seen in (b). Therefore, people from the South region and from Urban Areas of Southeast and Midwest regions are less vulnerable to poverty than people from the North and from the Northeast regions.

Two other remarkable factors are related to the 45° straight line in these figures. As it is observable, the proportion of highly vulnerable people was (fairly always) a little smaller than the proportion of the poor, although these two measures behaved the same way. This can be explained due to the fact that the proportion of the poor is a measure of (unconditional) vulnerability of society. In Figure 7-b, the biggest differences could be seen since an increase of the proportion of poor was not always followed by a similar increase in the vulnerability proportion.[16]

[16]See Appendix 10 (Supporting information 11): Estimated proportion of people vulnerable to poverty versus the estimated proportions of the poor calculated directly from POF by the 20 geographical contexts, which take part in Geographical Areas and Major Regions.

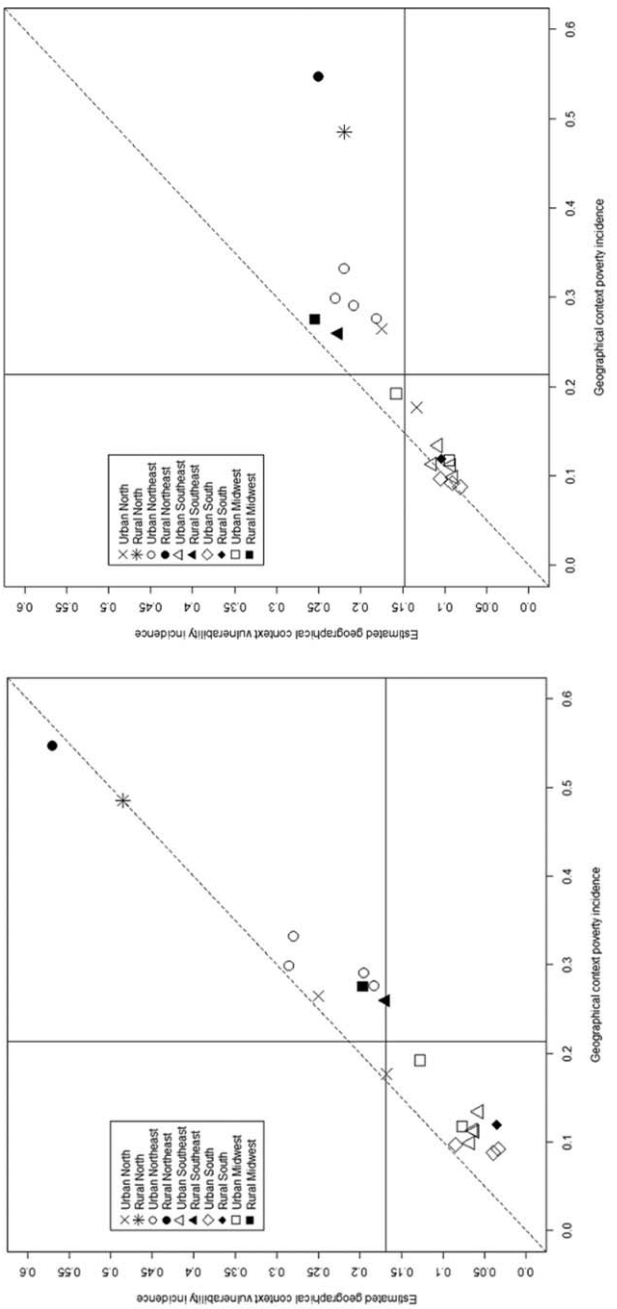


(a) Metropolitans, Urban and Rural Areas

(b) Major Regions

Figure 6. Estimated Proportion of People Vulnerable to Poverty According to Vulnerability Threshold Levels Established Between 0 and 1

Source: Brazilian Family Expenditure Survey POF -2008-2009, IBGE.



(a) Level above 0.5

(b) Level between 0.2 e 0.5

Figure 7. Estimated Proportion of People Vulnerable to Poverty Versus the Estimated Proportions of the Poor Calculated Directly from POF by the 20 Geographical Contexts which take part in Geographical Areas, and in Major Regions

Source: Brazilian Family Expenditure Survey POF 2008–2009, IBGE.

4. Final Considerations

The present article was proposed to examine the welfare, inequality, poverty and vulnerability to poverty of Brazilian families from the perspective of consumption. This issue is commonly addressed through the income perspective. However, the consumption pattern is brought as the most suitable for these studies, since it presents a better response to seasonal fluctuations, demonstrating how families behave according to their budget availability and, thus, better capturing their living conditions.

POF, conducted by IBGE, is the research that raises the information expenses of a range sufficient to determine the consumption pattern of Brazilian families. Nevertheless, the use of information from POF to conduct analysis on welfare, inequality, poverty and vulnerability to poverty can be further investigated. For this reason, we selected no sporadic expenses most likely to represent welfare gains and assigned values to the consumption of durable goods by an estimate of the user costs. The last step of the consumption aggregate consisted of correcting the values obtained by means of a spatial price deflator. As a result, we verified in this paper that through the construction of a consumption aggregate, which reflects multiple dimensions of the families consumption choices, such as food, housing, durable goods, health, education and transportation and other non-food items, it is possible to perform these studies with POF data.

Once defined the consumption aggregate, we advanced with the measurement of welfare, poverty, inequality and vulnerability. To do so, we examined the behavior of the Generalized Lorenz Curve, two (abbreviated) social welfare functions, we calculated Gini and Atkinson measures of inequality and mean logarithmic deviation, then we decomposed inequality by components and by population subgroup. To measure poverty, we analyzed the sensitivity of the exercise of identification to the different poverty lines, and we presented the results of the poverty severity per geographic areas and different population subgroups. Among the subgroups analyzed, we highlight that the difference in years of education had a fundamental role for both inequality and poverty.

Finally, we assessed the vulnerability to poverty, including area effects (clusters) based on the work of Chaudhuri *et al*. (2002) and Elbers *et al*. (2002).

We emphasize that the present paper is part of a larger work of poverty studies based on POF data. Possible extensions of this study are, among other exercises, to apply the evaluation measures and analysis presented here in other geographical divisions, the comparison of indicators over time (2002–2003 and 2008–2009), the creation of pseudo panels to improve the measurement of vulnerability and further study on the poverty lines.

References

Ackland, R., S. Dowrick, and B. Freyens, *Measuring Global Poverty: Why PPP Methods Matter*, mimeo, Research School of Social Sciences, The Australian National University, 2007.

Addison, T., D. Hulme, and R. Kanbur, *Poverty Dynamics: Interdisciplinary Perspectives*, New York, Oxford University Press, 2009.

Atkinson, A. B., "On the Measurement of Poverty," *Journal of Economic Theory*, 2, 244–63, 1970.

———, "Lecture 1. Political Arithmetic: Financial Poverty in the European Union," *Poverty in Europe*, 10–66, YrjöJahnsson Foundation, 1998.

Atkinson, A., B. Cantillon, E. Marlier, and B. Nolan, *Social Indicators: The EU and social inclusion*, New York, Oxford University Press, 2002.
Barbosa, F., *Microeconomia: teoria, modelos econométricos e aplicações à economia brasileira*, Rio de Janeiro, IPEA, 1985.
Blackorby, C. and D. Donaldson, "Welfare Ratios and Distributionally Sensitive Cost-Benefit Analysis," *Journal of Public Economics*, 34, 265–90, 1987.
Bourguignon, F., C. Goh, and D. Kim, "Estimating Individual Vulnerability to Poverty with Pseudo-Panel Data" in S. Morgan, D. Grusky and G. Field(ed.), *Mobility and Inequality*, 349–69, Stanford, Stanford University Press, 2006.
Calvo, C. and S. Dercon, "Risk and Vulnerability to Poverty" in Kakwani, N., Silber, J. (ed.), *The Many Dimensions of Poverty*, 215–28, New York, Palgrave Macmillan, 2008.
Chakravarty, S. "Inequality, Polarization and Poverty: Advances in Distributional Analisys," *Economic Studies in Inequality, Social Exclusion and Well-Being*, New York, Springer, 2009.
Chaudhuri, S., J. Jalan, and A. Suryahadi, "Assessing Household Vulnerability to Poverty from Cross-Sectional Data: A Methodology and Estimates from Indonesia," Discussion Paper 102, New-York, Economics Department, Columbia University, 2002.
Cowell, F. "Measurement of Inequality" in A.B. Atkinson and F. Bourguignon (ed.), *Handbook of Income Distribution*, 87–166, Amsterdam, Elsevier, 2000.
Dang, H-A. and P. Lanjouw, "Measuring Poverty Dynamics With Synthetic Panes Based on Cross-Sections," Policy Research Working Paper 6504, Washington DC, The World Bank, 2013.
Deaton, A., "6. Saving and Consumption Smoothing," *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, 335–400, World Bank, The Johns Hopkins University Press, 1997.
Deaton, A. and S. Zaidi, "Guidelines for Consctructing Consumption Aggregates for Welfare Analysis," Living Standards Measurement Survey Working Paper 135, Washington DC, The World Bank, 2002.
Duclos, J-Y. and A. Abdelkrim, "Poverty and Equity: Measurement, Policy and Estimation with DAD," *Economic Studies in Inequality, Social Exclusion and Well-Being*, New York, Springer, 2006.
Elbers, C., J. O. Lanjouw, and P. Lanjouw, "Micro-level Estimation of Welfare," Policy Research Working Paper no. WPS 2911, Washington DC, The World Bank, 2002.
Ferreira, F. G. H., P. Lanjouw, and M. Neri, "A New Poverty Profile for Brazil Using PPV, Pnad and Census Data," Texto para Discussão, PUC-Rio, 418, 1–66, 2000.
Ferreira, F. G. H., P. Leite, J. A. Litchfield, and G. Ulyssea, "Ascensão e queda da desigualdade de renda no Brasil," in R. Barros, M. Foguel and G. Ulyssea, *Desigualdade de Renda no Brasil*, 357–78, IPEA, 2006.
Ferreira, F. G. H., J. Melissa, J. Rigolini, L. Lopez-Calva, M. Lugo, and R. Vakis, *Economic Mobility and the Rise of the Latin America Middle Class*, Washington DC, The World Bank, 2013.
Foster, J., J. Greer, and E. Thorbecke, "A Class of Decomposable Poverty Measures," *Econometrica*, 52, 761–66, 1984.
Foster, J., S. Seth, M. Lokshin, and Z. Sajaia, *A Unified Approach to Measuring Poverty and Inequality Theory and Practice*, Washington DC, The World Bank, 2013.
Haughton, J., and S. Khandker, *Handbook on Poverty and Inequality*, Washington DC, The World Bank, 2009.
Hentschel, J. and P. Lanjouw, "Constructing an Indicator of Consumption for the Analysis of Poverty: Principles and Illustrations with Reference to Ecuador". LSMS Working Paper – Number 124, Washington DC, The World Bank, 1996.
IBGE, *Mapa de Pobreza e Desigualdade: municípios brasileiros 2003*. Rio de Janeiro, IBGE, 2008 (CD-ROM).
IBGE, *Pesquisa de Orçamentos Familiares 2008–2009: Despesas, Rendimentos e Condições de Vida*, Rio de Janeiro, 2010.
ILO, *Household income and expenditure statistics*, Report II, Seventeenth International Conference of Labour Statisticians, Geneva, December, 2003.
———, *Consumer Price Index Manual: Theory and practice*, 2004 (revisions as at 15/08/2010) http://www.ilo.org/public/english/bureau/stat/guides/cpi/.
Journal of Economic Inequality, *Special Issue: Inequality, New Directions*. 9, 1569–1721 (Print), 1573-8701 (Online), September, 2011.
———, 10, 137–297, May, 2012.
Kakwani, N. and J. Silber (ed.) *The Many Dimensions of Poverty*, New York, Palgrave Macmillan, 2007.
———, (ed.), *Quantitative Approaches to Multidimensional Poverty Measurement*, New York, Palgrave Macmillan, 2008.

Lambert, P., *The Distribution and Redistribution of Income*. New York, Manchester University Press, 3rd ed., 2001.
Lanjouw, J. and P. Lanjouw, "How to Compare Apples and Oranges: Poverty Measurement Based on Different Definitions of Consumption," *Review of Income and Wealth*, 47, 25–42, 2001.
Lanjouw, P., *Constructing a Consumption Aggregate for the Purpose of Welfare Analysis: Principles, Issues and Recommendations Arising from the case of Brazil*, Paris, OECD Conference, 2009.
Lerman, R. and S. Yitzhaki, "Income Inequality Effects by Income Source: A New Approach and Applications to the United States," *The Review of Economics and Statistics*, 67, 151–56, 1985.
Lopez-Calva, L. and E. Ortiz-Juarez, "A Vulnerability Approach to the definition of the Middle Class", Policy Research Working Paper 5902, Washington DC, The World Bank, 2011.
OECD, *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*, 2013.
Oliveira, L. *Três Ensaios Sobre Pobreza Multidimensional*, Rio de Janeiro, UFRJ, Tese de Doutorado, 2010.
Nordhaus W. and J. Tobin, "Is Growth Obsolete", in _____, *Economic Research: Retrospect and Prospect*, 5, 1–80, Economic Growth, NBER, 1972.
Ravallion, M. "Poverty Lines in Theory and Practice". LSMS Working Paper – Number 133. Washington DC, The World Bank, 1998.
———, *Setting Poverty Lines: Economic Foundations of Current Practices*, Washington, Development Research Group, World Bank, 2001.
———, "On multidimensional indices of poverty," *Journal of Economic Inequality*, 9, 235–48, 2011.
Ravallion, M. and J. Jalan, "Is Transient Poverty Different? Evidence for Rural China," *Journal of Development Studies*, 36, 82–99, 2000.
Ribas, R., *Vulnerabilidade à pobreza no Brasil: Medindo Risco e Condicionalidade a partir da Função de Consumo das Famílias*, PPE, 37, 2007.
Rosenbaum, P. R., and D. B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55, 1983.
SAE, *Comissão para Definição da Classe Média no Brasil*, Secretaria de Assuntos Estratégicos, Brasil, 2011.
Sen, A., "Poverty: An Ordinal Approach to Measurement," *Econometrica*, 44, 219–31, 1976.
———, *Poverty and famines an essay on entitlement and deprivation*. Oxford, Claredon Press, 1982.
———, *Desenvolvimento como Liberdade*, São Paulo, Companhia das Letras, 2004.
———, *Desigualdade Reexaminada*, 2ª ed., Rio de Janeiro, Editora Record, 2008.
———, *The Idea of Justice*, Cambridge, Belknap Harvard University Press, 2009.
Sen, A. and J. Foster, "On Economic Inequality after a Quarter Century," *Annex to the enlarger edition of On Economic Inequality* by Sen, A., New York, Oxford University Press, 1997.
Shorrocks A., *Ranking Income Distributions*, Econometrica, 50, 3–17, 1983.
Slesnick, D., *Consumption and Social Welfare: Living Standards and their Distribution in the United States*, Cambridge University Press, 2001.
Soares, S., *Metodologias para estabelecer a linha de pobreza: objetivas, subjetivas, relativas e multidimensionais*, Rio de Janeiro, IPEA, Texto para Discussão 138, 2009.
SRF. *Instrução Normativa da Secretaria da Receita Federal n° 162 de 31 de dezembro de 1998*, disponível em http://www.receita.fazenda.gov.br/Legislacao/ins/Ant2001/1998/in16298.htm, visited on 04/06/2013.
Stiglitz, J. E., A. Sen, and J.P. Fitoussi, *Report by the Commission on the Measurement of Economic Performance and Social Progress*, 2009.
Varian, H., *Microeconomic Analysis*, New York, W. W. Norton & Company, Inc. (3rd edition), 1992.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Supporting information 1: Consumption Aggregate Codes; Database; Deflector Codes
Supporting information 2: Appendix 1: *Per Capita* Consumption Aggregate versus Total *Per Capita* Expense and *Per Capita* General Consumption Expenditure calculated by POF–Brazil
Supporting information 3: Appendix 2: Imputation of food expenditures
Supporting information 4: Appendix 3: User cost of durable goods by Federative Units (R$ annualized values)

Supporting information 5: Appendix 4: Price Deflator by Geographical Context

Supporting information 6: Appendix 5: Mean logarithmic deviation - decomposition by population subgroup

Supporting information 7: Appendix 6: FGT Class of Poverty Measures

Supporting information 8: Appendix 7: Poverty severity and relative contribution by consumption lines, according to Geographical Areas, Major Regions and Population subgroups

Supporting information 9: Appendix 8: Explanatory Variables and Estimated Coefficients (α) of the Heterocedastic Model

Supporting information 10: Appendix 9: Explanatory variables and estimated coefficients (β)

Supporting information 11: Appendix 10: Estimated proportion of people vulnerable to poverty versus the estimated proportions of the poor calculated directly from POF by the 20 geographical contexts which take part in Geographical Areas and Major Regions (Poverty line = R$185)