Measurement of two services components of the basket of the Brazilian Consumer Price Index using the Continuous National Household Sample Survey (PNADC) as data source

Lincoln T. da Silva, Vladimir G. Miranda, Marcelo R. do Amaral, Plinio L. dos Santos, Gustavo V. Leite

gustavo.leite@ibge.gov.br

16th Ottawa Group Meeting
Rio de Janeiro – Brazil
May, 2019

Instituto Brasileiro de Geografia e Estatística

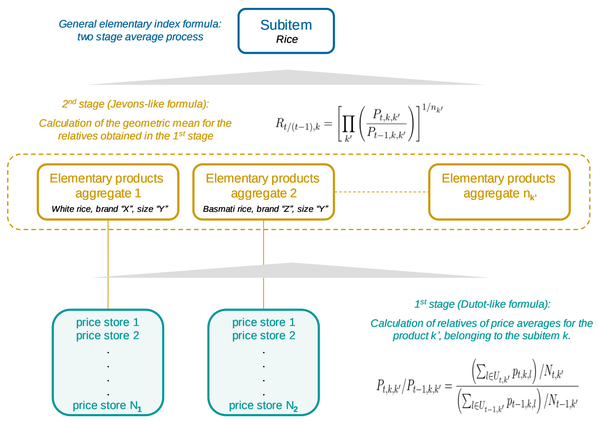
83 anos do IBGE

Introduction

- The IPCA is the Brazilian official CPI produced within the scope of IBGE's System of National Consumer Price Indices (SNIPC) used for the Central Bank to establish its monetary policies.
- As in other CPIs, the IPCA's basket is composed of goods and services representative of consumer habits. The elementary aggregates of the IPCA are denoted *subitems.*
- Among such subitems, two service components, housekeeper (ED) and services for home maintenance and repair (MORD), amounts to approximately, 4% and 1%, respectively, of the IPCA's basket weights.
- Since those services are mainly characterized by Labour's informal market, the measurement of prices relies for these components relies on household surveys.
- Until 2016 IBGE's Monthly Employment Survey (PME) was the source for price data of ED and MORD services.
- With the end of the PME, Labour's force information should be provided by the Continuous Nationial Household Sample Survey (PNADC).
- However, the PNADC sample design is diferente than the one of the PME. And more importantly, the PNADC was designed to provide estimateres on a quarterly basis, whereas the PME used to provide monthly data.
- Due such changes, adoption of the PNADC for calculation of inflation for the subitem ED and MORD is not straightforward.
- The main challenges and approaches adopted are discussed in this paper.

Parameter of interest

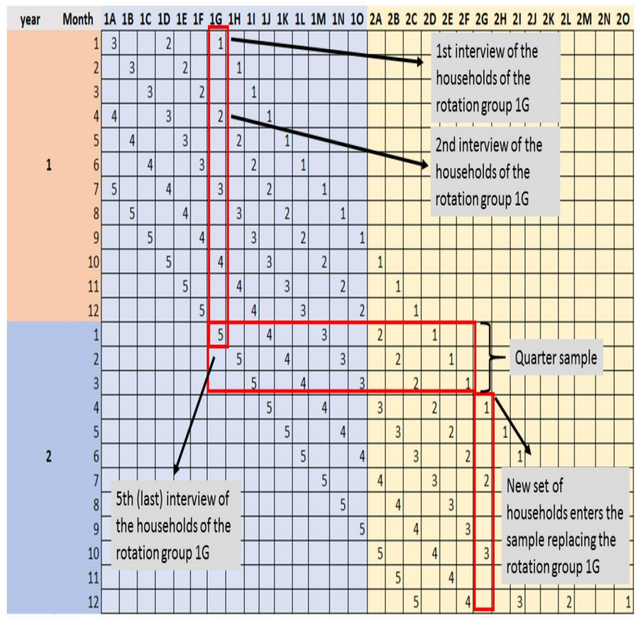
Elementary level formula: index for a subitem *k*
(two-stage formula: Dutot + Jevons)



Since in our case $n_{k'} = 1$, the parameter of interest is given by a Dutot-like formula.

PNADC sample features

PNADC sample rotation scheme



Choice of estimator

Why not use a monthly sample estimator?

$$\hat{R}_{t/(t-1),k} = \hat{P}_{t,k}/\hat{P}_{t-1,k}$$

$$\hat{P}_{t,k} = \left(\sum_{l\in a_{t,k}} w_{t,k,l} \times p_{t,k,l}\right) / \sum_{l\in a_{t,k}} w_{t,k,l}$$

Lack of sobreposition leads to high variance

$$V(\hat{R}_{t/(t-1),k}) = V(\hat{P}_{t,k}) + V(\hat{P}_{t-1,k}) - 2Cov(\hat{P}_{t,k}, \hat{P}_{t-1,k})$$

A better choice

$$\tilde{R}_{t/(t-1),k} = (\tilde{P}_{t,k}/\tilde{P}_{t-3,k})^{1/3}$$

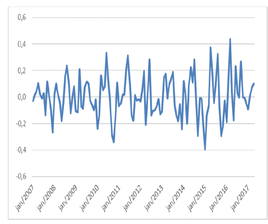
$$\tilde{P}_{t,k} = \left(\sum_{l\in e_{t,k}} w_{t,k,l} \times p_{t,k,l}\right) / \sum_{l\in e_{t,k}} w_{t,k,l}$$

Drawback: bias

$$Bias(\tilde{R}_{t/(t-1),k}) \cong (P_{t,k}/P_{t-1,k}) - (P_{t,k}/P_{t-3,k})^{1/3}$$

Bias analisys

Summary Statistics	Bias	Absolute Bias
Min	-0.4	0.0
Q1	-0.1	0.04
Median	-0.01	0.1
Mean	0.0	0.12
Q3	0.1	0.18
Max	0.44	0.44



Data treatment

Sample weights treatment for the domains of interest

Original PNADC PSU's quarter weights

$$w_{gi} = \frac{1}{m_g} \times \frac{N_g}{N_{gi}}$$

Quarterly weights for selection of a household

$$w_{gij} = w_{gi} \times w_{j|gi} = \frac{1}{m_g} \times \frac{N_g}{N_{gi}} \times \frac{N^*_{gi}}{n_{gi}}$$

Monthly weights correction for selection of a household

$$w^t_{gij} = w_{gij} \times \frac{\sum_{g,i,j} w_{gij}}{\sum_{g,i,j,} w_{gij} \times I(j \in a_t)}$$

Monthly weights with correction for non-responses

$$w^{t_a}_{gij} = w^t_{gij} \times \frac{n^*_{gi}}{n^{**}_{gi}}$$

Correction for matching households

$$w^{t_p}_{gij} = w^{t_a}_{gij} \times \frac{\sum_{g,i,j} w^{t_a}_{gij}}{\sum_{g,i,j} w^{t_a}_{gij} \times I(j \in e_t)}$$

Post stratification correction (final weights)

$$w^{tSNIPC}_l = w^{t_p}_{gij} \times \frac{P_b}{\hat{P}_b}$$

Outliers detection and treatment

Detection:

1) Use of Box-Cox transformation

$$\mathcal{F}_{BC}(p_{t,k,l}) = p'_{t,k,l} = \begin{cases} \frac{p^{\lambda_{BC}}_{t,k,l} - 1}{\lambda_{BC}} & \text{if } \lambda_{BC} \neq 0 \\ \ln(p_{t,k,l}) & \text{if } \lambda_{BC} = 0 \end{cases}$$

2) Construction of tolerance regions

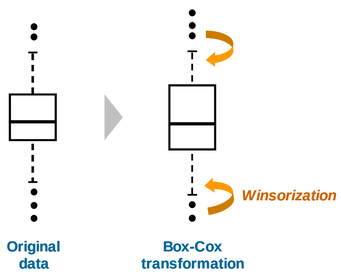
$$LI = Q1 - 2(Q3 - Q1)$$

$$LS = Q3 + 2(Q3 - Q1)$$

Treatment: Winsorization

$$p'_{t,k,l} = \begin{cases} LI & \text{if } p'_{t,k,l} < LI \\ p'_{t,k,l} & \text{if } LI \leq p'_{t,k,l} \leq LS \\ LS & \text{if } p'_{t,k,l} > LS, \end{cases}$$

Schematic view



"Direct" estimates

Direct estimates from the PNADC

Prices estimates after data treatment

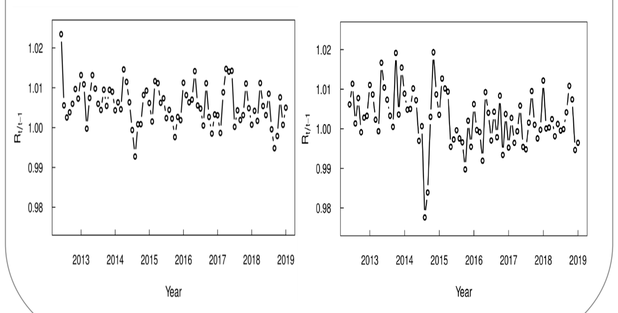
$$\tilde{P}_{t,k} = \left(\sum_{l\in e_{t,k}} w_{t,k,l} \times p_{t,k,l}\right) / \sum_{l\in e_{t,k}} w_{t,k,l} = \left(\sum_{l\in e_{t,k}} w^{tSNIPC}_l \times \mathcal{F}^{-1}_{BC}(p'_{t,k,l})\right) / \sum_{l\in e_{t,k}} w^{tSNIPC}_l$$

$$\tilde{P}_{t-3,k} = \left(\sum_{l\in e_{t,k}} w^{tSNIPC}_l \times \mathcal{F}^{-1}_{BC}(p'_{t-3,k,l})\right) / \sum_{l\in e_{t,k}} w^{tSNIPC}_l$$

Relatives estimates

$$\tilde{R}_{t/(t-1),k} = (\tilde{P}_{t,k}/\tilde{P}_{t-3,k})^{1/3}$$

Series of "direct" estimates



Time series approach

Structural time series (STS) approach

$$ST_t = Trend_t + Seasonal_t + Cicle_t + Irregular_t$$

Desired component

$$Trend_t = ST_t - Seasonal_t - Irregular_t$$

State space equations

$$\tilde{R}_t = \mathbf{Z_t}\boldsymbol{\alpha}_t + \epsilon_t \quad \text{with } \epsilon_t \sim N(0, \mathbf{H}_t)$$

$$\boldsymbol{\alpha}_t = \mathbf{T_t}\boldsymbol{\alpha}_{t-1} + \mathbf{U}_t\boldsymbol{\xi}_t \quad \text{with } \boldsymbol{\xi}_t \sim N(0, \mathbf{Q}_t)$$

Basic structural model

$$\tilde{R}_t = \mu_t + \gamma_t + \epsilon_t \quad \text{with } \epsilon_t \sim N(0, \sigma^2_\epsilon),$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad \text{with } \eta_t \sim N(0, \sigma^2_\eta),$$

$$\beta_t = \beta_{t-1} + \zeta_t \quad \text{with } \zeta_t \sim N(0, \sigma^2_\zeta),$$

$$\gamma_t = \sum_{f=1}^{s/2} \gamma_{f,t}.$$

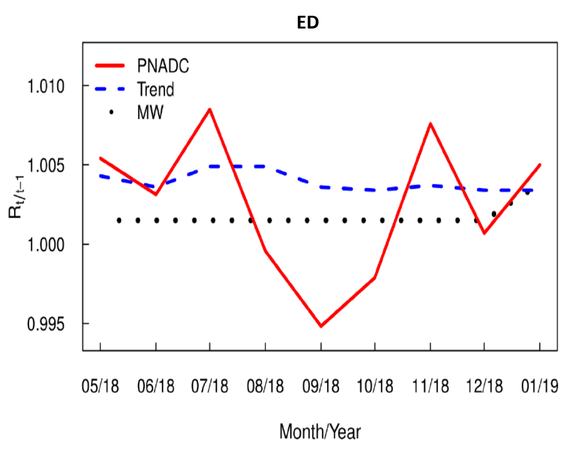
Hyperparameters based on the Kalman filter and maximum likelihood method.

Seasonal component treated via trigonometric functions.

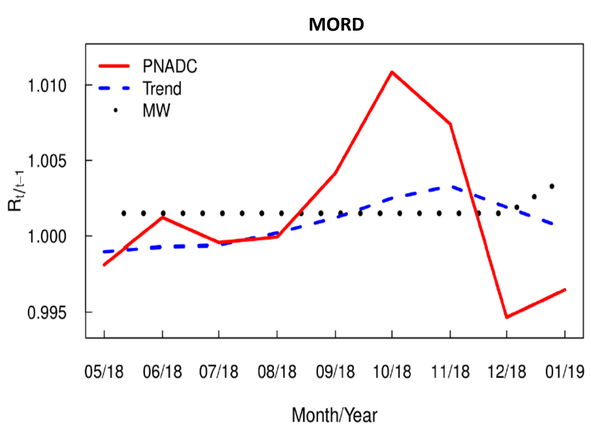
Results

Comparison of estimates derived from the "direct PNADC", STS for PNADC and minimum wage (MW) at the national level

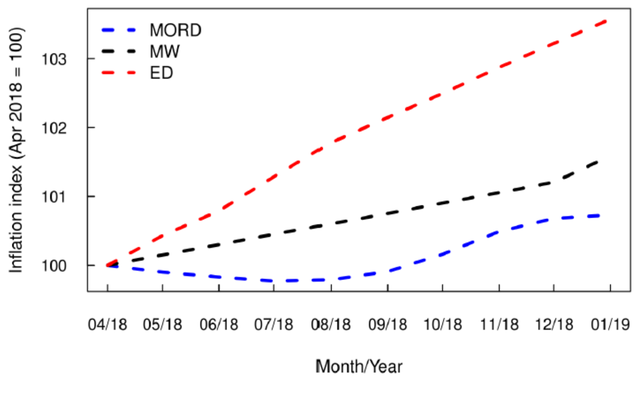
ED



MORD



Comparison of monthly inflation for the STS and minimum wage (MW) estimates



Microdata

Results are released monthly in the page of the PNADC (PNAD Contínua) at the IBGE site

Example of price and weights microdata in the txt files

$p_{t-3,k,l}$	$p_{t,k,l}$	w^{tSNIPC}_l
1000	1200	547.7415699
956	964	559.7468645
954	600	726.3381933
2500	1000	956.7144679
937	954	1013.7662481
960	954	1013.7662481
800	850	646.0075808
954	900	728.5953983
500	500	658.3385862
955	500	794.9736150
250	280	794.9736150
280	2500	1451.1268236
840	1000	1451.1268236
300	300	800.9278299
900	1200	701.3398607
954	954	1499.6492980
400	200	1325.3028530
480	70	1736.5144949
600	600	2552.4055598
800	400	2552.4055598
600	560	1447.6757622
400	400	725.2000739
954	954	141.0628322
100	100	197.0514501
954	900	281.8889155
1200	1200	289.2362006
954	954	277.9214131
700	800	359.6288822
954	500	250.5969280
400	900	250.5969280
700	700	320.8735527
954	800	278.1578971
1000	954	389.2214253
954	2000	406.3560259
480	600	236.3068951
954	954	323.1973792
954	954	426.2240515

Conclusions

- The paper presents the main challenges in moving from the PME survey, previous source of labour's force information, to the PNADC survey for derivation of inflation for the subitems housekeeper (ED), and home and maintenance repairs (MORD).
- The derivation of an estimator taking the complex sample design of the PNADC is discussed.
- Sample weighs treatment and the outlier detection and treatment for the use of the PNADC data are presented.
- STS approach is employed in order to deal with the fluctuations rising from the small samples obtained.
- The results obtained via the "raw" PNADC estimator and the trend component estimator are compared with the estimates obtained via the Brazilian official minimum wage.
- The estimates obtained via STS modelling approach show good agreement with the national economical scenario for the period analysed. The analysis reveals the power of the method and its superiority respective the minimum wage approach which is insensitive to monthly conjunctural changes in the economy which affects the labour's market.
- This work also serves as a tutorial for users interested in working with the prices microdata released for the ED and MORD subitems.
- The data is only available for the national level of aggregation since for finer levels the results still presented high volatility and their use for the calculation of the CPI was not considered appropriate.