

Nota Metodológica 01/2018

Metodologia de cálculo da inflação dos subitens do SNIPC Empregado Doméstico e Mão de Obra para Reparos Domésticos a partir das informações de rendimentos da PNAD Contínua.

1. Introdução

O Sistema Nacional de Índices de Preços ao Consumidor (SNIPC) requer a estimação da **variação média mensal** dos preços dos subitens que compõem uma cesta definida. A variação dos preços de cada um dos subitens da cesta recebe um peso de forma a produzir índices de preços que representem a inflação para uma determinada população de referência.

Na nomenclatura do SNIPC, os subitens são constituídos por um conjunto de produtos e serviços homogêneos representativos das diferentes formas de comercialização cujos preços são coletados regularmente através da seleção de amostras de produtos e locais de comercialização. Para dois subitens em particular, a saber, **Empregado Doméstico (ED)** (código 7101010) e **Mão de Obra para Reparos Domésticos (MORD)** (código 2103042), no entanto, os dados eram até recentemente obtidos de informações coletadas na Pesquisa Mensal de Emprego (PME) do IBGE (IBGE, 2007) (ver também (IBGE, 2013, pág. 37)).

Entretanto, a PME foi encerrada em fevereiro de 2016, sendo desse mês os últimos dados disponíveis oriundos desta fonte. Com o fim da PME, uma solução definitiva ainda não estava disponível para resolver o problema da estimação da inflação dos subitens ED e MORD. Desde então, uma solução temporária vem sendo utilizada na imputação das variações médias de preços destes dois subitens, conforme descrito na nota técnica 03/2016 (IBGE, 2016a), disponível no portal do IBGE da internet.

Devido ao encerramento da PME e sua substituição pela Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC) (IBGE, 2014a; IBGE, 2017) na produção de

indicadores conjunturais sobre o mercado de trabalho no Brasil, pretende-se adotar esta pesquisa como fonte das informações necessárias para estimar a variação média de preços dos dois subitens acima mencionados. Este texto descreve em detalhes a abordagem metodológica desenvolvida e os procedimentos necessários para utilização dos dados da PNADC para estimar os relativos mensais de preços dos subitens ED e MORD.

O documento está estruturado da seguinte forma. Na seção 2 define-se o problema de interesse partindo-se da teoria de índices de preços para estabelecer o parâmetro populacional para medida da 'inflação' de um subitem, a saber, o relativo de preços do subitem entre dois meses.

Na seção 3 detalham-se as considerações metodológicas para a construção do estimador de relativos a partir dos dados da PNADC. Para tanto, discutem-se brevemente as características da amostra da PNADC e as implicações do caráter de periodicidade trimestral da PNADC para a escolha de um estimador de relativos mensais. Mostra-se que, por serem disjuntas as amostras da PNADC em meses consecutivos, o uso de um estimador que seja construído a partir de informações de meses consecutivos não é recomendável por possuir variância elevada. A alternativa sugerida é usar um estimador que leve em conta informações de domicílios pareados da amostra da PNADC, ou seja, informações do subconjunto de mesmos domicílios informantes nos meses t e $t - 3$.

Na seção 4 são discutidas todas as etapas para seleção e tratamento dos dados da PNADC para cálculo das estimativas do estimador proposto na seção 3. A primeira etapa é a seleção das ocupações da PNADC adequadas para comporem os subitens ED e MORD. Em seguida, apresenta-se o processo usado para parear as informações de um mesmo indivíduo. A terceira etapa de tratamento das informações da PNADC envolve a recalibração dos pesos de um indivíduo, necessária uma vez que o uso das informações mensais e pareadas considera uma subamostra da PNADC. Por fim, o último item da seção 4 apresenta o processo de detecção e tratamento de observações atípicas.

Mesmo após os tratamentos descritos, as séries das estimativas de rendimentos apresentam uma variabilidade alta, típica de dados oriundos de amostras pequenas, e verifica-se a necessidade de adoção de modelos estruturais de séries temporais para obter as tendências dos relativos dos subitens aqui estudados. A seção 5 descreve brevemente os modelos estruturais utilizados. As conclusões do document são apresentados na seção 6.

2. Definição do problema

A obtenção do índice de preços de um item qualquer a partir dos subitens componentes emprega a fórmula de Laspeyres para medida do movimento de preços

entre dois momentos T (período corrente) e 0 (período base), que pode ser escrita da seguinte forma (vide (IBGE, 2013, pág. 39)):

$$L_{T/0} = \sum_{k=1}^K \mathcal{W}_{0,k} R_{T,k} = \sum_{k=1}^K \mathcal{W}_{0,k} \times \left(\frac{P_{T,k}}{P_{0,k}} \right), \quad (1)$$

onde

$P_{T,k}$ é o preço do subitem k no período T ;

$P_{0,k}$ é o preço do subitem k no período base 0 ;

$R_{T,k} = P_{T,k}/P_{0,k}$ é o relativo de preços do subitem k no período T comparado com o período base 0 ;

K é o número total de subitens que compõem o item considerado; e

$$\mathcal{W}_{0,k} = Q_{0,k} P_{0,k} / \sum_{l=1}^K Q_{0,l} P_{0,l} \quad (2)$$

é o peso do subitem k obtido a partir da Pesquisa de Orçamentos Familiares - POF, e representa a parcela dos gastos com o subitem k no total dos gastos do item (conjunto) considerado no período base, pois $Q_{0,k}$ é a quantidade consumida do subitem k no período base 0 .

Vale notar que os pesos $\mathcal{W}_{0,k}$ não variam com o período T , sendo fixados no período base, e satisfazem:

- a) $0 \leq \mathcal{W}_{0,k} \leq 1$, para todo k ;
- b) $\sum_{k=1}^K \mathcal{W}_{0,k} = 1$.

Ainda conforme (IBGE, 2013, pág 39), 'o interesse maior, em termos de cálculo dos índices de preços ao consumidor, é obter variações de preços em períodos menores, por exemplo, **em ciclos mensais.**' Por este motivo, adota-se a estratégia de calcular variações mensais e acumular estas variações no tempo para obtenção do índice de Laspeyres.

Uma forma de ver esta operação de calcular variações mensais para depois acumulá-las é dada pela expressão:

$$L_{T/0} = \sum_{k=1}^K \mathcal{W}_{0,k} \prod_{t=1}^T (P_{t,k}/P_{t-1,k}) = \sum_{k=1}^K \mathcal{W}_{0,k} \prod_{t=1}^T R_{t/(t-1),k}, \quad (3)$$

onde

$$R_{t/(t-1),k} = P_{t,k}/P_{t-1,k}. \quad (4)$$

A expressão (3) mostra que o índice de Laspeyres para variação de preços entre 0 e T para o conjunto dos subitens que formam o item de interesse depende dos pesos constantes do período base ($\mathcal{W}_{0,k}$) para cada subitem, e dos produtos dos relativos de preços ($R_{t/(t-1),k}$) calculados a cada par de meses para cada subitem.

Esta outra forma de ver o índice de Laspeyres para o item (ou conjunto de subitens) de interesse é atraente por explicitar um alvo de inferência em nível de subitem a cada mês que é dado pelo **relativo simples** de preços do subitem no mês atual t e no mês anterior $t - 1$.

Em tese, seria possível então obter o índice de Laspeyres obtendo a cada mês t o preço do subitem k no mês t ($P_{t,k}$), e depois calculando as razões entre tais preços a cada par de meses adjacentes, para obter os relativos simples, e finalmente multiplicando estes relativos para acumular as variações mensais.

Acontece que o preço do subitem k no mês t ($P_{t,k}$) **não é conhecido** por diversos motivos: primeiro, na verdade o subitem não é um produto bem especificado, e sim uma coleção de produtos; tampouco há um único preço praticado em todos os locais onde produtos deste subitem são ofertados; e finalmente, não seria possível na prática fazer censos mensais para coleta de preços de todos os produtos componentes do subitem k em todos os locais onde este produto é ofertado – será sempre necessário usar uma amostra de locais e de produtos para **estimar** essa quantidade.

Fica então evidente a necessidade de explicitar as quantidades populacionais que se quer obter em nível de subitem antes de propor estimadores para estas quantidades, a serem obtidos com base na amostra de uma fonte de dados adequada.

Uma primeira alternativa seria definir como alvo de inferência a cada mês simplesmente o **preço médio** do subitem k no mês t ($P_{t,k}$). Nesse caso, a quantidade populacional de interesse a cada mês seria definida como:

$$P_{t,k} = \left(\sum_{l \in U_{t,k}} P_{t,k,l} \right) / N_{t,k} \quad (5)$$

onde $P_{t,k,l}$ é o preço praticado no mês t para uma combinação de produto x local l , para um produto pertencente ao subitem k ;

$U_{t,k}$ é a população (conjunto completo) de combinações produto x local correspondentes ao subitem k no mês t ;

$N_{t,k}$ é o tamanho da população $U_{t,k}$ de combinações produto x local correspondentes ao subitem k no mês t .

No entanto, o preço médio por subitem a cada mês $P_{t,k}$ não é o alvo direto da inferência, mas sim o relativo de preços médios a cada par de meses consecutivos ($R_{t/(t-1),k}$), dado por uma razão dos preços médios populacionais definidos na Eq. (5):

$$R_{t/(t-1),k} = \frac{P_{t,k}}{P_{t-1,k}} = \frac{\left(\sum_{l \in U_{t,k}} P_{t,k,l} \right)}{N_{t,k}} \bigg/ \frac{\left(\sum_{l \in U_{t-1,k}} P_{t-1,k,l} \right)}{N_{t-1,k}} \quad (6)$$

O problema de interesse reside na construção de estimadores adequados para o parâmetro populacional definido na Eq. (6). Na seção seguinte apresentam-se as discussões e detalhes para a escolha de um estimador para os relativos dados pela Eq. (6), para os subitens ED e MORD, levando-se em conta as peculiaridades da amostra da PNADC.

3. Considerações para a escolha do estimador

Conforme descrito anteriormente, a fonte de dados que vinha sendo utilizada para obtenção dos estimadores para cálculo da inflação dos subitens ED e MORD era a PME. Com a extinção desta pesquisa esta fonte será substituída pela PNADC. Antes de propor os estimadores a serem utilizados com base na PNADC, é necessário entender os principais aspectos metodológicos desta pesquisa e suas diferenças em relação à PME (para maiores detalhes vide (IBGE, 2014a; IBGE, 2017; Freitas e Antonaci, 2014)).

A Pesquisa Nacional por Amostra de Domicílios Contínua - PNADC é realizada através de uma amostra probabilística de domicílios (Freitas e Antonaci, 2014) e destina-se a produzir informações contínuas sobre a inserção da população na força de trabalho, associada a características demográficas e de educação, e, também, para o estudo do desenvolvimento socioeconômico do País. As informações são produzidas para os seguintes níveis geográficos: Brasil, Grandes Regiões, Unidades da Federação, Municípios das Capitais, Regiões Metropolitanas que contêm os municípios das capitais, e Região Integrada de Desenvolvimento da Grande Teresina.

A amostra da PNADC foi planejada visando produzir informações trimestrais. Isto implica que são necessários três meses para concluir a coleta da amostra selecionada em todo o país. Para garantir que as variações observadas nas estimativas de dois trimestres consecutivos não sejam demasiado voláteis, foi adotado um esquema de rotação com sobreposição parcial da amostra em trimestres consecutivos. Neste esquema, um conjunto de domicílios selecionado é entrevistado por no máximo 5 trimestres, e entre uma entrevista e outra há um intervalo de três meses, conforme

apresentado na Figura 1. Para operacionalizar o esquema de rotação, as Unidades Primárias de Amostragem (UPAs) de cada estrato foram divididas em 15 grupos de rotação (denotados pelas letras A, B, ..., O na Figura 1). A cada trimestre é selecionada uma nova amostra de domicílios nas UPAs de 3 destes grupos e, nas UPAs dos demais 12 grupos, os domicílios que já faziam parte da amostra farão nova entrevista. Este esquema proporciona que a amostra de um trimestre tenha uma sobreposição esperada de 80% dos domicílios com o trimestre anterior e de 20% com o mesmo trimestre do ano anterior.

Mês	1A	1B	1C	1D	1E	1F	1G	1H	1I	1J	1K	1L	1M	1N	1O	2A	2B	2C	2D	2E	2F	2G	2H	2I	2J	2K	2L	2M	2N	2O
1	3			2			1															1								
2		3			2			1																						
3			3			2			1																					
4	4			3			2			1																				
5		4			3			2			1																			
6			4			3			2			1																		
7	5			4			3			2			1																	
8		5			4			3			2			1																
9			5			4			3			2			1															
10				5			4			3			2			1														
11					5			4			3			2			1													
12						5			4			3			2			1												
1							5			4			3			2			1											
2								5			4			3			2			1										
3									5			4			3			2			1									
4										5			4			3			2			1								
5											5			4			3			2			1							
6												5			4			3			2			1						
7													5			4			3			2			1					
8														5			4			3			2			1				
9															5			4			3			2			1			
10																5			4			3			2			1		
11																	5			4			3			2			1	
12																		5			4			3			2			1

FIGURA 1: ESQUEMA DE ROTAÇÃO 1-2(5) DA PNADC DESTACADO PELOS RETÂNGULOS VERMELHOS NA FIGURA.

FONTE: (FREITAS E ANTONACI, 2014).

Na construção de estimadores para os preços médios e o relativo de preços é importante notar que, na linguagem da PNADC, as combinações de produto x local de interesse para o índice de preços são os trabalhadores na faixa etária de interesse que estão ocupados e pertencem ao grupo de empregados domésticos ou ao grupo dos que oferecem mão de obra para reparos em domicílios, num determinado mês de referência para uma dada região de interesse.

Estimativas mensais para as médias populacionais $P_{t,k}$ podem então ser obtidas a partir de uma subamostra da PNADC onde o domínio de interesse são os grupos de trabalhadores que representam os subitens ED e MORD a cada mês. O estimador amostral para essas médias populacionais é dado por:

$$\hat{P}_{t,k} = \left(\sum_{l \in a_{t,k}} w_{t,k,l} \times p_{t,k,l} \right) / \sum_{l \in a_{t,k}} w_{t,k,l} \quad (7)$$

onde $a_{t,k}$ é a amostra de pessoas pesquisadas no mês t pertencentes ao subitem ou domínio k (ED ou MORD), $p_{t,k,l}$ é o valor do ‘preço’ praticado pelo trabalhador l do subitem k no mês t , e $w_{t,k,l}$ é o peso amostral do trabalhador l do subitem k do mês t . Vale notar que aqui o ‘preço’ praticado é tomado como sendo o rendimento bruto mensal habitualmente recebido pelo trabalhador ocupado numa das duas categorias ocupacionais dos subitens de interesse, – ver (IBGE, 2016b) e a discussão na seção 4.2.

O uso do estimador dado pela Eq. (7) possui características atrativas, pois não requer produzir pesos especiais ou qualquer outra intervenção nas rotinas de produção de estimativas a partir da amostra da PNADC, além de permitir usar as amostras completas coletadas a cada mês para estimar os domínios definidos para os trabalhadores dos subitens ED ou MORD.

Além disso, estimativas obtidas a partir da Eq. (7) não dependem de modelos, sendo baseadas em estimador consistente e assintoticamente normal, cujas estimativas de precisão são facilmente obtidas usando software padrão que permite levar em conta a estrutura complexa do plano amostral empregado na PNADC.

O estimador para preços médios mensais dado pela Eq. (7) pode ser usado para a construção de um estimador para a variação mensal de preços, quantidade de interesse no cálculo da inflação, utilizando-se a razão de estimadores de preços médios entre pares de meses consecutivos

$$\hat{R}_{t/(t-1),k} = \hat{P}_{t,k} / \hat{P}_{t-1,k} \quad (8)$$

Porém, uma análise mais cuidadosa das características da PNADC revela que a adoção do estimador de relativos dado pela Eq. (8) não é aconselhável. Conforme o processo de rotação de amostras da PNADC discutido acima, as amostras coletadas a cada par de meses adjacentes na PNADC são disjuntas. Este fato acarreta que a variância aproximada do estimador definido na Eq. (8), dada por

$$V(\hat{R}_{t/(t-1),k}) = V(\hat{P}_{t,k}) + V(\hat{P}_{t-1,k}) - 2Cov(\hat{P}_{t,k}, \hat{P}_{t-1,k}) \quad (9)$$

é ‘máxima’. Isto decorre do terceiro termo do lado direito da Eq. (9) ser sempre nulo devido à ausência de correlação entre os indivíduos amostrados em meses adjacentes. Portanto, a variância da razão será ‘máxima’ no sentido de que qualquer sobreposição das amostras de meses adjacentes reduziria a variância, caso os preços nos meses adjacentes sejam positivamente correlacionados, algo que é esperado no caso em questão.

Outro ponto importante é que as amostras coletadas a cada mês na PNADC têm tamanho aproximadamente igual a um terço da amostra completa, cuja coleta é planejada para durar um trimestre inteiro.

A combinação do tamanho reduzido das amostras mensais e a ausência de sobreposição entre as amostras de meses subsequentes implicam numa elevada variabilidade amostral das estimativas das razões de preços médios entre pares de meses subsequentes, causando volatilidade das séries de estimativas obtidas dessa maneira.

Análises iniciais (não mostradas aqui) realizadas com estimativas calculadas usando o estimador definido na Eq. (8) confirmaram a volatilidade esperada nas séries de estimativas resultantes. Sendo assim, concluiu-se que esta opção não era recomendável, e decidiu-se buscar alternativas para a estimação das quantidades de interesse.

Um estimador alternativo pode ser desenvolvido a partir da inspeção da expressão da variância do estimador de relativos para meses adjacentes dada pela Eq. (9). Esta expressão sugere que se as amostras fossem sobrepostas a variância amostral do estimador poderia ser substancialmente reduzida, sendo esta redução tanto maior quanto maior for a sobreposição e a correlação entre preços praticados nos mesmos produtos x locais em momentos distintos.

Sendo assim, decidiu-se investigar um estimador alternativo para o parâmetro dado na Eq. (6) que busca tirar proveito da estrutura longitudinal da PNADC. A ideia é usar amostras emparelhadas de pessoas nos subgrupos de interesse, e usar como estimador para o parâmetro de interesse dado na Eq. (6) o seguinte estimador:

$$\tilde{R}_{t/(t-1),k} = (\tilde{P}_{t,k} / \tilde{P}_{t-3,k})^{1/3}. \quad (10)$$

O estimador dado pela Eq. (10) estima uma variação mensal no mês t , porém esta variação não é oriunda de meses adjacentes e sim a *variação trimestral mensalizada* dos relativos de preços entre os meses t e $t - 3$. Sendo assim, os preços médios a cada mês devem levar em conta o pareamento entre indivíduos nos instantes t e $t - 3$ e são dados por

$$\tilde{P}_{t,k} = \left(\sum_{l \in e_{t,k}} w_{t,k,l} \times p_{t,k,l} \right) / \sum_{l \in e_{t,k}} w_{t,k,l} \quad (11)$$

sendo $e_{t,k}$ a amostra do subitem do mês t pareada com a amostra do mesmo subitem no mês $t - 3$. Isto sugere que, a cada mês, apenas 4/5 da amostra total estariam disponíveis após o pareamento para a estimação de variações nos ‘preços’ dos serviços prestados.

A despeito desta redução do tamanho da amostra disponível para a estimação dos relativos trimestrais, espera-se que a variância do estimador dado pela Eq. (10) seja menor do que a do estimador proposto na Eq. (8).

Vale notar que, sob condições usuais de regularidade, o estimador proposto na Eq. (10) será consistente para a quantidade populacional dada por:

$$(P_{t,k}/P_{t-3,k})^{1/3} \quad (12)$$

que corresponde a um valor distinto do alvo de inferência requerido e definido na Eq. (6).

Portanto, pode-se afirmar que o estimador dado pela Eq. (10) será enviesado para o parâmetro na Eq. (6), sendo o viés assintótico dado por:

$$Vies(\tilde{R}_{t/(t-1),k}) \cong (P_{t,k}/P_{t-1,k}) - (P_{t,k}/P_{t-3,k})^{1/3} \quad (13)$$

Esse viés tenderá a ser positivo em casos de preços com tendência de aceleração de alta, e negativo em caso de preços com tendência de desaceleração de alta. Análises similares podem ser feitas para o caso da tendência de queda dos preços.

Apesar do viés esperado nesse novo estimador, espera-se que o erro quadrático médio do estimador proposto na Eq. (10) seja menor que o do estimador proposto na Eq. (8), pois a redução de variância devida ao uso de amostras pareadas deve mais que compensar o viés incorrido por estimar as variações mensais de preços com base em uma razão de preços médios com espaçamento de 3 meses entre os preços considerados no numerador e no denominador da razão.

Para ilustrar a questão relacionada com o viés da estimação, são utilizados dados da série histórica do IPCA para calcular o viés incorrido pelo uso do tipo de estimador aqui proposto, em comparação com o estimador 'direto' da variação mensal de preços. A Tabela 1 apresenta as estatísticas resumo da distribuição do viés e do **módulo** do viés calculados usando a Eq. (13) aplicada à série do IPCA, considerando o período de janeiro de 2007 a abril de 2017.

Observa-se da Tabela 1 que a distribuição da série do **módulo** do viés fica em torno de 0,1p.p para até 50% dos dados da série e $\approx 0,18$ p.p para até 75% dos dados da série. As estatísticas resumo do viés também apresentam importantes aspectos como valor médio da série igual a zero. Isto indica que o viés não é cumulativo ao longo do tempo e tende a se cancelar com o passar dos meses, o que é altamente desejável para a medida da inflação.

A Figura 2 mostra a evolução das estimativas do viés para todo o período considerado. A análise desta figura mostra que o viés tem variação modesta, nunca tendo superado 0,5 pp em valor absoluto e que vieses negativos e positivos tendem a se alternar ao longo do tempo, em acordo com as estatísticas resumo apresentadas na Tabela 1.

Resumos da Distribuição	Viés	Viés Absoluto
Mínimo	-0,40	0,00
Q1	-0,10	0,04
Mediana	-0,01	0,10
Média	0,00	0,12
Q3	0,10	0,18
Máximo	0,44	0,44

TABELA 1: RESUMOS DA DISTRIBUIÇÃO DA SÉRIE DO VIÉS E DO MÓDULO DO VIÉS, EM VALORES PERCENTUAIS, CALCULADOS USANDO A EQ. (13) APLICADA À SÉRIE DO IPCA, CONSIDERANDO O PERÍODO DE JANEIRO DE 2007 A ABRIL DE 2017.

Com base na discussão e nos dados apresentados, onde observa-se que o viés encontrado ao utilizar o estimador dado pela Eq. (10) é pequeno, decidiu-se adotar este estimador para o relativo de preços dos subitens ED e MORD.

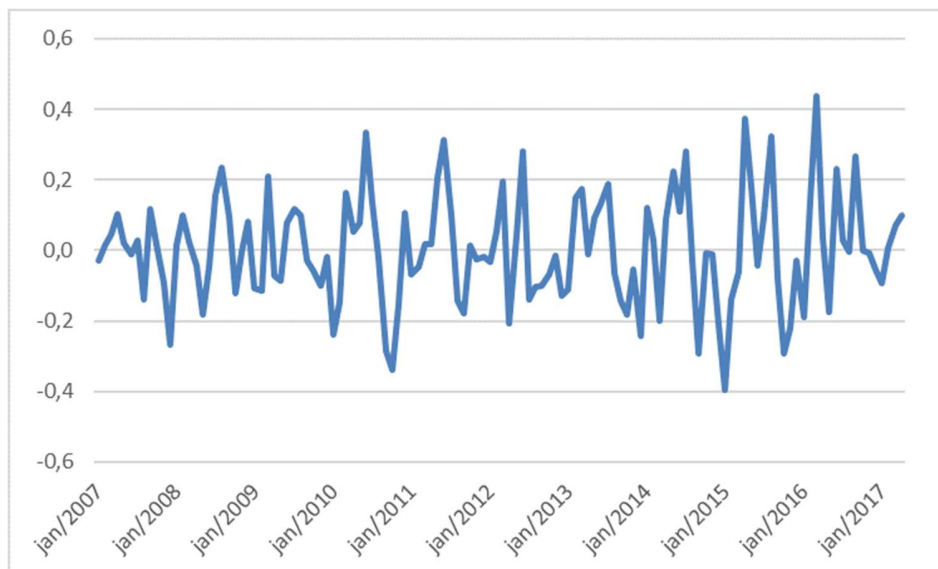


FIGURA 2: SÉRIE DO VIÉS CALCULADO USANDO A EQ. (13) APLICADA À SÉRIE DO IPCA, CONSIDERANDO O PERÍODO DE JANEIRO DE 2007 A ABRIL DE 2017.

4. Seleção e tratamento dos dados da PNADC para uso no cálculo do estimador adotado

A presente seção detalha todos os processos de seleção e tratamento das informações da PNADC para uso do estimador adotado para o cálculo da inflação dos subítem ED e MORD do SNIPC.

4.1. Critérios para seleção das ocupações pertinentes da PNADC

A PNADC considera como *trabalhador doméstico* o indivíduo que trabalha na prestação de serviço doméstico remunerado em uma ou mais unidades domiciliares (IBGE, 2017). A escolha das categorias profissionais representativas do subitem ED levou em consideração um grupo bem diversificado de categorias ocupacionais, mas a seleção das mesmas envolveu uma avaliação qualitativa da relevância de cada uma dessas ocupações nos hábitos de consumo das famílias, conforme a população-objetivo do IPCA. Como ilustração, a participação relativa do gasto com empregados domésticos no orçamento das famílias corresponde a 4,0033% na estrutura de ponderação do IPCA (IBGE, 2014b), que tem como fonte de informação a Pesquisa de Orçamentos Familiares – POF 2008-2009. Em paralelo, para uma data específica, o estudo também considera o levantamento das frequências das ocorrências de cada uma dessas ocupações na lista de categorias identificadas na PNAD Contínua.

Com base nos critérios acima, observa-se que o subitem ED do SNIPC pode ser adequadamente representado na PNADC pelo grupamento ocupacional *trabalhadores dos serviços domésticos em geral*, cód. 9111 (na variável V4010) da classificação de ocupações de pesquisas domiciliares¹, pois este engloba a maior parte de ocupações de interesse para o caso do SNIPC.

Critérios análogos de representatividade no padrão de consumo das famílias do IPCA e frequência de ocorrência foi adotado para a escolha das categorias laborais representativas do subitem MORD. Assim, o *trabalhador que presta serviços em obras de acabamento, instalações e serviços auxiliares da construção* é definido como mão de obra para serviços de acabamento, manutenção e pequenos reparos, em uma ou mais unidades domiciliares, desde que ocupadas na semana de referência da pesquisa, cujo empreendimento da atividade principal não empregue mais do que cinco pessoas.

No sistema de classificação da PNADC (IBGE, 2014a), as ocupações apropriadas para o cálculo dos índices de preços do subitem MORD foram obtidas da Seção Construção, a partir da qual duas classes de atividades da CNAE Domiciliar 2.0 foram identificadas (variável V4013) como adequadas para posterior seleção das ocupações, a saber:

¹ As codificações utilizadas estão disponíveis no endereço na ref (IBGE, 2010).

“construção de edifícios” (41000) e “serviços especializados para construção” (43000).

Dentre as CNAE selecionadas, é realizado um outro filtro para as ocupações de interesse² através do código da classificação de ocupações de pesquisas domiciliares (variável V4010), a saber, pedreiros (cód. 7112), pintores e empapeladores (cód.7131), eletricitas de obras e afins (cód. 7411) e trabalhadores elementares da construção de edifícios (cód. 9313).

Para cálculo da inflação, foi considerada como preço do serviço prestado a variável da PNADC *rendimento bruto mensal habitualmente recebido* (variável VD4016) (IBGE, 2016b). Assim, quando se considera a remuneração habitual, significa descartar as remunerações de natureza esporádica, por exemplo, décimo terceiro salário e adicional por horas extras. Ademais, pelo fato do rendimento ser bruto, não são descontadas as contribuições para a previdência ou descontos ocasionais por falta ao emprego.

A escolha de tal variável baseia-se no julgamento de que esta é mais apropriada para captar a tendência da inflação dos subitens de interesse e mais robusta em relação a flutuações de salários atípicas não representativas da dinâmica do mercado.

4.2. Procedimentos de pareamento

O uso do estimador dado pela Eq. (10) requer as informações de pessoas moradoras de domicílios entrevistados em dois diferentes momentos t e $t - 3$. No entanto, a utilização das informações dos moradores residentes nos domicílios presentes em t e $t - 3$ não é direta pois exige que os mesmos sejam pareados para garantir que os dados são oriundos dos mesmos indivíduos, e também que estes indivíduos não mudaram de ocupação no período.

O processo de pareamento de indivíduos envolve diversas etapas sendo o pareamento do domicílio a primeira a ser executada. Para ser realizado o pareamento de um mesmo domicílio entre duas entrevistas, faz-se uso da identificação do domicílio (*id_dom*) que é composta pela concatenação das variáveis UPA, V1008 (nº do domicílio) e V1014 (grupo de rotação).

A segunda etapa de pareamento corresponde ao pareamento dos moradores dentro dos domicílios. Dentro de cada domicílio existe a possibilidade de acompanhar os moradores caso estes permaneçam como moradores do mesmo domicílio entre as

² A princípio também tinham sido considerados os carpinteiros (cód. 7115), telhadores (cód. 7121), aplicadores de revestimentos cerâmicos, pastilhas, pedras e madeiras (cód. 7122), gesseiro (cód. 7123), vidraceiro (cód. 7125), bombeiros e encanadores (cód. 7126). No entanto, como estas categorias apresentaram um número muito baixo de ocorrências nas amostras analisadas da PNADC, elas foram descartadas.

diversas entrevistas. Para tanto, usa-se a variável V2004 que é o código de identificação dos moradores de cada domicílio. Esse código é gerado automaticamente no Dispositivo Móvel de Coleta (DMC) quando o morador é cadastrado pela primeira vez no domicílio e esse número é único, não se repetindo para nenhum outro morador do domicílio. Ao longo das demais entrevistas, o morador já cadastrado carrega seu código único até sair da amostra.

Após a conclusão do pareamento de todos os moradores entre os meses t e $t - 3$, aplicam-se os filtros de ocupação, definidos na seção anterior, considerando os dados do mês t , para os indivíduos pareados entre t e $t - 3$ e depois localizam-se os registros de pessoas pareadas que se mantiveram na mesma ocupação em t e $t - 3$.

4.3. Reponderação

Os dados disponíveis para o cálculo do estimador dado pela Eq. (10) constituem apenas uma parte da amostra da PNADC, uma vez que provêm de um dado mês t . Além disso, ainda há uma perda devido à não resposta e ao processo de pareamento com os domicílios e moradores pesquisados em t que estavam em $t - 3$. Para obter inferências sobre os parâmetros populacionais é necessário levar em conta o desenho amostral da pesquisa e os respectivos pesos oriundos deste processo. Como as informações utilizadas para o cálculo das estimativas da inflação de rendimentos não usam a amostra completa da PNADC, os pesos precisam ser recalibrados para obtenção de estimativas mensais corretas.

A seguir são apresentadas as etapas de ponderação e os ajustes nos pesos amostrais da PNADC necessários para a apropriação no SNIPC. Para simplificar a notação usada na apresentação, todas as expressões apresentadas nesta subseção se referem a um estrato genérico usado para a seleção da amostra de UPAs da PNADC, mas sem incluir índice adicional para designar este estrato.

4.3.1. Cálculo dos pesos trimestrais

O peso básico de cada UPA para estimar características produzidas a cada trimestre é dado por

$$w_{gi} = \frac{1}{m_g} \times \frac{N_g}{N_{gi}}, \quad (14)$$

onde w_{gi} é o peso básico da UPA i do grupo de rotação g , para um trimestre da PNADC, englobando os meses t , $t - 1$ e $t - 2$ do esquema de trimestre usual da PNADC;

m_g é o número de UPAs selecionadas no grupo de rotação g ;

N_{gi} é o número de domicílios particulares permanentes ocupados, fechados e vagos na UPA i do grupo de rotação g , conforme os dados do Cadastro Nacional de Endereços para Fins Estatísticos - CNEFE existentes no **momento da seleção das UPAs**;

N_g é o número de domicílios particulares permanentes ocupados, fechados e vagos no grupo de rotação g , conforme os dados do CNEFE existentes **no momento da seleção das UPAs**.

O peso básico de cada domicílio é dado pelo inverso da probabilidade de seleção do domicílio e pode ser decomposto na parcela correspondente à seleção da UPA e na parcela correspondente à seleção dos domicílios dentro da UPA, sendo expresso por:

$$w_{gij} = w_{gi} \times w_{j|gi} = \frac{1}{m_g} \times \frac{N_g}{N_{gi}} \times \frac{N_{gi}^*}{n_{gi}} \quad (15)$$

$w_{j|gi}$ é o peso básico do domicílio j na UPA i do grupo de rotação g , num trimestre da PNADC;

N_{gi}^* é o número de domicílios particulares permanentes ocupados e fechados na UPA i do grupo de rotação g , conforme os dados da última atualização do CNEFE **no momento da seleção dos domicílios**;

n_{gi} é o número de domicílios selecionados na UPA i do grupo de rotação g , dentro de um dado estrato.

4.3.2. Cálculo dos pesos mensais

Para utilização dos dados de um único mês t (IBGE, 2015), a correção nos pesos é dada pela seguinte expressão:

$$w_{gij}^t = w_{gij} \times \frac{\sum_{g,i,j} w_{gij}}{\sum_{g,i,j} w_{gij} \times I(j \in a_t)} \quad (16)$$

Na Eq. (16), w_{gij}^t é o peso básico do domicílio j , da UPA i , do grupo de rotação g , para o mês t . Note que o superíndice t foi introduzido para denotar o mês de referência.

O numerador na fração no lado direito da Eq. (16) é o total dos pesos básicos dos domicílios para um trimestre da PNADC e o denominador é o total dos pesos básicos dos domicílios pesquisados no mês t . Para seleção do mês t no denominador da fração introduz-se o indicador para filtrar dentre os domicílios do trimestre de pesquisa apenas os que pertencem ao mês t .

4.3.3. Cálculo dos pesos mensais com ajuste para não resposta

Devido à perda de entrevistas decorrentes de não resposta, os pesos básicos mensais recebem um novo ajuste utilizando a fórmula a seguir:

$$w_{gij}^{t_a} = w_{gij}^t \times \frac{n_{gi}^*}{n_{gi}^{**}}, \quad (17)$$

onde $w_{gij}^{t_a}$ é o peso do domicílio j , da UPA i do grupo de rotação g , no mês t , ajustado devido à não resposta. O superíndice t_a é usado para denotar peso relativo ao mês t ajustado por não resposta.

n_{gi}^* é o número de domicílios selecionados com morador na UPA i do grupo de rotação g ;

n_{gi}^{**} é o número de domicílios selecionados com entrevista realizada na UPA i do grupo de rotação g .

Em estratos onde não há domicílios pareados, esta etapa de ajuste é realizada no pós-estrato para todos os estratos componentes do pós-estrato.

4.3.4. Cálculo dos pesos mensais para domicílios pareados

Para incorporar a modificação na amostra devida ao pareamento, os pesos são novamente ajustados por:

$$w_{gij}^{t_p} = w_{gij}^{t_a} \times \frac{\sum_{g,i,j} w_{gij}^{t_a}}{\sum_{g,i,j} w_{gij}^{t_a} \times I(j \in e_t)} \quad (18)$$

onde $w_{gij}^{t_p}$ é o peso dos domicílio j , da UPA i , do grupo de rotação g , para o mês t com ajuste de pareamento, para os domicílios com pelo menos um de seus moradores pareados entre os meses t e $t - 3$. O superíndice t_p é usado para denotar as quantidades associadas aos domicílios pareados no mês t .

O numerador da fração na Eq. (18) é o total dos pesos ajustados considerando a não resposta, no mês t e o denominador é o total dos pesos ajustados considerando a não resposta, no mês t , dos domicílios com pelo menos um de seus moradores pareados entre os meses t e $t - 3$. Para filtrar os domicílios pareados no cálculo do total dos pesos no denominador da Eq. (18), introduz-se o indicador de domicílios pertencentes à amostra pareada.

4.3.5 Pós-estratificação dos pesos mensais para domicílios pareados

Como na PNADC os pesos são ajustados por pós-estratificação, considerando a população estimada nos pós-estratos para o dia 15 do mês de referência (IBGE 2015), este ajuste é mantido, sendo realizado em cada estrato pertencente a cada pós-estrato b . Com isto, o último ajuste nos pesos dá origem ao peso final de um indivíduo l , pertencente ao domicílio j , da UPA i , do grupo de rotação g , pareado entre os instantes t e $t - 3$, para o cálculo das estimativas dos relativos de preços do SNIPC:

$$w_l^{tSNIPC} = w_{gij}^{t_p} \times \frac{P_b}{\hat{P}_b}. \quad (19)$$

P_b é a estimativa populacional produzida pelo IBGE para o pós-estrato b para o dia 15 do mês de referência e \hat{P}_b é a estimativa populacional obtida com os dados dos domicílios com pelo menos um de seus moradores pareados para o pós-estrato b no mês de referência.

4.4. Detecção e tratamento de *outliers* dos dados da PNADC apropriados no SNIPC

Um outro processo pelo qual os dados de rendimentos extraídos da PNADC devem passar, após os passos descritos acima, consiste na detecção e tratamento de observações discrepantes (ou *outliers*). Como o cálculo de estimativas para os relativos de preços dado pela Eq. (10) depende das estimativas de preços médios, a presença de *outliers* na distribuição de preços pode ter grande influência sobre as estimativas calculadas a partir da Eq. (11). Para evitar distorções nas estimativas de preços médios e relativos adota-se então um processo automatizado de detecção e tratamento para *outliers*.

O primeiro passo da abordagem é a definição de critérios objetivos para estabelecer quais observações serão consideradas *outliers*. Para uma dada distribuição de preços (rendimentos), considera-se como *outliers* os valores que sejam menores (maiores) que o limite inferior LI (superior LS) da distribuição. Os limites inferior e superior são definidos como

$$LI = Q1 - 2(Q3 - Q1) \text{ e } LS = Q3 + 2(Q3 - Q1), \quad (20)$$

onde, na Eq. (20), $Q1$ ($Q3$) corresponde ao primeiro (terceiro) quartil da distribuição sob análise.

Em cada instante t serão consideradas separadamente duas distribuições de preços, as distribuições de preços em t e $t - 3$ obtidas a partir do pareamento de indivíduos presentes nas amostras nestes dois instantes de tempo.

Para o cálculo dos quartis e limites LI e LS na Eq. (20) é necessário levar em conta os pesos dos indivíduos emparelhados dados pela Eq. (19) para obtenção da distribuição de preços populacional. No entanto, devido ao processo de rotação da amostra e ajustes de não resposta o peso de um indivíduo l emparelhado entre t e $t - 3$ pode ter um valor na amostra do mês t diferente do mês $t - 3$. Com isto, para garantir que a inflação aqui calculada é oriunda exclusivamente de uma variação pura de preços, adotam-se, para os indivíduos emparelhados l entre t e $t - 3$, os pesos dados no mês de referência t para cálculo das quantidades na Eq. (20), tanto para a distribuição do mês t quanto para a do mês $t - 3$.

O segundo passo da abordagem é decidir, uma vez identificados os *outliers*, qual será o tratamento feito com estas observações. Decidiu-se adotar a técnica de winsorização para tratar as informações detectadas como *outliers*. Neste caso, valores considerados *outliers* tem seu valor substituído/imputado pelo valor de *LI* (*LS*) caso sejam menores (maiores) que *LI* (*LS*).

Como distribuições de preços em geral são assimétricas, aplica-se uma transformação de Box-Cox (Box e Cox, 1964) nos dados para aproximá-los de uma distribuição normal para execução dos processos de detecção e tratamento de outliers descritos acima.

Matematicamente, as regras definidas são expressas por

$$p'_{t,k,l} = \begin{cases} LI & \text{se } p'_{t,k,l} < LI \\ p'_{t,k,l} & \text{se } LI \leq p'_{t,k,l} \leq LS \\ LS & \text{se } p'_{t,k,l} > LS, \end{cases} \quad (21)$$

onde, na Eq. (21), $p'_{t,k,l}$ são os preços obtidos a partir da transformação de Box-Cox, dados por (Box e Cox, 1964)

$$\mathcal{F}_{BC}(p_{t,k,l}) = p'_{t,k,l} = \begin{cases} \frac{p_{t,k,l}^{\lambda_{BC}} - 1}{\lambda_{BC}} & \text{se } \lambda_{BC} \neq 0 \\ \ln(p_{t,k,l}) & \text{se } \lambda_{BC} = 0. \end{cases} \quad (22)$$

$p_{t,k,l}$ corresponde ao preço do serviço prestado pelo indivíduo l , do subitem k no instante t . O parâmetro λ_{BC} determina a transformação de Box-Cox (\mathcal{F}_{BC}) a ser usada nos dados e é estimado para cada uma das amostras de preços analisada.

Note que para o cálculo das estimativas de relativos dadas na Eq. (10) é necessário aplicar estes tratamentos para duas distribuições de preços em separado (a dos indivíduos pesquisados no instante t e a dos pesquisados no instante $t - 3$) para obtenção dos valores médios dados pela Eq. (11). Nota-se também que devido aos indivíduos pareados entre t e $t - 3$ e os pareados entre t e $t + 3$ formarem subamostras distintas de indivíduos, cada amostra da PNADC pesquisada num instante t tem tratamento de *outliers* em dois instantes distintos t e $t + 3$. Porém, em cada um destes instantes o tratamento é feito para os dados originais, ou seja, quando os dados pesquisados em t passam pela crítica de *outliers* para cálculo do relativo entre t e $t + 3$ os dados não consideram a crítica feita em t no cálculo do relativo entre t e $t - 3$.

Por fim, como o subitem MORD é composto por várias ocupações, conforme descrito na seção 4.1, para este subitem o tratamento descrito pelas Eqs. (21) e (22) é realizado em separado para cada uma das ocupações.

Com base no peso de um indivíduo pareado das subamostras do SNIPC, no instante t , dado pela Eq. (19), e com os critérios para tratamentos de *outliers* dados pelas Eqs. (20)-(22), as estimativas de preços médios e relativos para os subitens ED e MORD podem ser calculadas a partir das Eqs. (10) e (11) e são dadas por:

$$\tilde{R}_{t/(t-1),k} = (\tilde{P}_{t,k} / \tilde{P}_{t-3,k})^{1/3}, \quad (23)$$

Com

$$\tilde{P}_{t,k} = \left(\sum_{l \in e_{t,k}} w_{t,k,l} \times p_{t,k,l} \right) / \sum_{l \in e_{t,k}} w_{t,k,l} = \left(\sum_{l \in e_{t,k}} w_l^{tSNIPC} \times \mathcal{F}_{BC}^{-1}(p'_{t,k,l}) \right) / \sum_{l \in e_{t,k}} w_l^{tSNIPC}, \quad (24)$$

$$\tilde{P}_{t-3,k} = \left(\sum_{l \in e_{t,k}} w_l^{tSNIPC} \times \mathcal{F}_{BC}^{-1}(p'_{t-3,k,l}) \right) / \sum_{l \in e_{t,k}} w_l^{tSNIPC} \quad (25)$$

onde nas Eqs. (24) e (25) os conjuntos $e_{t,k}$ indicam que as somas são restritas aos indivíduos l pertencentes a um dos subitens k de interesse (ED ou MORD) e que tenham sido pareados entre os meses t e $t - 3$. \mathcal{F}_{BC}^{-1} denota a transformação inversa de Box-Cox dada na Eq. (22). Observa-se que os pesos para os indivíduos pareados que entram no cálculo das Eqs. (24) e (25) são os dados no instante t , tanto para o cálculo de $\tilde{P}_{t,k}$ quanto para $\tilde{P}_{t-3,k}$.

5. Modelo de série temporal para estimação da tendência da série de relativos mensais dos subitens ED e MORD

A série de estimativas de relativos mensais $\tilde{R}_{t/(t-1),k}$ obtidas com procedimento descrito na seção 4 apresenta variações típicas de pesquisas com amostras pequenas. Sendo assim, optou-se por utilizar a componente de tendência da série de estimativas de $\tilde{R}_{t/(t-1),k}$ como medida da variação de preços dos subitens em questão.

A estimativa da tendência é obtida a partir de modelos estruturais que representam a série temporal através de componentes não observáveis de tendência, sazonalidade e irregular que evoluem estocasticamente no tempo.

5.1. Modelos estruturais

Os modelos estruturais para séries temporais (Harvey, 1989) permitem a decomposição de uma série temporal em componentes não observáveis que possuem no entanto, uma interpretação clara, tais como tendência, sazonalidade, ciclo e irregular. O componente de tendência capta a direção e o nível da série. A sazonalidade, por sua vez, representa o movimento periódico da série (neste caso intra-anual). O ciclo está associado a movimentos recorrentes de mais longo prazo e o

componente irregular reflete movimentos não sistemáticos e residuais devidos à variação da série não explicada pelos demais componentes do modelo. A forma aditiva da decomposição, usada para representar uma série de variações simples, como no presente caso, tem a seguinte representação:

$$ST_t = Tendência_t + Sazonalidade_t + Ciclos_t + Irregular_t \quad (26)$$

sendo ST_t a observação da série temporal no período t e os demais componentes são os definidos acima.

No modelo estrutural, os componentes geralmente assumem formas estocásticas, sendo cada componente influenciado por um termo de perturbação (erro aleatório). O foco do processo de modelagem é a estimação dos componentes de interesse, nesse caso a tendência da série em questão. Dessa forma, considerando uma série para qual não há o componente cíclico, tem-se:

$$Tendência_t = ST_t - Sazonalidade_t - Irregular_t \quad (27)$$

Para a estimação destes componentes, os modelos estruturais são expressos sob a formulação de espaço de estados (Harvey, 1989) e utiliza-se o filtro de Kalman (Kalman, 1960), que estabelece um procedimento recursivo para realização de inferências sobre os componentes não observáveis.

Sendo assim, a principal característica da abordagem de espaço de estados está na capacidade de fornecer estimativas atualizadas dos componentes não observáveis e de prever valores futuros das observações.

5.2. Representação de modelos estruturais na formulação de espaço de estados

Na presente seção, a notação \tilde{R}_t será usada tanto para denotar uma série temporal genérica como para fazer conexão com a série de estimativas dos relativos de preços para o mês t , para o subitem k , $\tilde{R}_{t/(t-1),k}$.

Considere uma série temporal \tilde{R}_t , com $t = 1, \dots, T$. Denota-se por α_t o vetor³ dos componentes não-observáveis de \tilde{R}_t , denominado vetor de estados. O foco da modelagem é a realização de inferências sobre o vetor de estados α_t com base na informação contida nos valores observados \tilde{R}_t . O estado atual (ou corrente) do sistema pode ser definido como o conjunto mínimo de informação que, juntamente com as observações futuras, seja suficiente para descrever o comportamento futuro do sistema.

A formulação de espaço de estados é definida por duas equações:

³ Vetores e matrizes são denotados por letras em negrito.

$$\tilde{R}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \epsilon_t \quad \text{com } \epsilon_t \sim N(0, \mathbf{H}_t) \quad (28)$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{U}_t \boldsymbol{\xi}_t \quad \text{com } \boldsymbol{\xi}_t \sim N(0, \mathbf{Q}_t) \quad (29)$$

A primeira equação (Eq. 28), denominada equação de observação, representa a relação entre as observações e os estados atuais dos componentes não-observáveis. A segunda (Eq. 29), equação de transição (ou do sistema), descreve a forma como os componentes não-observáveis evoluem estocasticamente ao longo do tempo e pode também incorporar variáveis que permitam a especificação de processos dinâmicos dos componentes. Informações mais detalhadas estão disponíveis em (Durbin e Koopman, 2001).

As matrizes que determinam as equações do sistema \mathbf{Z}_t , \mathbf{T}_t , \mathbf{H}_t , \mathbf{Q}_t e \mathbf{U}_t geralmente dependem de um conjunto de parâmetros desconhecidos que necessitam ser estimados e terão sua forma detalhada na seção 5.2.2 após apresentação do modelo estrutural básico adotado.

As etapas do processo de previsão de \tilde{R}_t , e de estimação dos componentes não-observáveis $\boldsymbol{\alpha}_t$, são realizadas de forma recursiva com base no filtro de Kalman (Harvey, 1989) apresentado em seguida.

5.2.1 O filtro de Kalman

O filtro de Kalman (FK) é um algoritmo recursivo condicional a um particular conjunto de informações. Se tal conjunto inclui todo o passado e exclui os valores atuais dos dados observados, $\mathbf{D}_{t-1} = (\tilde{R}_1, \dots, \tilde{R}_{t-1})$, permite a realização de previsões. Quando novas medidas de \tilde{R}_t estão disponíveis, o conjunto de informações passa a incluir, além de todo o passado, os valores atuais dos dados observados (\mathbf{D}_t), constituindo-se assim um problema de extração de sinal.

O algoritmo possui três estágios de estimação, a saber:

Previsão: $P[\boldsymbol{\alpha}_t | \tilde{R}_{t-1}]$, permite obter uma estimativa para os componentes do vetor de estados sem incorporar as observações mais recentes;

Atualização: $P[\boldsymbol{\alpha}_t | \tilde{R}_t]$. Processo também conhecido como extração de sinal que permite recuperar, no tempo t , informação sobre um componente não observável do sistema utilizando todas as informações disponíveis até t ;

Suavização: $P[\boldsymbol{\alpha}_t | \tilde{R}_T]$. Fornece uma estimativa no tempo t utilizando toda a informação disponível, anterior e posterior ao tempo t (considerando uma série de tamanho T).

A suavização está associada a uma importante característica do filtro de Kalman: a capacidade de fornecer uma estimativa suavizada retrospectiva para os componentes não observáveis, quando nova informação está disponível. O processo de suavização

permite, assim, revisar as inferências sobre o vetor de estados no passado a partir das informações disponibilizadas até o tempo T . Nesta fase, o algoritmo é processado da observação mais recente até a mais antiga.

Destaca-se que a estimativa da tendência utilizada na solução do problema em questão é aquela obtida no estágio final de suavização.

A estimação dos hiperparâmetros envolvidos num modelo em espaço de estados com base nas equações do FK é realizada utilizando-se o método de máxima verossimilhança descrito em (Harvey, 1989) e (Durbin e Koopman, 2001).

5.2.2 Modelo estrutural básico

O modelo estrutural básico (MEB) contempla, na equação das observações, a soma dos componentes de nível e sazonalidade. Sua forma geral é a seguinte:

$$\tilde{R}_t = \mu_t + \gamma_t + \epsilon_t \quad \text{com } \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad (30)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad \text{com } \eta_t \sim N(0, \sigma_\eta^2), \quad (31)$$

$$\beta_t = \beta_{t-1} + \zeta_t \quad \text{com } \zeta_t \sim N(0, \sigma_\zeta^2), \quad (32)$$

$$\gamma_t = \sum_{f=1}^{s/2} \gamma_{f,t}. \quad (33)$$

O componente μ_t se refere ao componente de tendência (ou nível), β_t é o componente de inclinação (*slope*) e γ_t o de sazonalidade. O termo ϵ_t se refere ao componente irregular do modelo, ao passo que η_t , e ζ_t são os ruídos aleatórios da equação de estados, respectivamente, do nível e da inclinação.

Aqui, tratar-se-á a sazonalidade com o uso de funções trigonométricas seguindo a formulação

$$\begin{bmatrix} \gamma_{f,t} \\ \gamma_{f,t}^* \end{bmatrix} = \begin{bmatrix} \cos(\lambda_f) & \text{sen}(\lambda_f) \\ -\text{sen}(\lambda_f) & \cos(\lambda_f) \end{bmatrix} \begin{bmatrix} \gamma_{f,t-1} \\ \gamma_{f,t-1}^* \end{bmatrix} + \begin{bmatrix} \Omega_{f,t} \\ \Omega_{f,t}^* \end{bmatrix}, \quad \text{com } \begin{bmatrix} \Omega_{f,t} \\ \Omega_{f,t}^* \end{bmatrix} \sim N(0, \sigma_\Omega^2 \mathbf{I}_{2 \times 2}). \quad (34)$$

Na Eq. (34), $\lambda_f = 2\pi f/s$ e f varia entre 1 e $s/2$ para a equação de $\gamma_{f,t}$ e entre 1 e $s/2 - 1$ para a equação de $\gamma_{f,t}^*$. No caso de dados mensais, $s = 12$.

Para este modelo, as matrizes do sistema \mathbf{Z}_t , \mathbf{T}_t , \mathbf{H}_t , \mathbf{Q}_t e \mathbf{U}_t seguem um processo determinístico invariante ao longo do tempo e são dadas por

$$\mathbf{Z}_t = [1 \ 0 \ 1 \ 0 \ 1 \ \dots \ 1]_{1 \times 13}, \quad (35)$$

$$\mathbf{Q}_t = \begin{bmatrix} \sigma_\eta^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_\zeta^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_\Omega^2 \mathbf{I}_{11 \times 11} & & \\ & & & & \\ & & & & \end{bmatrix}_{13 \times 13}, \quad (36)$$

$$\mathbf{U}_t = \mathbf{I}_{13 \times 13}, \quad (37)$$

$$\mathbf{H}_t = \sigma_\epsilon^2, \quad (38)$$

$$\boldsymbol{\xi}_t = \begin{bmatrix} \eta_t \\ \zeta_t \\ \Omega_{1,t} \\ \Omega_{1,t}^* \\ \vdots \\ \Omega_{5,t} \\ \Omega_{5,t}^* \\ \Omega_{6,t} \end{bmatrix}, \quad (39)$$

$$\mathbf{T}_t = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \mathbf{C}_1 & & & \vdots \\ \vdots & \vdots & & \dots & & \\ 0 & 0 & \dots & & \mathbf{C}_5 & 0 \\ 0 & 0 & \dots & & & -1 \end{bmatrix}_{13 \times 13}, \quad (40)$$

onde $\mathbf{I}_{d \times d}$ são matrizes identidade de dimensão d e $\mathbf{C}_1 \dots \mathbf{C}_5$ são matrizes 2×2 dadas por

$$\mathbf{C}_f = \begin{bmatrix} \cos(\lambda_f) & \text{sen}(\lambda_f) \\ -\text{sen}(\lambda_f) & \cos(\lambda_f) \end{bmatrix}. \quad (41)$$

Nesta formulação o vetor de componentes não observáveis é dado por

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_{1,t} \\ \gamma_{1,t}^* \\ \vdots \\ \gamma_{5,t} \\ \gamma_{5,t}^* \\ \gamma_{6,t} \end{bmatrix}, \quad (42)$$

e o vetor de hiperparâmetros a ser estimado é dado por $\boldsymbol{\theta} = (\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\zeta^2, \sigma_\Omega^2)^T$.

Ao representar o modelo definido pelas Eqs. (30)-(34) na formulação de espaço de estados, e utilizando-se o filtro de Kalman para a estimação dos componentes não observáveis, é possível obter uma estimativa suavizada do componente de tendência

μ_t , da série de estimativas de relativos \tilde{R}_t , para utilização no cálculo da inflação dos subitens ED e MORD.

6. Conclusão

O presente documento descreve o novo método de estimação da inflação dos subitens ED e MORD do SNIPC após o fim da PME. Foram detalhadas todas as etapas de construção de um estimador para o relativo de preços dos subitens de interesse levando-se em conta as particularidades do desenho amostral e do esquema de rotação de painéis da PNADC, nova fonte de informações sobre rendimentos, que são então usados para estimar variações de preços dos subitens ED e MORD.

Todos os processos necessários para calcular estimativas dos estimadores adotados foram apresentados em detalhes, como o processo de seleção das ocupações de interesse, o pareamento para identificar indivíduos em trimestres consecutivos, a reponderação necessária para obtenção dos pesos adequados da subamostra da PNADC de interesse para o SNIPC e os métodos utilizados para detecção e tratamento dos outliers presentes nas amostras. Também foi descrita a metodologia de séries temporais adotada para tratar as flutuações encontradas nas séries de estimativas diretas obtidas da PNADC.

Devido à política de não revisão dos dados do SNIPC, os resultados finais das estimativas de tendência não são apresentados para evitar comparabilidade com resultados já divulgados nas séries do SNIPC. No entanto, os estudos mostram que para o nível Brasil os modelos de séries temporais fornecem um bom ajuste para a série de estimativas diretas da PNADC, porém devido às amostras serem diminutas para as grandes regiões, a modelagem perde em qualidade no ajuste e por isto decidiu-se adotar apenas o nível mais agregado Brasil na nova metodologia a ser implantada pelo IBGE.

Referências

- Box, G. E. P., e D. R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society. Series B (Methodological)* 26 (2). [Royal Statistical Society, Wiley]: 211–52. [HTTP://WWW.JSTOR.ORG/STABLE/2984418](http://www.jstor.org/stable/2984418).
- Durbin, J., e S.J Koopman. 2001. *Time Series Analysis by State Space Methods*. Cambridge University Press.
- Freitas, M. P. S, e A. A Antonaci. 2014. "Sistema Integrado de Pesquisas Domiciliares Amostra Mestra 2010 E Amostra Da Pnad Contínua." *Textos Para Discussão Diretoria de Pesquisas* 50. IBGE. [HTTPS://BIBLIOTECA.IBGE.GOV.BR/VISUALIZACAO/LIVROS/LIV86747.PDF](https://biblioteca.ibge.gov.br/visualizacao/livros/liv86747.pdf).
- Harvey, A.C. 1989. *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge University Press.
- IBGE. 2007. "Pesquisa Mensal Do Emprego" 2 ed. IBGE. [HTTPS://BIBLIOTECA.IBGE.GOV.BR/VISUALIZACAO/LIVROS/LIV37313.PDF](https://biblioteca.ibge.gov.br/visualizacao/livros/liv37313.pdf).
- . 2010. "Código de Ocupações Domiciliares." IBGE. [FTP://FTP.IBGE.GOV.BR/CENSOS/CENSO_DEMOGRAFICO_2010/METODOLOGIA/ANEXOS/ANEXO_7_OCUPACAO_COD.PDF](ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Metodologia/Anexos/Anexo_7_Ocupacao_Cod.pdf).
- . 2013. "SISTEMA Nacional de índices de Preços Ao Consumidor: Métodos de Cálculo." *Série Relatórios Metodológicos* 14. IBGE. [HTTPS://BIBLIOTECA.IBGE.GOV.BR/VISUALIZACAO/LIVROS/LIV65477.PDF](https://biblioteca.ibge.gov.br/visualizacao/livros/liv65477.pdf).
- . 2014a. "Pesquisa Nacional Por Amostra de Domicílios Contínua." *Notas Metodológicas* 1. IBGE. [FTP://FTP.IBGE.GOV.BR/TRABALHO_E_RENDIMENTO/PESQUISA_NACIONAL_POR_AMOSTRA_DE_DOMICILIOS_CONTINUA/NOTAS_METODOLOGICAS/NOTAS_METODOLOGICAS.PDF](ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_Continua/Notas_Metodologicas/Notas_Metodologicas.pdf).
- . 2014b. "SISTEMA Nacional de índices de Preços Ao Consumidor: Estruturas de Ponderação a Partir Da Pesquisa de Orçamentos Familiares 2008-2009." *Série Relatórios Metodológicos* 39. IBGE. [HTTPS://BIBLIOTECA.IBGE.GOV.BR/VISUALIZACAO/LIVROS/LIV86320.PDF](https://biblioteca.ibge.gov.br/visualizacao/livros/liv86320.pdf).
- . 2015. "Nota Técnica Pnadc 1: Esclarecimentos Sobre Os Resultados Da Pnad Contínua Produzidos Mensalmente" 1. IBGE. [FTP://FTP.IBGE.GOV.BR/TRABALHO_E_RENDIMENTO/PESQUISA_NACIONAL_POR_AMOSTRA_DE_DOMICILIOS_CONTINUA/MENSAL/NOTAS_TECNICAS/NOTA_TECNICA_01_PNADC_MENSAL.PDF](ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_Continua/Mensal/Notas_Tecnicas/Nota_Tecnica_01_Pnadc_Mensal.pdf).
- . 2017. "Pesquisa Nacional Por Amostra de Domicílios Contínua." *Notas Técnicas* 1.2. IBGE. [HTTPS://BIBLIOTECA.IBGE.GOV.BR/VISUALIZACAO/LIVROS/LIV101392.PDF](https://biblioteca.ibge.gov.br/visualizacao/livros/liv101392.pdf).

———. 2016a. “Nota Técnica 03/2016: Empregados Domésticos E Mão de Obra de Pequenos Reparos.” *SISTEMA Nacional de índices de Preços Ao Consumidor*. IBGE. FTP://FTP.IBGE.GOV.BR/PREÇOS_INDICES_DE_PREÇOS_AO_CONSUMIDOR/SISTEMA_DE_INDICES_DE_PREÇOS_AO_CONSUMIDOR/NOTAS_TECNICAS/SNIPC_NOTA_TECNICA_2016_03.PDF.

———. 2016b. “Glossário Da Pesquisa Nacional Por Amostra de Domicílios Contínua Mensal” 1. IBGE. FTP://FTP.IBGE.GOV.BR/TRABALHO_E_RENDIMENTO/PESQUISA_NACIONAL_POR_AMOSTRA_DE_DOMICILIOS_CONTINUA/MENSAL/GLOSSARIO_PNADC_MENSAL.PDF.

Kalman, R. E. 1960. “A New Approach to Linear Filtering and Prediction Problems.” *Transactions of the ASME – Journal of Basic Engineering*, no. 82 (Series D): 35–45.

Diretoria de Pesquisas

12 de março de 2018