

Taxa de desemprego no Brasil em quatro décadas: retroposição da PNAD contínua de 1976 a 2016 ♦

Rafael Bacciotti¹

Emerson Fernandes Marçal²

Resumo

Este trabalho tem por objetivo retroposicionar a série de desemprego de forma a compatibilizar a Pesquisa Nacional por Amostra de Domicílios Contínua na frequência trimestral com as demais séries de desemprego brasileiras. O trabalho gera, a partir da metodologia de espaço estado multivariada uma série única que permite estimar a taxa de desemprego prevalente no país desde 1976. A retroposição da série de emprego é feita utilizando as informações das seguintes pesquisas: Pesquisa de Emprego e Desemprego do Dieese, Pesquisa Nacional por Amostra de Domicílios. O trabalho contribui para a literatura ao retrapor a séries para um horizonte temporal muito longo que cobre quarenta anos na frequência trimestral. A retroposição sugere que o pico histórico de desemprego no Brasil no período de 1979 a 2016 deu-se no ano de 2016.

Palavras-Chave

Mercado de trabalho. Taxa de desemprego. Espaço-estado. Modelos de componentes não observados. Filtro de Kalman.

Abstract

In this paper we employ a retroposition method to Brazilian Quarterly National Households Survey unemployment data to generate a time series for a longer time span covering the period from 1976 to 2016. Our retroposition method consists on multivariate state space models and allows us to estimate Brazilian unemployment rate. The retroposition is performed using the information from DIEESE's Employment and Unemployment Survey and Annual National

♦ Os autores agradecem a Diogo de Prince e Pedro Luiz Valls Pereira por comentários e sugestões feitas a uma versão preliminar deste trabalho. Os erros remanescentes são de inteira responsabilidade dos autores. O segundo autor também agradece ao CNPq pelo auxílio parcial (Processos – 312017/2019-7 e 436510/2018-9).

¹ Analista – Instituição Fiscal Independente – Senado Federal – End.: Bloco 2 – Interlegis – 2º andar Sala 13 – Brasília/DF – Brasil – CEP 70165-900 – E-mail: rbacciotti@senado.leg.br. ORCID: <https://orcid.org/0000-0003-2908-1396>.

² Coordenador – Fundação Getúlio Vargas – Escola de Economia de São Paulo – Centro de Estudos em Macroeconomia – End.: Rua Itapeva, 286, 10º andar – São Paulo/SP – Brasil – CEP: 01332-000. E-mail: emerson.marcal@fgv.br – ORCID: <https://orcid.org/0000-0002-0841-5644>.

Recebido: 22/12/2017. Aceite: 31/01/2020.
Editor Responsável: Bruno de Paula Rocha



Esta obra está licenciada com uma Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Households Survey National Survey as well. Our work contributes to the literature by generating a series that covers forty years of Brazilian Macroeconomic History on a quarterly basis. To the best of our knowledge an exercise like this was not performed in any other work. Our estimates suggest that peak of unemployment was reached at the last quarter of 2016 for the analyzed time span.

Keywords

Labor market. Unemployment rate. State-space models. Kalman Filter.

Classificação JEL

C32. E24. J21.

1. Introdução

A análise do mercado de trabalho em perspectiva histórica é uma tarefa desafiadora, pois não há uma pesquisa longa, abrangente e ao mesmo tempo compatível em termos metodológicos e conceituais que permita acompanhar de maneira adequada o desempenho de uma variável macroeconômica relevante como a taxa de desemprego.¹

Além das limitações para a análise em perspectiva comparada, a ausência de séries longas e homogêneas pode restringir também a elaboração de modelos econométricos tipicamente utilizados em análises macroeconômicas como, por exemplo, o estudo da relação entre a inflação e a taxa de desemprego.

Essas questões foram acentuadas no início de 2016, quando a Pesquisa Mensal de Emprego (PME) foi descontinuada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Desde então, a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD contínua), resultado da integração entre a PME e da PNAD anual, passou a ser a única referência do instituto na produção e disseminação de estatísticas sobre o mercado de trabalho em alta frequência.²

¹ A taxa de desemprego, ou taxa de desocupação (nomenclatura utilizada pelo IBGE), é definida em termos da força de trabalho: representa a razão entre o número de pessoas desocupadas e o contingente de pessoas na força de trabalho (soma de pessoas desocupadas e ocupadas).

² A amostra da pesquisa foi desenhada para produzir informações trimestrais, embora o IBGE também passou a divulgar recentemente algumas variáveis mensais construídas na forma de trimestres móveis.

A PNAD contínua é mais abrangente que a PME, cuja cobertura se restringia a apenas seis regiões metropolitanas: Belo Horizonte, Porto Alegre, Recife, Rio de Janeiro, Salvador e São Paulo. Ao contrário da PNAD anual, com período de referência em setembro, a nova pesquisa possibilita o acompanhamento conjuntural das variáveis em âmbito nacional.

A despeito dos ganhos decorrentes da ampliação da cobertura geográfica comparativamente à PME, da divulgação mais tempestiva frente à PNAD anual, da atualização e modernização metodológica e conceitual de acordo com as recomendações da Organização Internacional do Trabalho (OIT), as séries históricas da PNAD contínua começam em 2012, momento a partir do qual a nova pesquisa passou a ser executada em caráter definitivo.

O presente estudo tem como objetivo contribuir com a retroposição³ da taxa de desemprego trimestral da PNAD contínua, em complemento aos trabalhos de Alves e Fasolo (2015), Carvalho (2016) e Vaz e Barreira (2016). O exercício realizado neste trabalho reconstrói a taxa de desemprego da PNAD desde 1976. A estimação dos pontos faltantes da taxa de desemprego da nova pesquisa do IBGE foi realizada a partir dos modelos de componentes não observados representados em espaço-estado estimado por filtro de Kalman. Os autores desconhecem trabalhos que tenham procurado realizar uma retroposição para período tão longo com qualquer técnica.

O trabalho está organizado da seguinte forma. Após esta introdução, a seção 2 contextualiza o problema da retroposição no espectro mais amplo de dados com irregularidades, apresenta alguns trabalhos que utilizam a metodologia de modelos de componentes não observados para lidar com essas questões, incluindo também estudos mais específicos aplicados a retroposição de séries históricas para o mercado de trabalho brasileiro. A seção 3 descreve brevemente a metodologia utilizada. A seção 4 faz uma revisão das diversas bases de dados de emprego existentes no Brasil. Na seção 5 os resultados principais do trabalho são apresentados. A seção 6 um balanço dos resultados e realiza-se testes de robustez. Na seção 7 apresenta-se as limitações do trabalho, compara-se com a literatura e sugere-se possíveis refinamentos. Por fim as conclusões finais do trabalho são apresentadas.

³ "Given a time-series for which values cover the time interval t to $t+k$, the backcalculation exercise aims to estimate the missing values $t-1, t-2, \dots, t-j$ for a certain j ." (Cannata et al 2015, p. 108).

2. Discussão da literatura

O tratamento técnico para lidar com questões de compatibilização de séries de maior frequência com séries de menor, interpolar informações faltantes ou retroplar quando apenas dados recentes estão disponíveis é similar. Pode ser feito por diferentes metodologias.⁴

Uma primeira abordagem abrange métodos que não consideram a utilização de séries relacionadas, baseando-se, ou em critérios matemáticos ou em modelos univariados especialmente da família ARIMA.⁵ A segunda abordagem diz respeito aos métodos que lançam mão de indicadores relacionados com a série a ser reconstruída. Exemplos são os procedimentos de Denton (1971), Chow e Lin (1971) e suas extensões conduzidas por Fernández (1981) e Litterman (1983). Esta abordagem pode também utilizar modelos formulados em termos de componentes não observados, na forma espaço-estado, de forma similar à proposta neste trabalho. O objetivo principal das técnicas da segunda abordagem é criar uma nova série temporal que seja consistente com os dados de baixa frequência e que preserve o comportamento de curto prazo da variável relacionada.

A metodologia de componentes não observados é aplicada em Harvey e Chung (2000) para estimar a variação da taxa de desemprego trimestral da *Labour Force Survey* (LFS) no Reino Unido com a utilização de uma medida relacionada - número de pedidos de auxílio-desemprego - mas disponível em termos mensais e que tiveram ganhos de desempenho em relação a outros modelos. Quando utilizam um modelo univariado, a redução do erro quadrático médio comparativamente à estimação mais simples baseada em diferenças foi de 10%. Com o modelo bivariado, que levou em consideração a dinâmica da série relacionada, a redução chegou a 50%.

O trabalho de Moauro e Savio (2005), na linha de Harvey e Chung (2000), utiliza modelos de componentes não observados para lidar com a interpolação e a distribuição de séries de atividade econômica de países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE).

⁴ Conforme Moauro e Savio (2005), Di Fonzo (2003), Angelini, Henry e Marcellino (2003) e Proietti (2005)

⁵ Harvey e Valls Pereira (1985) mostram que os dados faltantes de séries do tipo de fluxo e estoque podem ser estimados através de um modelo ARMA univariado escrito na forma de espaço de estados (com função de verossimilhança calculada pelo filtro de Kalman).

Também encontram evidência de melhor de desempenho frente a outras abordagens, como o próprio modelo de componentes não observados univariado e os métodos de Denton (1971), Chow-Lin (1971) e Litterman (1983).

Especificamente sobre a retopolação de dados no mercado de trabalho brasileiro, vale mencionar que Dias, Issler e Rodrigues (2010) reconstruíram as séries de emprego e renda da PME anteriormente a 2002, ano em que a pesquisa passou por uma grande revisão de metodologia.

Outros estudos já foram conduzidos para lidar especificamente com as séries curtas da PNAD contínua, mas a partir de outras técnicas de estimação. Alves e Fasolo (2015) usaram uma abordagem de vetores autorregressivos de frequência mista (MF-VAR) para simular como teria evoluído a taxa de desemprego da nova pesquisa desde maio de 1999. Em outras palavras, o MF-VAR, representado na forma de espaço de estados e estimado por métodos bayesianos, foi aplicado com o intuito de recuperar a distribuição conjunta da taxa de desemprego no período anterior a 2012 a partir de outras séries correlatas existentes nesse horizonte temporal.

O trabalho de Vaz e Barreira (2016), por outro lado, reconstruiu historicamente a PNAD contínua utilizando as informações da PME e da PNAD anual. Além da taxa de desemprego, a série de salários também foi prolongada. O estudo tratou de minimizar as diferenças entre a PNAD contínua e a PNAD anual através da utilização dos microdados.

A técnica de interpolação de Denton (1971) foi utilizada em Carvalho (2016) para construir séries longas e de alta frequência compatíveis com a PNAD contínua. O exercício foi aplicado para a população ocupada, a população desocupada e a população não economicamente ativa partindo da PNAD anual como referência e a PME como indicadora das oscilações mensais. A PNAD é usada como referência dos valores anuais, supondo que são “os dados que estimam com melhor precisão os valores populacionais dos indicadores de mercado de trabalho” (Carvalho 2016, p. 5).

3. Metodologia de econométrica

Nos modelos estruturais de componentes não observados as séries de tempo são descritas como a soma de componentes não observados.

Um modelo na representação espaço-estado consiste de uma equação de medida e uma equação de transição, definidas a seguir, conforme Durbin e Koopman (2012):

$$y_t = Z_t \alpha_t + \varepsilon_t \quad \varepsilon_t \sim N(0, H_t) \quad (1)$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \eta_t \sim N(0, Q_t) \quad (2)$$

A equação de medida (1) descreve a relação entre a série observada (y_t) e o vetor α_t que contém as variáveis de estado não observadas, tais como tendência, ciclo, sazonalidade. Os choques são supostos, em geral, não correlacionados entre si. A equação de transição (2), por sua vez, exibe a dinâmica dos componentes não observados no tempo. O subscrito t indica o tempo medido na frequência das séries que desejamos reconstruir. No trabalho são utilizadas séries de frequências mais altas que as séries que desejamos reconstruir.

A mensuração do estado, que é não observado, é feita a partir do filtro de Kalman, que de acordo com Koopman *et al.* (2006), desempenha um papel correlato aos mínimos quadrados no modelo clássico de regressão linear. O modelo dado pelas Equações (1) e (2) é estimado por máxima verossimilhança pela atualização recursiva via filtro de Kalman, descrita nas seguintes equações:

$$v_t = y_t - z_t a_t \quad (3)$$

$$a_{t+1} = T_t a_t + K_t v_t \quad (4)$$

Sendo que v_t representa o erro de previsão filtrado um passo a frente e α_t a contrapartida do estado filtrado (α_t) no período t . O mesmo vale para z_t . Maiores detalhes sobre o filtro de Kalman ver Harvey (1989) e Durbin e Koopmans (2012).

Dentro deste arcabouço, é possível fazer previsão extrapolando os componentes para os períodos subsequentes e estimar o estado em qualquer ponto do tempo, utilizando toda a informação amostral (suavização).

Uma propriedade conveniente do processo de estimação sequencial pelo filtro de Kalman é a capacidade de lidar com as informações faltantes de uma série histórica. Segundo Koopman e Hoogerheide (2013), os dados faltantes podem estar em toda a amostra, em pontos específicos ou em blocos. Nessas situações, quando y_t é tida como faltante, a estimação do estado filtrado é realizada substituindo os termos aleatórios pelos seus valores esperados. De maneira análoga, quando o objetivo é fazer previsão, as projeções são obtidas pelo tratamento de $t, t + 1, t + 2 \dots$ como faltantes (Commandeur e Koopman, 2007).

O filtro de Kalman pode ser utilizado para modelar processo multivariados que admitam representação espaço-estado. A análise, contudo, pode se tornar computacionalmente desafiadora, uma vez que as interações dinâmicas entre as séries temporais podem levar a uma rápida elevação no número de parâmetros estimados (Koopman e Hoogerheide 2014). O filtro permite a construção de uma função de máximo-verossimilhança que tem de ser maximizada a partir de técnicas de otimização. Após a convergência é necessário checar em que medida há evidência de que o máximo global tenha sido atingido. Também técnicas de estimação alternativas como Métodos de Cadeias de Markov baseados em Monte Carlo (MCMC) podem ser utilizadas.

No caso deste trabalho, os componentes utilizados são tendência, ciclo (para captar movimentos regulares da série), sazonalidade (para modelar os movimentos na frequência sazonal) e um componente irregular, sem memória temporal.

Esses componentes podem ser modelados de diversas formas. A tendência - componente de longo prazo da série - indica a direção em que a série está se movendo, pode conter o nível e a inclinação, mas também pode ser especificada apenas com o nível. A sazonalidade, quando existir, pode ser modelada através de variáveis *dummies* ou por uma função trigonométrica com amplitude fixa ou variável. O ciclo pode ser especificado por uma forma trigonométrica com períodos curto, médio ou longos ou como um processo autorregressivo de ordem superior a 1.

O modelo estrutural com seus componentes descritos acima pode ser ilustrado da seguinte forma:

$$t, t + 1, t + 2 \dots y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \Omega_\varepsilon) \quad (5)$$

em que μ_t representa a tendência, γ_t a sazonalidade, ψ_t o ciclo ou componente autorregressivo e ε_t o componente irregular, y_t representa um vetor de variáveis, no caso multivariado, ou um escalar no caso univariado. No presente trabalho utilizaremos um modelo multivariado que contém os termos acima.

A seguinte especificação foi estimada:

$$\check{y}_t = \check{Z}_t \check{\alpha}_t + \check{\varepsilon}_t \quad (6)$$

$$\check{\alpha}_{t+1} = \check{T}_t \check{\alpha}_t + \check{R}_t \check{\eta}_t \quad (7)$$

Nas quais $\check{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix}$, $\check{Z}_t = [I \quad \xi]$, $\check{R}_t = [I \quad 0 \quad I]$, $\check{\alpha}_t = \begin{bmatrix} \mu_t \\ \mu_1 \\ AR_t \end{bmatrix}$,

$\check{T}_t = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \phi \end{bmatrix}$, $\check{\varepsilon}_t \sim NI(0, \check{H}_t)$, $\check{\eta}_t = \begin{bmatrix} \check{\eta}_{1t} \\ 0 \\ \check{\eta}_{2t} \end{bmatrix}$, $\check{\eta}_{it} \sim NI(0, \check{Q}_{it})$. A dimensão

de AR_t é de 3×1 , μ_t é de 2×1 , ξ é de 3×2 .⁶

Note que o modelo descrito em (6) e (7) é uma versão restrita de um modelo multivariado. Maiores detalhes podem ser encontrados em Koopmans *et al.* (2006), página 172.

4. Principais bases de emprego no Brasil

As diversas pesquisas passaram por reformulações e aperfeiçoamentos ao longo dos anos tendo sido algumas delas descontinuadas. A Pesquisa de Emprego e Desemprego (PED) do Dieese é única fonte cuja metodologia

⁶ Os termos associados a sazonalidade foram omitidos para simplicidade.

permaneceu inalterada desde seu início em 1984. A PNAD (em 1992), os Censos Demográficos (em 2000) e a PME (em 2002) sofreram revisões que alteraram a dinâmica de suas séries históricas. A compatibilização das séries antes e depois das revisões não costuma ser realizada pelo IBGE, mas no caso da PNAD anual o instituto disponibilizou um conjunto de dados harmonizados iniciados em 1992 com a cobertura geográfica existente até 2003.⁷

A Tabela 1 sintetiza as principais características das séries levantadas no trabalho, cujas séries históricas são apresentadas conjuntamente na Figura 1.

Tabela 1 - Resumo das características das pesquisas

Pesquisa	Fonte	Divulgação	Abrangência	Intervalo	População investigada
PNAD Contínua	IBGE	Trimestral	Nacional	2012 a 2016	14 anos ou mais
PNAD	IBGE	Anual*	Nacional	1976 a 2015	10 anos ou mais
PED	Dieese	Mensal	Região metropolitana de São Paulo	1984 a 2016	10 anos ou mais

Nota: * Com exceção de 1994 e dos anos censitários

Fonte: DIEESE e IBGE. Elaboração própria.

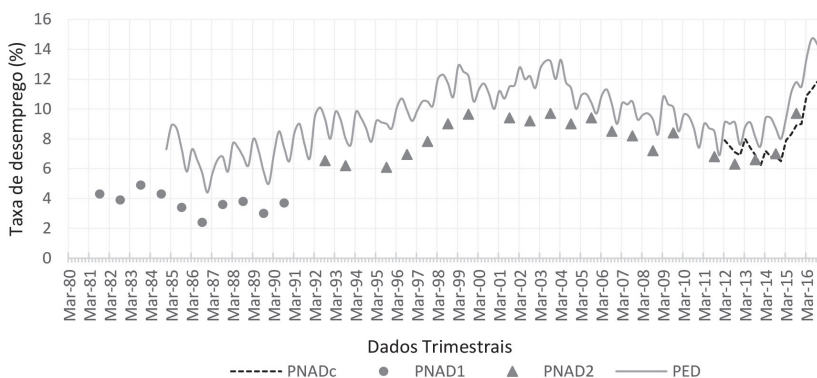


Figura 1 - Taxa de desemprego (% da força de trabalho) desde 1980

Fonte: dados obtidos no IBGE, na OIT, Dieese e em Mattos (2011).

⁷ Disponível desde 2001 em http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2014/default_indicadores_harmonizados.shtm. Acesso em 07/03/2017. Para o período anterior, desde 1992, os dados também harmonizados da taxa de desemprego foram gentilmente cedidos pela pesquisadora do IBGE Vandeli dos Santos Guerra.

5. Resultados

Nesta seção reportam-se os resultados e descrevem-se os dados utilizados nas estimações feitas com o objetivo de reconstruir a série da PNAD a partir de modelos estruturais multivariados de séries de tempo. As variáveis utilizadas nas estimações foram as taxas de desemprego obtidas na PNAD, separada em PNAD 1 (1976 a 1991), PNAD 2 (1992 a 2015); e PED (desde dezembro de 1984).⁸

Na Tabela 2 são apresentadas as principais estatísticas descritivas para cada uma das fontes investigadas. No período em que as séries coexistem, como se observa na Figura 2, nota-se uma diferença de nível, mas as séries parecem ter uma tendência comum, fato muito importante para a retro-polação da série da PNAD contínua.

Tabela 2 - Principais estatísticas descritivas

	Nº de observações	Média	Desvio padrão	Mínimo	Máximo
PNAD Contínua	20	8,3	1,8	6,2	12,0
PNAD anual 1 (1976 - 1991)	14	3,3	0,9	1,8	4,9
PNAD anual 2 (1992 - 2015)	21	8,0	1,3	6,1	9,7
PED	129	9,5	2,1	4,4	14,7

Fonte: Elaboração própria.

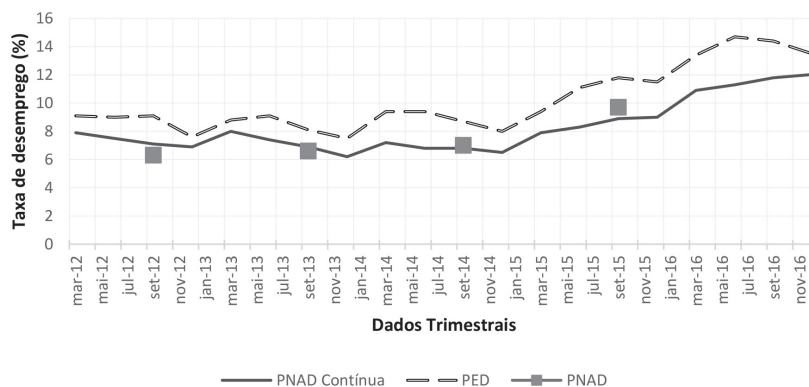


Figura 2 - Taxa de desemprego (% da força de trabalho) desde 2012

Fonte: IBGE e Dieese. Elaboração própria

⁸ Modelos com os dados do Censo foram utilizados, mas os resultados não foram satisfatórios e não reportados para poupar espaço.

Na sequência, foram estimados dois modelos de componentes não observados. Todas as estimações foram feitas no programa *OxMetrics* na versão 7.1 através do módulo STAMP.⁹ As variáveis modeladas conjuntamente foram a taxa de desocupação da PNAD Contínua, a PNAD anual e a PED. A opção pela PED está associada ao maior número de informações disponíveis no período em que há coexistência com a PNAD Contínua. A série da PED é um dos raros exemplos de pesquisa cuja a metodologia permaneceu constante ao longo de toda a série histórica.

Na primeira estimação, a amostra de dados se inicia em 1976. No modelo estrutural, as taxas de desemprego da PNAD contínua, da PNAD anual e da PED foram analisadas conjuntamente considerando os componentes de tendência especificada com o nível, sazonalidade e um processo autorregressivo de primeira ordem.¹⁰

As três séries de desemprego foram consideradas no vetor de variáveis observadas. Na sequência, foram introduzidas restrições nas matrizes de variância dos distúrbios dos componentes para explicitar a dependência da PNAD contínua em relação às dinâmicas da PNAD e da PED e a presença de componentes comuns.

Os dados faltantes obtidos pelo filtro de Kalman compreendidos no período do terceiro trimestre de 1976 ao quarto trimestre de 2011 estão sintetizados na série PNADc - Reconstruída, que pode ser visualizada na Figura 3.

⁹ <http://stamp-software.info/>

¹⁰ A série de desemprego possui um limite inferior e superior, o que torna a introdução de um componente de tendência temporal inadequado.

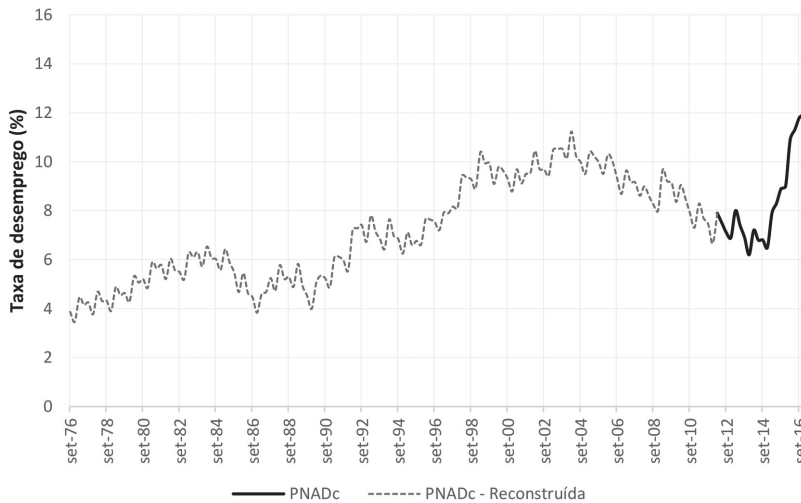


Figura 3 - PNADc - Reconstruída

Fonte: elaboração própria.

Com base nas Equações (6) e (7), as três variáveis utilizadas no vetor são dadas por PNADc, PNAD e PED, nesta ordem.

O vetor de estado no último instante de tempo da amostra é apresentado na Tabela 3. A partir dele, com exceção do ciclo AR (1), os valores dos demais componentes são significativos e, portanto, diferentes de zero. Há também os coeficientes associados às *dummies* sazonais. Como se trata de dados trimestrais, há a necessidade de 3 variáveis *dummies*. As variáveis utilizadas são centradas na média, ou seja, a primeira *dummy* possui valor $\frac{3}{4}$ se o dado é de primeiro trimestre e $-\frac{1}{4}$ se o dado se refere a outro trimestre. Note que no ano soma-se zero. A vantagem desta forma de construir as *dummies* é que o valor do nível independente da sazonalidade. O coeficiente $-0,2894$ está associado com a primeira *dummy* centrada.

Nesta tabela há também informação sobre como os componentes são modelados. Por exemplo, na equação da PNAD anual não há componente sazonal, fato este denotado por zero na Tabela 3. Por serem as séries mais longas, a PED e a PNAD foram utilizadas como base para a estimação do componente de duas tendências comuns utilizadas para reconstruir a tendência não observada da PNADc. O valor um na Tabela 3 indica que há um

componente de tendência para a série e o valor dois indica que a equação contém os componentes comum estimados.

Tabela 3 - Modelo multivariado PNADc – Reconstruída

Vetor Y : PNADc, PED, PNAD			
Y = Nível + Sazonalidade + Irregular + AR(1)			
Seleção dos componentes: 0=Sem, 1=Com, 2=Dependente, 3=Fixo			
	PNADc	PED	PNAD
Nível	2	1	1
Sazonalidade	2	1	0
AR(1)	2	1	1
Irregular	1	1	1
	Vetor de estado (4ºT/2016)	Valor	Prob
	Nível	12,83561	[0,00000]
	Sazonalidade 1	-0,28939	[0,00178]
	Sazonalidade 2	0,33905	[0,00038]
	Sazonalidade 3	-0,23608	[0,00009]
	AR(1)	-0,28136	[0,67748]

Nota: 0 -> Sem o componente na equação

1 -> Com componente presente e base de normalização

2 -> Com componente presente

3 -> Componente com variância zero.

Fonte: Elaboração própria.

6. Testes de robustez e balanço dos resultados

Tendo em vista avaliar a robustez dos resultados, um modelo similar ao formulado anteriormente foi estimado a partir de 1984, utilizando-se os dados harmonizados da PNAD anual disponibilizados pelo IBGE a partir de 1992.

Dos resultados apresentados na Tabela 4 é possível notar que os valores dos componentes do vetor de estado estimado no último instante de tempo - quarto trimestre de 2016 – são estatisticamente significativos. No caso do AR(1), este foi a pelo menos 10%, sinalizando que a formulação do modelo com a presença desses componentes parece adequada. Todas as *dummies* sazonais são significativas, indicando a presença de sazonalidade na série de desocupação como seria esperado.

Tabela 4 - Modelo multivariado PNADc – Reconstruída 1

Vetor Y: PNADc, PED, PNAD2			
Y = Nível + Sazonalidade + Irregular + AR(1)			
Seleção dos componentes: 0=Sem, 1=Com, 2=Dependente, 3=Fixo			
	PNADc	PED	PNAD2
Nível	2	1	1
Sazonalidade	2	1	0
AR(1)	2	1	1
Irregular	2	1	1
Vetor de estado (4ºT/2016)		Valor	Prob
Nível		12,48701	[0,00000]
Sazonalidade 1		-0,30276	[0,00359]
Sazonalidade 2		0,31432	[0,00231]
Sazonalidade 3		-0,22883	[0,00016]
AR(1)		0,07335	[0,08247]

Nota: 0 -> Sem o componente na equação

1 -> Com componente presente e base de normalização

2 -> Com componente presente

3 -> Componente com variância zero.

Fonte: Elaboração própria.

Os dados faltantes, recuperados com filtro de Kalman, geraram uma nova série retropolada (PNADc – Reconstruída 1) bem similar, no tempo de coexistência, com a PNADc – Reconstruída, embora uma diferença de patamar mais persistente ao longo da década de 1980 e 1990 possa ser observado. Em média, a PNADc - Reconstruída é 0,3 ponto percentual inferior à PNADc - Reconstruída 1.

As séries PNADc - Reconstruída e PNADc - Reconstruída 1 estimadas a partir da abordagem de modelos de componentes não observados apresentaram correlação elevada com as séries reconstruídas por Alves e Fasolo (2015), Vaz e Barreira (2016) e Carvalho (2016).¹¹ Os resultados detalhados estão apresentados na Tabela 5.

¹¹ Essa comparação tornou-se possível pois os dados foram gentilmente cedidos pelos analistas Bruno Ottoni Eloy Vaz (FGV/IBRE) e Angelo Marsiglia Fasolo (Banco Central).

Tabela 5 - Correlação entre as séries reconstruídas neste e em outros trabalhos

	PNADc - Reconstruída	PNADc - Reconstruída 1	Vaz e Barreira (2016)	Alves e Fasolo (2015)
PNADc - Reconstruída	1			
PNADc - Reconstruída 1	0,9864	1		
Vaz e Barreira (2016)	0,9528	0,9562	1	
Alves e Fasolo (2015)	0,9245	0,9720	0,9556	1
Carvalho (2016)	0,9090	0,9388	0,9490	0,9434

Fonte: Elaboração própria

Na Figura 4 pode-se observar a evolução temporal das séries reconstruídas. Uma das vantagens da metodologia utilizada neste trabalho reside no fato de que a série reconstruída nos pontos nos quais não há dados faltantes sempre será igual ao dado real.

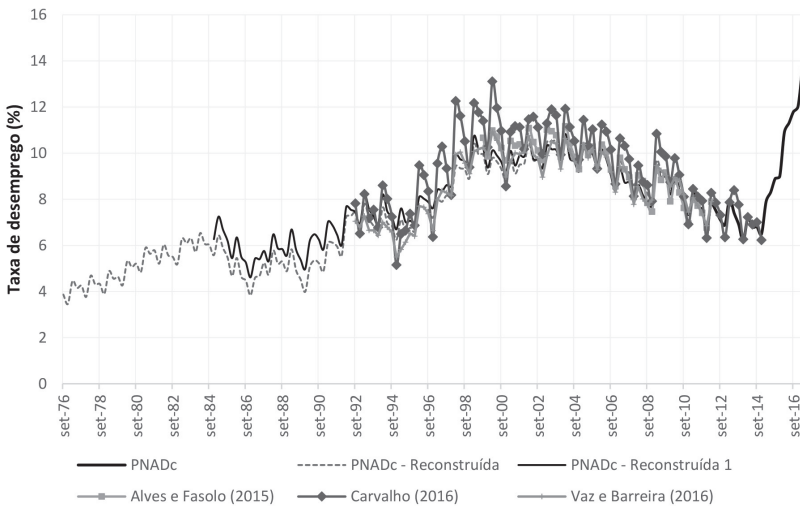


Figura 4 - Comparação da série reconstruída com Alves e Fasolo (2015), Vaz e Barreira (2016) e Carvalho (2016)

Fonte: elaboração própria a partir dos dados cedidos pelos autores.

A observação das séries em conjunto permite afirmar que o patamar atual do desemprego é o mais alto desde 1976. Observando-se a evolução do desemprego ao longo dos anos, o efeito do ciclo recessivo iniciado em 2014 sobre o mercado de trabalho foi mais pronunciado em relação a outros períodos de recessão pelos quais a economia brasileira atravessou, como nos anos de 1981-1983, 1987-1988, 1989-1992, 1995, 1998-1999, 2001, 2003 e 2008-2009 (de acordo com datações estabelecidas pelo Comitê de Datação de Ciclos Econômicos da FGV).¹²

Se adicionarmos os limites de dois desvios padrão da série PNADc – Reconstruída 1 (Figura 5), considerando, portanto, a incerteza associada à estimação, o nível atual seria compatível em magnitude apenas com o limite superior do primeiro trimestre de 1999.

Entre 2014 e 2016, pelo dado oficial do IBGE, a taxa subiu 4,7 pontos percentuais, passando de 6,8% para 11,5% da força de trabalho. Considerando os valores estimados na PNADc - Reconstruída 1, os picos anteriores haviam sido registrados em 1999 (10,1%) e 2003 (10,1%).

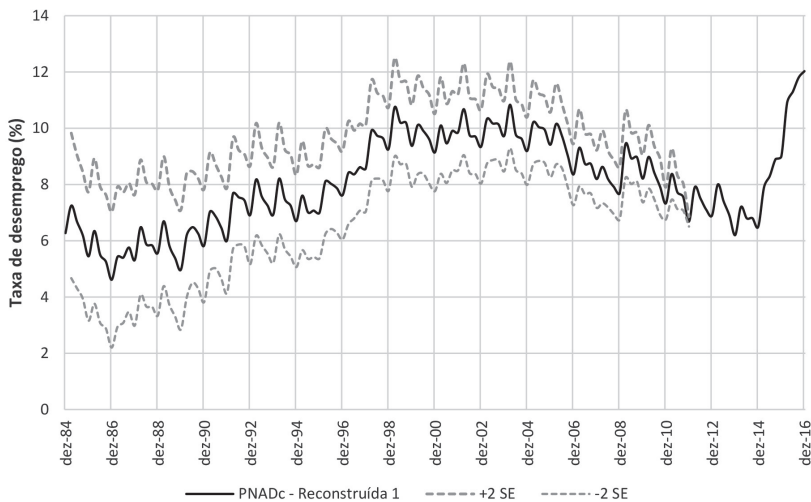


Figura 5 - Série reconstruída e o respectivo intervalo de confiança

Fonte: Elaboração própria.

¹² Ver <http://portalivre.fgv.br/main.jsp?lumChannelId=4028808126B9BC4C0126BEA1755C6C93>. Acessado em 26/04/2017.

7. Comparação com a literatura, extensões e limitações

Alves e Fasolo (2015) utilizam metodologia interessante para reconstruir série com dados faltantes de forma similar a este trabalho. Questões de metodologia econométrica à parte, o prazo de reconstrução é pequeno. A reconstrução inicia-se em 1999.

Já Carvalho (2016) utiliza o algoritmo clássico de interpolação de Deaton (1971) para reconstruir a série da PNAD. O método utilizado permite a comparação de estimativas para os dados faltantes com a série original. Aqui também o espectro temporal de reconstrução não é tão grande quanto o deste trabalho. O período engloba os anos de 1992 a 2014.

Vaz e Barreira (2016) cobrem o período de 1992 até 2012 e utilizam microdados. O período de reconstrução das séries neste tipo de abordagem é limitado pela existência de microdados de qualidade. A inexistência do dado primário limita o período máximo de reconstrução. Apesar da limitação inerente à possível inexistência de fonte primária para períodos muito longos, o trabalho realizado pelos autores é de grande importância ao prover um excelente *benchmark* para avaliação dos resultados aqui apresentados, dado que a PNAD Contínua está disponível apenas para um período muito curto.

A obtenção de uma série harmonizada da PNAD anual entre 1976 e 1991 a partir dos microdados disponibilizados pelo IBGE traria mais conforto sobre os resultados obtidos no primeiro exercício (iniciado em 1976). Esse refinamento seria relevante para minimizar discrepâncias geradas pela mudança de metodologia, sobretudo no que diz respeito à composição e à conceituação da força de trabalho. O trabalho de Barbosa (2014) fornece os subsídios para compatibilizar informações de diferentes períodos, tanto para as pesquisas da PNAD quanto para os Censos Demográficos, mas não temos certeza da possibilidade de estender a metodologia para períodos longos e mais remotos.

Além da possibilidade de trabalho com os microdados, outros passos podem ser listados como sugestões de pesquisa e desdobramentos. Um complemento aos resultados apresentados diretamente para a taxa de desemprego poderia vir da comparação do resultado oriundo de reconstruções das séries da população ocupada e desocupada.¹³

Outro exercício possível de ser conduzido diz respeito à reconstrução da taxa de desemprego também em bases mensais. Embora a pesquisa tenha sido desenhada para produzir informações trimestrais, o IBGE passou a divulgar na frequência recentemente variáveis na forma de trimestres móveis, ou seja, o dado original é agregado temporalmente. Como o agregador temporal da série é conhecido, seria possível aplicar a mesma metodologia na reconstrução da taxa de desemprego mensal. Infelizmente, o IBGE divulga apenas a série agregada temporalmente.

Adicionalmente, a metodologia de modelos de componentes não observados poderia ser utilizada na reconstrução de outras variáveis relevantes da PNAD contínua como o salário médio e a massa de rendimentos que podem ser úteis na medida em que representam subsídios relevantes para o entendimento da dinâmica do consumo, crédito e arrecadação de certos tributos.

8. Conclusão

Este trabalho realizou um exercício para retroagir a taxa de desemprego trimestral da PNAD contínua. A nova pesquisa domiciliar do IBGE é resultado da integração entre a PME e a PNAD anual com série histórica iniciada apenas em 2012. Apesar dos ganhos derivados do aperfeiçoamento conceitual e de abrangência geográfica em nível nacional, a decisão do IBGE trouxe um custo associado à redução da abrangência do espectro temporal das séries históricas em alta frequência que só serão recuperados com o tempo.

¹³ O IBGE não disponibiliza as informações dos valores absolutos em termos harmonizados.

Os dados faltantes, anteriores ao primeiro trimestre de 2012, foram reconstruídos a partir de modelos de componentes não observados. O exercício permite analisar a dinâmica histórica do mercado de trabalho de maneira mais apropriada.

Apesar da ampliação do horizonte temporal, uma ressalva importante a ser feita com relação à série reconstruída a partir de 1976 diz respeito à utilização de dados da PNAD anual não ajustadas completamente à revisão metodológica de 1992. De todo modo, os resultados obtidos mostraram-se condizentes com os encontrados em outros trabalhos já publicados que exploraram o mesmo problema com outras metodologias.

Como conclusão derivada do trabalho, a taxa de desemprego na métrica da PNAD contínua atingiu, na amostra analisada, seu patamar mais elevado da série histórica retropolada desde 1976 até o final de 2016, justamente no final do período superando em magnitude os resultados obtidos em outros períodos recessivos da história econômica em quatro décadas.

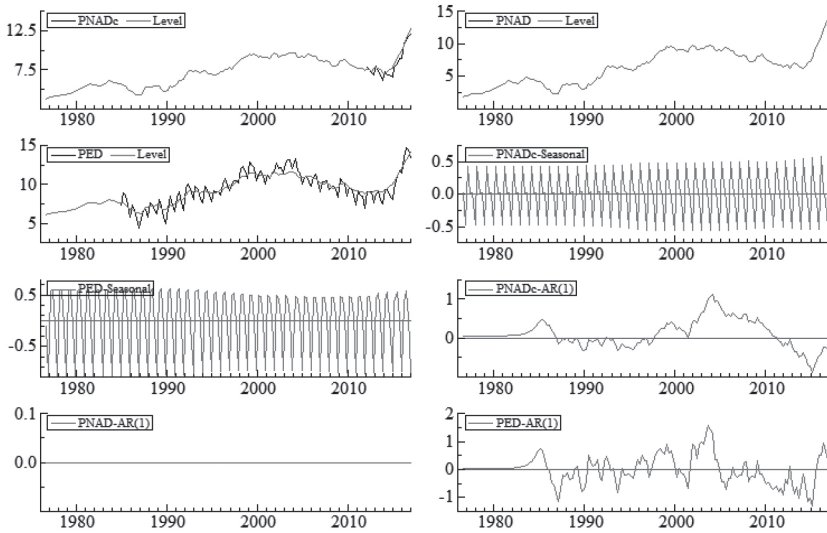
Referências

- Alves, Sergio Afonso Lago e Fasolo, Angelo Marsiglia. *Not Just Another Mixed Frequency Paper*. Brasília: Banco Central do Brasil, 2015. (Working Series Paper, no. 400).
- Angelini, Elena; Henry, Jérôme; Marcellino, Massimiliano. *Interpolation and backdating with a large information set*. Frankfurt am Main: European Central Bank, 2003. (Working Paper, n. 252).
- Barbosa, Rogério Jerônimo. *Comparabilidade das informações disponíveis nos ss (1960-2010) e PNADs (1976, 1985, 1995 e 2005)*. São Paulo: Centro de Estudos da Metrópole, 2014. (Texto para discussão N. 017/2014).
- Cannata, Rosa. R.; Buono, Dario; Biscosi, Ferdinando. The Macroeconomic Imbalances Procedure and the scoreboard: ensuring data coverage. *EURONA - Eurostat review on National Accounts and Macroeconomic Indicators*, Luxemburgo, no. 2: 101-122, Fev. 2015.
- Carvalho, Sandro Sacchet de. Construção de séries longas de alta frequência de indicadores do mercado de trabalho com a PME e a PNADC. *Carta de Conjuntura*, Brasília, no. 32, Jul./Set. 2016. Nota técnica.
- Chow, Gregory C.; Lin, An-loh. Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The review of Economics and Statistics*, v. 53, no. 4: 372-375, Nov. 1971.
- Commandeur, Jacques F.; Koopman, Siem J. *An introduction to state space time series analysis*. New York: Oxford University Press, 2007. 174 p.
- Denton, Frank T. Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the American Statistical Association*, v. 66, n. 333: 99-102, Mar. 1971.
- Di Fonzo, Tommaso. *Temporal disaggregation of economic time series: towards a dynamic extension*. Luxemburgo: European Commission – Eurostat, 2003. (Working Papers and Studies).

- Dias, Victor Pina; Issler, João Victor; Rodrigues, Cláudia. Interpolação de variáveis fiscais brasileiras usando representação de espaço de estados. In: ENCONTRO BRASILEIRO DE ECONOMETRIA, 32., 2010. Salvador. *Anais eletrônicos...* Salvador. EBE, 2010. Disponível em: < <http://bibliotecadigital.fgv.br/ocs/index.php/sbe/EBE10/paper/viewPDFInterstitial/2342/1136>>. Acesso em: 18 mai. 2017.
- Durbin, James; Koopman, Siem Jan. *Time series analysis by state space methods*. 2nd ed. Oxford: Oxford University Press, 2012, 346 p.
- Harvey, Andrew. *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge: Cambridge University Press, 1989, 554 p.
- Harvey, Andrew C.; Chung, Chia-Hui. Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society*, v. 163, no. 3: 303-309, 2000.
- Harvey, Andrew C.; Koopman, Siem J.; Penzer, Jeremy. Messy time series: a unified approach. *Advances in Econometrics*, New York, v. 13: 103-143, 1998.
- Harvey, Andrew C.; Valls Pereira, Pedro L. The Estimation of dynamic models with missing observation. *Brazilian Review of Econometrics*, v.5, no.2: 81-95, 1985.
- IBGE – Instituto Brasileiro de Geografia e Estatística. *Esclarecimentos sobre os resultados da PNAD contínua produzidos mensalmente*. Rio de Janeiro: IBGE, 2015. (Notas Técnicas, n.1).
- IBGE – Instituto Brasileiro de Geografia e Estatística. *Pesquisa Nacional por Amostra de Domicílios*. Rio de Janeiro: IBGE, 2014. (Notas Metodológicas, v.1).
- IBGE – Instituto Brasileiro de Geografia e Estatística. *Sistema integrado de pesquisas domiciliares - SIPD*. Rio de Janeiro: IBGE, 2007. (Texto para Discussão, no. 24).
- Koopman, Siem Jan; Hoogerheide, Lennart. Analysis of Historical Time Series with Messy Features: The Case of Commodity Prices in Babylonia. In: Van Der Spek, Robartus Johannes; Van Zanden, Jan Luiten; Van Leeuwen, Bas. *A History of Market Performance: From Ancient Babylonia to the Modern World*. Abingdon: Routledge, 2013, cap. 3: 45-67.
- Koopman, Siem Jan et al. *STAMP: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants Press, 2006, 231 p.
- Litterman, Robert B. A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics*, v. 1, no. 2: 169-173, 1983.
- Mattos, Fernando Augusto Mansor de. *Emprego público no Brasil: aspectos históricos, inserção no mercado de trabalho e evolução recente*. Brasília: Ipea, 2011. (Texto para Discussão, no. 1582).
- Moauero, Filippo; Savio, Giovanni. Temporal disaggregation using multivariate structural time series models. *The Econometrics Journal*, v. 8, no. 2: 214-234, Jul. 2005.
- Proietti, Tommaso. *Temporal disaggregation by state space methods: Dynamic regression methods revisited*. Luxemburgo: European Commission – Eurostat, 2005. (Working Papers and Studies).
- Vaz, Bruno Ottoni Eloy; Barreira, Thiago Cabral. *Metodologia de reinterpolação da Pesquisa Nacional por Amostra de Domicílios Contínua de 1992 a 2012*. Rio de Janeiro, RJ: FGV-IBRE, 2016, 13 p. (Nota técnica).

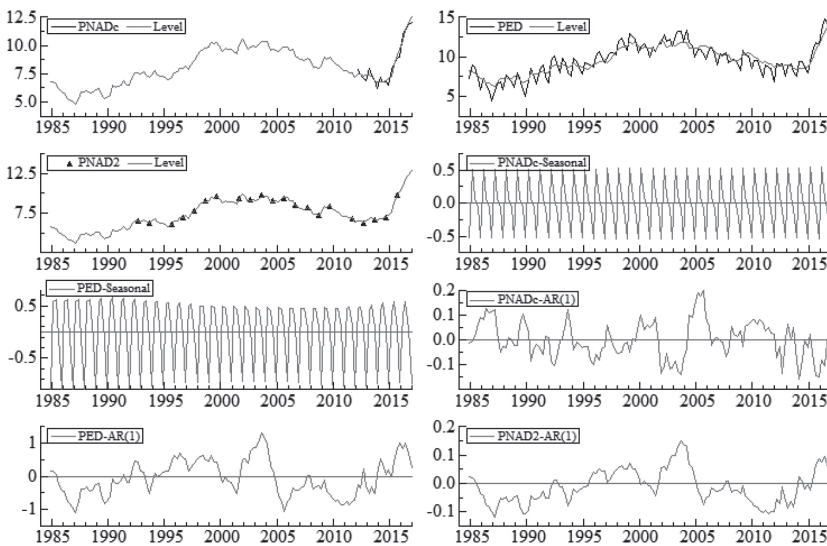
Anexo

Componentes estimados: PNADc - Reconstruída



Fonte: Elaboração própria.

Componentes estimados: PNADc – Reconstruída I



Fonte: Elaboração própria.

Taxa de desemprego média (% da força de trabalho)

	PNADc	PNADc - Reconstruída	PNADc - Reconstruída 1
1977	-	4,16	-
1978	-	4,31	-
1979	-	4,59	-
1980	-	5,11	-
1981	-	5,63	-
1982	-	5,57	-
1983	-	6,10	-
1984	-	6,05	-
1985	-	5,61	6,39
1986	-	4,60	5,43
1987	-	4,81	5,47
1988	-	5,29	5,94
1989	-	4,82	5,72
1990	-	5,13	6,19
1991	-	5,94	6,59
1992	-	7,16	7,39
1993	-	7,06	7,48
1994	-	6,93	7,41
1995	-	6,77	7,17
1996	-	7,51	7,90
1997	-	8,02	8,49
1998	-	9,24	9,64
1999	-	9,84	10,13
2000	-	9,38	9,69
2001	-	9,46	9,83
2002	-	9,82	9,87
2003	-	10,42	10,09
2004	-	10,24	9,86
2005	-	10,02	9,90
2006	-	9,58	9,32
2007	-	9,14	8,74
2008	-	8,46	8,10
2009	-	9,09	8,88
2010	-	8,19	8,16
2011	-	7,52	7,58
2012	7,3	7,3	7,3
2013	7,1	7,1	7,1
2014	6,8	6,8	6,8
2015	8,5	8,5	8,5
2016	11,5	11,5	11,5

Fonte: IBGE. Elaboração própria.