

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

RAFAEL DA ROCHA MENDONÇA BACCIOTTI

**RETROPOLAÇÃO DA TAXA DE DESEMPREGO DA PNAD CONTÍNUA ATRAVÉS
DE MODELOS DE COMPONENTES NÃO OBSERVADOS**

SÃO PAULO

2017

RAFAEL DA ROCHA MENDONÇA BACCIOTTI

**RETROPOLAÇÃO DA TAXA DE DESEMPREGO DA PNAD CONTÍNUA ATRAVÉS
DE MODELOS DE COMPONENTES NÃO OBSERVADOS**

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Mestre em Economia.

Campo de conhecimento: Economia

Orientador: Prof. Dr. Emerson Fernandes Marçal

SÃO PAULO

2017

Bacciotti, Rafael da Rocha Mendonça.

Retropolação da taxa de desemprego da PNAD contínua através de modelos de componentes não observados / Rafael da Rocha Mendonça

Bacciotti. - 2017.

44 f.

Orientador: Emerson Fernandes Marçal

Dissertação (MPFE) - Escola de Economia de São Paulo.

1. Modelos econométricos. 2. Mercado de trabalho - Brasil. 3. Pesquisa nacional por amostra de domicílios. 4. Desemprego. I. Marçal, Emerson Fernandes. II. Dissertação (MPFE) - Escola de Economia de São Paulo. III. Título.

CDU 331.6(81)

RAFAEL DA ROCHA MENDONÇA BACCIOTTI

**RETROPOLAÇÃO DA TAXA DE DESEMPREGO DA PNAD CONTÍNUA ATRAVÉS
DE MODELOS DE COMPONENTES NÃO OBSERVADOS**

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Mestre em Economia.

Campo de conhecimento: Economia

Data da aprovação:

__/__/__

Banca Examinadora:

Prof. Dr. Emerson Fernandes Marçal
(Orientador)
FGV – EESP

Prof. Dr. Pedro Luiz Valls Pereira
FGV – EESP

Prof. Dr. Diogo de Prince Mendonça
UNIFESP

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Dr. Emerson Fernandes Marçal, pela paciência e apoio técnico.

Agradeço à minha família, em especial meus pais e minha querida Mari, pelo carinho e apoio emocional.

Agradeço aos colegas da Tendências Consultoria Integrada, pelo incentivo e apoio financeiro.

Sou grato ainda aos pesquisadores Vandeli Guerra dos Santos (IBGE), Bruno Ottoni Eloy Vaz (FGV/IBRE), Angelo Marsiglia (BCB) e Filippo Moauro (ISTAT), que gentilmente compartilharam informações importantes para a condução deste trabalho; aos membros da banca examinadora, Prof. Dr. Pedro Luiz Valls Pereira e Prof. Dr. Diogo de Prince Mendonça, pelas comentários e sugestões e aos amigos Igor Velecico e Antonio Morales pelas elucidações.

RESUMO

A análise do mercado de trabalho em perspectiva histórica com base em séries de alta frequência no Brasil é uma tarefa desafiadora, pois não há uma pesquisa longa, abrangente e ao mesmo tempo compatível em termos metodológicos e conceituais que permita acompanhar o desempenho das diversas variáveis de maneira adequada. Essas questões foram exacerbadas no início de 2016, quando a Pesquisa Mensal de Emprego (PME) foi interrompida pelo IBGE. Desde então, a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) passou a ser a única referência do instituto sobre o mercado de trabalho em alta frequência. Nesse contexto, este trabalho tem como objetivo retroagir a taxa de desemprego trimestral da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), iniciada em março de 2012. Duas séries foram geradas com base na metodologia de modelos de espaço de estados e o filtro de Kalman: uma iniciada em 1976 e outra em 1984, utilizando-se como referência, nos dois casos, as taxas de desemprego obtidas na PNAD anual e na Pesquisa de Emprego e Desemprego (PED - Dieese).

Palavras-chave: mercado de trabalho; taxa de desemprego; PNAD Contínua; espaço-estado; modelos de componentes não observados; filtro de Kalman

ABSTRACT

The analysis of the labor market in historical perspective in Brazil is a challenging task, since there is not a long, comprehensive and at the same time methodologically and conceptually compatible survey that allows monitoring the performance of the various variables in an appropriate manner. These issues were exacerbated in early 2016, when the Pesquisa Mensal de Emprego (PME) was interrupted by IBGE. Since then, the Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) has become the Institute's only reference to the high-frequency labor market monitoring. In this context, the objective of this study is to retropolate the unemployment rate from PNAD Contínua, begun in March 2012. Two series were produced from the state space model methodology, with the Kalman filter, one since 1976 and another since 1984, using as reference the annual PNAD and the Pesquisa de Emprego e Desemprego (PED-Dieese).

Keywords: job market; PNAD Contínua; unemployment rate; state space method; unobserved components time series models; Kalman filter

LISTA DE FIGURAS

FIGURA 1. TAXA DE DESEMPREGO (%) DESDE 1950	27
FIGURA 2. TAXA DE DESEMPREGO (%) DESDE 2012	27
FIGURA 3. RETROPOLAÇÃO INICIADA EM 1984	31
FIGURA 4. INTERVALO DE CONFIANÇA DA SÉRIE RECONSTRUÍDA	31
FIGURA 5. PNADC REC 1 (1984) E PNADC REC 2 (1976).....	34
FIGURA 6. COMPARAÇÃO COM ALVES E FASOLO (2015) E VAZ E BARREIRA (2016).....	35

LISTA DE TABELAS

TABELA 1: RESUMO DAS CARACTERÍSTICAS DAS PESQUISAS	26
TABELA 2: PRINCIPAIS ESTATÍSTICAS DESCRITIVAS	28
TABELA 3: MODELO MULTIVARIADO PNAD REC 1	30
TABELA 4: MODELO MULTIVARIADO PNAD REC 2	33

SUMÁRIO

1. INTRODUÇÃO E MOTIVAÇÃO	10
2. DISCUSSÃO DA LITERATURA	11
2.1 RETROPOLAÇÃO COMO UM PROBLEMA DE DADOS COM IRREGULARIDADES	11
2.2 MODELOS DE COMPONENTES NÃO OBSERVADOS	13
2.3 RETROPOLAÇÃO DE SÉRIES PARA O MERCADO DE TRABALHO BRASILEIRO	15
3. METODOLOGIA DE ESTIMAÇÃO.....	17
4. FONTES DE INFORMAÇÃO SOBRE A FORÇA DE TRABALHO NO BRASIL	19
4.1 ASPECTOS GERAIS.....	19
4.2 PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS CONTÍNUA	21
4.3 PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS.....	22
4.4 CENSOS DEMOGRÁFICOS	23
4.5 PESQUISA MENSAL DE EMPREGO.....	24
4.6 PESQUISA DE EMPREGO E DESEMPREGO.....	25
4.7 BASE DE DADOS E ESTATÍSTICAS DESCRITIVAS	26
5. RESULTADOS	29
5.1 RETROPOLAÇÃO INICIADA EM 1984	29
5.2 RETROPOLAÇÃO INICIADA EM 1976	32
5.3 COMPARAÇÕES	34
6. POSSÍVEIS EXTENSÕES E LIMITAÇÕES	37
7. CONCLUSÃO.....	38
8. REFERÊNCIAS.....	38
ANEXOS	42

1. Introdução e motivação

A análise do mercado de trabalho em perspectiva histórica com base em séries de alta frequência no Brasil é uma tarefa desafiadora, pois não há uma pesquisa longa, abrangente e ao mesmo tempo compatível em termos metodológicos e conceituais que permita acompanhar o desempenho de variáveis relevantes, como a taxa de desemprego¹, de maneira adequada.

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, por exemplo o estudo da relação entre a inflação e a taxa de desemprego (Curva de Phillips).

Essas questões foram exacerbadas no início de 2016, quando a Pesquisa Mensal de Emprego (PME) foi interrompida 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 a PNAD anual, passou a ser a única referência do instituto na produção e disseminação de estatísticas sobre o mercado de trabalho em alta frequência².

A PNAD Contínua é mais abrangente que a PME, cuja cobertura se restringia a seis regiões metropolitanas (Belo Horizonte, Porto Alegre, Recife, Rio de Janeiro, Salvador e São Paulo) e, ao contrário da PNAD publicada uma vez por ano (mas referente a setembro), 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 e da atualização e modernização metodológica e conceitual em acordo às 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.

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.

Diante da necessidade de séries longas e bases homogêneas para as análises históricas e modelagens econométricas, o presente trabalho tem como objetivo contribuir com a retropolação (ou *backcalculation*)³ 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 utiliza a taxa de desemprego da PNAD desde 1976 e da Pesquisa de Emprego e Desemprego (PED), publicada pelo Departamento Intersindical de Estatística e Estudos Socioeconômicos (Dieese), desde 1984, como variáveis indicadoras da dinâmica da PNAD Contínua. 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 via filtro de Kalman.

O artigo está organizado do seguinte modo. Após esta introdução, a seção 2 contextualiza o problema da retropolação 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 à retropolação de séries históricas para o mercado de trabalho brasileiro. A parte 3 descreve brevemente a metodologia utilizada. A parte 4 apresenta as características de cada pesquisa e a base de dados utilizada na estimação. Os resultados são apresentados na parte 5, enquanto as últimas seções expõem as limitações e possíveis extensões do exercício, e as conclusões do trabalho.

2. Discussão da literatura

2.1 Retropolação como um problema de dados com irregularidades

A retropolação pode ser contextualizada como parte de um problema comum na análise de séries temporais denominadas *messy times series* - em Harvey, Koopman e Penzer (1998). O termo é atribuído às variáveis que apresentam, por exemplo, valores faltantes, *outliers* e quebras estruturais.

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).

Particularmente, a estimação de observações faltantes (*missing observation*) ao longo de uma série cuja natureza é de estoque, como a taxa de desemprego, caracteriza um exemplo de interpolação quando os pontos ausentes estão situados entre dois pontos disponíveis na série histórica; de extrapolação quando existe a necessidade de se estimar informações desconhecidas para um período além da amostra de dados realizados, ou de retropolação quando o objetivo é estimar valores passados, anteriores ao período conhecido.

O mesmo problema de estimação de valores faltantes no caso em que a variável tem natureza de fluxo é conhecido como distribuição (ou desagregação temporal). Um exemplo conhecido na literatura é a construção de uma série mensal (de alta frequência), como um indicador mensal de atividade econômica, a partir de uma variável construída originalmente em termos trimestrais (de menor frequência), como o PIB, levando em conta a restrição de que a soma ou a média dos três meses na série construída corresponda ao valor do trimestre na série original.

O tratamento técnico para contornar essas questões - criar séries de maior frequência condizentes com séries de menor frequência, interpolar informações faltantes ou retropolar quando apenas dados recentes estão disponíveis - é similar (Angelini, Henry e Marcellino, 2003) e pode ser feito sob diferentes metodologias, basicamente dentro de duas abordagens⁴.

A primeira abrange métodos que não consideram a utilização de séries relacionadas, baseando-se, ou em critérios matemáticos ou em modelos ARMA⁵. A segunda diz respeito aos métodos que lançam mão de indicadores relacionados com a série a ser construída, como os procedimentos de Denton (1971), Chow e Lin (1971) e suas extensões conduzidas por Fernández (1981) e Litterman (1983), além dos modelos formulados em termos de componentes não observados na forma espaço-estado.

O objetivo principal das técnicas da segunda abordagem é criar uma nova série temporal que seja consistente com os dados de baixa frequência (*benchmarks*) e que preserve o comportamento de curto prazo da variável relacionada.

O método de ajuste matemático de Denton (1971), restrito ao problema de interpolação ou desagregação temporal, consiste em minimizar as diferenças entre a

⁴ Conforme Moauro e Savio (2005), Di Fonzo (2003a) 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).

série de menor frequência estimada e a série indicadora, dada uma restrição de agregação. No método estatístico de Chow e Lin (1971), que também pode ser utilizado para extrapolação ou retropolação, a série é construída a partir dos parâmetros estimados em uma regressão com as variáveis relacionadas, assumindo-se que os termos de erros são serialmente correlacionados⁶.

De maneira um pouco distinta de Chow e Lin (1971), cuja natureza é estática, Santos Silva e Cardozo (2001) propõe um modelo dinâmico atribuindo estrutura para as variáveis de alta frequência (com inclusão de um parâmetro auto-regressivo).

Di Fonzo (2003b) aplica o método dinâmico, propondo algumas adaptações, na retropolação das séries trimestrais de consumo norte-americano entre 1953 e 1957. Nesse período, havia disponibilidade dos dados anuais de consumo e trimestrais da renda disponível (variável relacionada). O autor ainda menciona que os modelos formulados em termos de componentes não observados representados em espaço-estado também podem ser utilizados para a reconstrução de séries temporais, fazendo menção a trabalhos da Insee (*Institut national de la statistique et des études économiques*) aplicados às séries de contas nacionais francesas.

2.2 Modelos de componentes não observados

Alguns trabalhos que aplicam este arcabouço serão descritos na sequência. Do ponto de vista teórico é válido mencionar que a utilização de modelos espaço-estado para a estimação de *missing observation* pode ser vista em Harvey e Pierse (1984) e, posteriormente, já no arcabouço dos modelos estruturais de componentes não observados em Harvey (1989) - de maneira aprofundada -, e em Commadeur e Koopman (2007) - em modo introdutório.

Koopman e Hoogerheide (2013) seguem a metodologia dos modelos estruturais para analisar conjuntamente a dinâmica de preços de commodities disponíveis para a economia babilônica⁷ durante o período de 385 a 61 a.c. A amostra disponível é irregular, de modo que diversos pontos ao longo do tempo foram tratados como faltantes e estimados pelo filtro de Kalman.

⁶ “A suposição de não correlação serial nos resíduos nas estimativas de alta frequência não é suportada pela evidência empírica. O processo Chow-Lin propõe um método para estimar a matriz de covariância sob a suposição de que os erros seguem um processo autoregressivo de primeira ordem AR(1)” (Chen, 2007, p. 16). Tradução livre.

⁷ Disponível em <http://www.iisg.nl/hpw/babylon.php>. Acesso em 27/02/2017

A análise foi realizada através de decomposições univariadas (de cada série de preço individualmente sem o uso de variáveis indicadoras), mas também foi conduzida pela extensão multivariada dos modelos estruturais conhecida como *seemingly unrelated time series equations models* (SUTSE), estimulada pela presença de características comuns na evolução desses preços.

A generalização multivariada nos modelos estruturais, seguindo a categorização proposta acima (sem o uso ou com o uso de variáveis relacionadas), se encaixa na segunda abordagem.

Embora o termo “relacionado” tenha um significado pouco diferente. Assume-se, nesse contexto, que a série a ser desagregada e o conjunto de variáveis relacionadas são afetadas por um ambiente similar. Consequentemente, elas devem se mover juntas, embora nenhuma delas necessariamente cause a outra (Moauero e Savio, 2005, p.3).⁸

Harvey e Chung (2000) aplicam a metodologia para estimar a variação da taxa de desemprego trimestral da *Labour Force Survey* – LFS no Reino Unido com a utilização de uma medida relacionada com o desemprego (número de pedidos de auxílio-desemprego), mas disponível em termos mensais, e tiveram ganhos de performance 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%.

Moauero e Savio (2005), na linha de Harvey e Chung (2000), utilizam modelos multivariados para lidar com a interpolação e a distribuição de séries de atividade econômica de países da Organização para a Cooperação e Desenvolvimento Económico (OCDE). Também encontram evidências de melhor performance (através da comparação de estatísticas como o erro quadrático médio) frente a outras abordagens, como o próprio modelo estrutural univariado e os métodos de Denton (1971), Chow-Lin (1971) e Litterman (1983).

Moauero (2014) aplica a abordagem multivariada para derivar uma série mensal de emprego para a Zona do Euro, entre janeiro de 1992 e agosto de 2012. A Eurostat

⁸ Tradução livre

divulga informações trimestrais de emprego nas contas nacionais e uma série mensal de desemprego, que foi utilizada como variável relacionada na estimação de um modelo com tendência linear local. As estimações foram realizadas para diversos cortes de tempo e a comparação com outros métodos que também lançam mão de variáveis relacionadas para estimar informações faltantes mostrou que a performance do modelo multivariado foi mais satisfatória ao apresentar erro quadrático médio mais baixo.

2.3 Retropolação de séries para o mercado de trabalho brasileiro

Vale dizer que a metodologia de modelos de componentes não observados também já foi aplicada em trabalhos nacionais. Especificamente sobre a retropolação de dados no mercado de trabalho, 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. As estimações dos valores anteriores a 2002 foram realizadas a partir de decomposições univariadas com a adição de variáveis de controle disponíveis em toda a extensão temporal – os autores utilizaram o PIB e a produção de energia para retropolar o nível de emprego e a produção de papel ondulado e a produção de cimento na regressão da renda.

Alguns estudos, por sua vez, 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 vetor auto regressivo 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.

Os autores, em primeiro lugar, ajustaram a taxa de desemprego da PNAD anual assumindo que a diferença de nível em relação à PNAD Contínua decorre de um fator multiplicativo. Assim, construíram uma variável ajustada da PNAD anual desde 1999, que foi encadeada às observações mensais da PNAD Contínua e utilizada como variável dependente nos modelos MF-VAR - em um total de trinta e seis simulações.

No processo de estimação dos pontos faltantes da variável dependente foram consideradas como variáveis explicativas as taxas de desemprego e as taxas de participação⁹ obtidas na PME e na PED, além do saldo de emprego formal (obtido no Cadastro Geral de Emprego e Desemprego do Ministério do Trabalho - CAGED) e do número de pessoas ocupadas no setor industrial (da Confederação Nacional da Indústria - CNI).

Vaz e Barreira (2016), por outro lado, reconstruíram historicamente as informações da PNAD Contínua utilizando as informações da PME e da PNAD anual. Além da taxa de desemprego, os autores prolongaram também a série de salários. Em um primeiro momento, o estudo tratou de minimizar as diferenças entre a PNAD Contínua e a PNAD anual a partir de uma compatibilização dos conceitos das variáveis no período em que elas coexistem. O nível da PNAD anual para os anos de 2012, 2013 e 2014 foi aproximado ao da PNAD Contínua através de ajustes realizados nos microdados da pesquisa para os seguintes critérios: população em idade ativa; trabalho remunerado; trabalho na produção para próprio consumo/uso e trabalhador; e trabalhador afastado temporariamente.

No processo de retropolação, os autores realizaram novamente ajustes pelos microdados para ajustar o nível da PNAD anual nos anos anteriores à PNAD Contínua (de 1992 a 2011) e construíram uma série mensal a partir de variações da PME, com correções necessárias para garantir os níveis da PNAD em setembro.

Já Carvalho (2016) propõe a utilização da técnica de interpolação de Denton (1971) 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, assumindo-se que são “os dados que estimam com melhor precisão os valores populacionais dos indicadores de mercado de trabalho” (CARVALHO, 2016, p. 5).

As séries interpoladas mostraram-se consistentes com as séries reais, uma vez que as variáveis da PNAD e da PME foram aproximadas conceitualmente à PNAD Contínua também através de uma aproximação conceitual pelos microdados. As

⁹ A taxa de participação é a relação entre as pessoas economicamente ativas (PEA), que compõem a força de trabalho, e as pessoas em idade ativa (PIA).

séries longas iniciadas em 1992 resultam do encadeamento das séries reais da PNAD Contínua com as séries interpoladas (no período anterior a 2012).

3. Metodologia de estimação

A ideia por trás dos modelos estruturais de componentes não observados é que as séries de tempo podem ser descritas como a soma de componentes não observados, como tendência (nível e inclinação), ciclo (para captar movimentos regulares da série), sazonalidade (movimentos periódicos) e um componente irregular (movimentos aleatórios).

Esses componentes podem ser modelados de diversas formas. A tendência, componente de longo prazo da série que 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. De maneira equivalente ao modelo clássico de regressão, é possível dizer que o nível cumpre o papel de intercepto, enquanto a inclinação corresponde ao coeficiente angular da linha de regressão em relação ao eixo x .

Uma vantagem do modelo estrutural diz respeito à possibilidade de os parâmetros associados às variáveis de estado mudarem no tempo, como um processo estocástico, ao invés de serem determinísticos (Commadeur e Koopman, 2007).

A frequência sazonal, quando apropriada, pode ser modelada através de variáveis *dummies* ou por uma função trigonométrica, fixa ou variável.

O ciclo presente em diversas séries pode ser especificado por uma forma trigonométrica com curta, média ou longa duração ou como um processo autorregressivo. O ciclo estocástico é construído a partir de uma perturbação dada por um choque e a introdução de um fator de amortecimento para forçar um processo estacionário.

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

$$y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma^2_\varepsilon) \quad (1)$$

em que μ_t representa a tendência, γ_t a sazonalidade, ψ_t o ciclo e ε_t o componente irregular (erro observacional).

O tratamento estatístico é conduzido pela representação na forma espaço-estado, que consiste de uma equação de medida e uma equação de transição, definidas a seguir, conforme Durbin e Koopman (2012):

$$y_t = Z_t \alpha_t + \varepsilon_t \quad \varepsilon_t \sim N(0, H_t) \quad (2)$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \eta_t \sim N(0, Q_t) \quad (3)$$

$$\alpha_1 \sim N(a_1, P_1) \quad (4)$$

A equação de medida (2) descreve a relação entre a série observada (y_t) e o vetor α_t que contém as variáveis de estado não observadas (tendência, ciclo, sazonalidade). A equação de transição (3), por sua vez, exibe a dinâmica dos componentes não observados no tempo. As premissas do modelo são que os termos de erro das equações de medida e de transição (ε_t e η_t) sejam independentes e que as matrizes Z_t , T_t , R_t , H_t e Q_t sejam inicialmente conhecidas.

A mensuração do estado não observado a partir das observações e a estimação dos parâmetros desconhecidos é conduzida pelo filtro de Kalman, que de acordo com Koopman *et al.* (2006), desempenha um papel correlato aos mínimos quadrados no modelo clássico de regressão.

Os parâmetros de (2) e (3) são estimados por máxima verossimilhança pela atualização recursiva do filtro de Kalman, descrita nas seguintes equações:

$$v_t = y_t - z_t a_t \quad (5)$$

$$a_{t+1} = T_t a_t + K_t v_t \quad (6)$$

Sendo que v_t representa o erro de previsão um passo a frente e a_t o estado filtrado no período t .

O objetivo do filtro do Kalman é encontrar valores ótimos da estimativa do estado filtrado no próximo período (a_{t+1}) a partir das informações disponíveis até o período corrente t , que dependerá da estimativa no período anterior (a_t) baseada nas observações passadas, mas será atualizada pela incorporação da inovação (v_t). O grau de influência que a inovação exercerá

sobre a estimativa a_{t+1} dependerá do termo K_t , denominado ganho de Kalman (Orair e Silva, 2012, p.13).

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

Uma propriedade conveniente do processo de estimação sequencial pelo filtro de Kalman é a capacidade para se lidar com as informações faltantes de uma série histórica. Segundo Koopman e Hoogerheide (2013), os dados faltantes podem estar em toda a amostra, individualmente ou em blocos. Nessas situações, quando y_t é tida como faltante, a estimação do estado filtrado é realizada considerando v_t e K_t iguais a zero. De maneira análoga, quando o objetivo é fazer previsão, as projeções são obtidas pelo tratamento de $t, t + 1, t + 2 \dots$ como *missing* (Commadeur e Koopman, 2007).

A análise dos componentes não observados também pode ser realizada por modelos multivariados, quando y_t é um vetor de $p \times 1$ observações. Nos modelos estruturais multivariados, conhecidos por *seemingly unrelated times series models* (SUTSE), as p diferentes séries são modeladas como no caso univariado, mas dependem de componentes não observados que são vetores também. A relação entre as séries ocorre através das correlações dos distúrbios de cada componente.

Dentro desse ambiente existe a possibilidade de se considerar que a tendência, o ciclo e a sazonalidade das diferentes séries sejam comuns, a partir da imposição de restrições sobre a matriz de covariância dos distúrbios de cada componente.

A aplicação do filtro de Kalman é válida na generalização multivariada dos modelos espaço-estado, e a análise 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).

4. Pesquisas sobre a força de trabalho no Brasil

4.1 Aspectos gerais

O IBGE realizou ao longo dos anos de 2006 e 2014 dezessete fóruns de discussão com especialistas na análise de indicadores socioeconômicos e usuários das informações estatísticas com o objetivo de avançar no projeto de reformulação metodológica e de integração de suas pesquisas amostrais domiciliares, denominado Sistema Integrado de Pesquisas Domiciliares – SIPD¹⁰.

A construção do SIPD resultou em termos práticos para o acompanhamento do tema trabalho e rendimentos na PNAD Contínua no final de 2011, a partir da integração da PME e da PNAD. Em linhas gerais este sistema integrado tem por finalidade aperfeiçoar e modernizar os processos de produção das estatísticas derivadas das pesquisas amostrais domiciliares. Dentre as motivações que envolvem sua criação é possível listar a divulgação de indicadores nacionais para o acompanhamento conjuntural do mercado de trabalho, o que atende à demanda de pesquisadores e analistas quanto à possibilidade de se avaliar impactos de curto prazo derivados de políticas públicas e acompanhar as flutuações da massa de salários da economia, por exemplo; a produção contínua de informações sobre consumo e orçamento das famílias¹¹; e a harmonização metodológica e conceitual das variáveis investigadas conforme recomendações atualizadas da OIT e da Comissão de Estatísticas das Nações Unidas (IBGE, 2007).

Apesar dos ganhos de qualidade e abrangência, as mudanças no âmbito do SIPD, sobretudo com a descontinuidade da PME, resultaram no encurtamento das séries históricas de alta frequência (que permitem a análise conjuntural do mercado de trabalho), o que tende a limitar análises comparadas da evolução do mercado de trabalho, bem como dificultar a elaboração de análises econométricas dada a redução dos graus de liberdade.

É válido mencionar, nesse contexto, que as pesquisas naturalmente passam por reformulações e aperfeiçoamentos ao longo dos anos. A PED do Dieese é única fonte cuja metodologia permaneceu inalterada desde seu início em 1984, enquanto 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

¹⁰ As apresentações e os textos produzidos para os fóruns estão disponíveis em <http://www.ibge.gov.br/home/estatistica/indicadores/sipd/default.shtm>. Acesso em 20/02/2017

¹¹ Também no âmbito do SIPD, a Pesquisa de Orçamento Familiar – POF, divulgada normalmente a cada cinco anos, fonte para a análise dos hábitos de consumo das famílias e utilizada no cálculo da estrutura de ponderações do Índice de Preços ao Consumidor Amplo - IPCA, passará a ser publicada anualmente pelo IBGE com o nome de POF Simplificada.

séries antes e depois das revisões não é uma prática comumente ofertada pelo IBGE, tarefa que acaba sendo realizada de forma independente pelos usuários das estatísticas.

A seguir serão exploradas, de maneira sintetizada, as características gerais de cada pesquisa para o entendimento da composição da taxa de desemprego em cada uma delas e a maneira como as séries derivadas de pesquisas com periodicidades distintas (mensais, trimestrais ou anuais) foram incorporadas ao exercício de retroposição da PNAD Contínua.

4.2 Pesquisa Nacional por Amostra de Domicílios Contínua

A PNAD Contínua iniciada em 2012 é a primeira pesquisa produzida no âmbito do SIPD, permitindo a investigação conjuntural do tema trabalho e rendimentos. Ela é realizada através de uma amostra probabilística de domicílios e foi construída para produzir informações trimestrais. Em cada trimestre, o IBGE investiga aproximadamente 211.00 domicílios em 16.000 setores censitários e 3.500 municípios, com representatividade para os diversos níveis geográficos: Brasil, Grandes Regiões, Unidades da Federação e Regiões Metropolitanas que incluem os municípios das capitais.

Na pesquisa, a população investigada para o tema trabalho são as pessoas com 14 anos ou mais de idade. São definidas como ocupadas as pessoas que ao menos durante uma hora completa na semana trabalharam em algum emprego remunerado (dinheiro ou produtos), ou de forma independente (conta-própria ou empregador) ou sem remuneração em auxílio de algum membro da família na produção de bens e serviços.

Também são considerados ocupadas as pessoas temporariamente afastadas do trabalho remunerado (por motivos de férias, jornada variável, licença maternidade, por motivo de doença etc.). Não são consideradas ocupadas as pessoas que trabalharam na produção e na construção para consumo próprio, nem aquelas que trabalharam sem remuneração para instituições religiosas, beneficentes ou de cooperativismo.

As pessoas desocupadas são aquelas que na semana de referência estavam sem trabalho, mas tinham disponibilidade de assumir uma posição e que no período de 30 dias haviam tomado algum tipo de providência efetiva para conseguir uma

colocação. O conjunto de pessoas ocupadas e desocupadas compõe a força de trabalho na economia. (Hypólito, 2010)

Na PNAD Contínua, portanto, a taxa de desemprego é definida como o percentual de pessoas desocupadas na semana de referência (semana que precede a semana de entrevista) em relação à força de trabalho composta por pessoas de 14 anos ou mais.

A série histórica, obtida no Sistema IBGE de Recuperação Automática (SIDRA), está compreendida no intervalo do primeiro trimestre de 2012 até o quarto trimestre de 2016.

4.3 Pesquisa Nacional por Amostra de Domicílios

A PNAD, pesquisa que investiga as características socioeconômicas da população, foi iniciada em 1967 abrangendo um conjunto restrito de regiões do País com realização trimestral. A sua frequência passou a ser anual a partir de 1971, e não é levada a campo nos anos em que o IBGE realiza os censos demográficos.

Apesar da abrangência na coleta de informações também ser nacional, o número de domicílios investigados na edição de 2012 foi 147.203 distribuídos em 9.116 setores censitários e 1.100 municípios, abaixo, portanto do escopo da PNAD Contínua.

Na pesquisa, a população investigada para o tema trabalho são as pessoas com 10 anos ou mais de idade. A definição de pessoas ocupadas é próxima à da PNAD Contínua, com pequenas diferenças como não haver restrições quanto ao número mínimo de horas trabalhadas na semana; considerar como ocupada quem trabalha sem remuneração em ajuda a instituição religiosa, beneficente ou de corporativismo, e também quem trabalha na produção para o próprio consumo ou construção para o próprio uso.

Já as pessoas desocupadas são aquelas que na semana de referência estavam sem emprego e que haviam tomado algum tipo de providência para conseguir trabalho. (IBGE, 2014)

Os dados para a taxa de desemprego da PNAD de 1976 a 1991 foram obtidos na base de dados da OIT, embora não sejam estritamente compatíveis com as informações a partir de 1992, momento em que o IBGE ampliou a abrangência do conceito de trabalho, ao incluir as atividades destinadas para o próprio consumo ou

construção para o próprio uso. Nessa oportunidade, o IBGE também reduziu para 10 anos o limite de definição para as pessoas em idade ativa, ao invés de 15 como era anteriormente (Vargas, 2008), embora na OIT os dados já estão ajustados para o universo de 10 anos ou mais.

A exclusão das atividades para próprio consumo ou uso na PNAD Contínua no conceito de ocupação, como visto anteriormente, condiz com a maioria das pesquisas sobre a força de trabalho no mundo, que não consideram como ocupadas as pessoas que realizam essas atividades de maneira exclusiva (Hypólito, 2010).

Vale dizer ainda que ao longo dos anos, a pesquisa teve seu escopo de abrangência geográfica estendido, e se tornou representativa de todo o País apenas em 2004, com a inclusão das áreas rurais dos estados do Acre, Amazonas, Pará, Rondônia, Roraima e Amapá (Barbosa, 2014).

Com o intuito de permitir uma comparação mais precisa no tempo das variáveis investigadas, o IBGE disponibilizou um conjunto de dados harmonizados iniciados em 1992 com a cobertura geográfica da PNAD existente até 2003¹².

A taxa de desemprego captada na pesquisa, portanto, diz respeito ao percentual de pessoas desocupadas na semana de referência (última semana de setembro) em relação à população economicamente ativa composta por pessoas de 10 anos ou mais. O único ponto de informação da PNAD em cada ano (relativo a setembro) foi encaixado em todo terceiro trimestre.

4.4 Censos Demográficos

Os Censos Demográficos representam a principal fonte de dados sobre a população e os domicílios em todo o território nacional. O primeiro recenseamento foi realizado no Brasil em 1872, antes mesmo da criação do IBGE (em 1936), e foi seguido pelas versões de 1890, 1900, 1920, 1940, 1950, 1960, 1970, 1980, 1991, 2000 e 2010. A tabulação da população economicamente ativa passou a ser iniciada na publicação de 1950¹³.

¹² 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 Guerra dos Santos.

¹³ Ver http://biblioteca.ibge.gov.br/visualizacao/periodicos/67/cd_1950_v1_br.pdf. Acesso em 07/03/2017

Como descrito em Barbosa (2014), o questionário até 1991 perguntava sobre a condição de atividade e ocupação exercida durante os doze meses anteriores à realização do Censo, o que significava que ao conjunto da população economicamente ativa eram incluídas as pessoas que trabalharam (em alguma atividade remunerada ou como membro não remunerado na ajuda de algum negócio ou estabelecimento) ou procuraram emprego em algum momento do período de 365 dias.

No Censos de 2000, por outro lado, além da ampliação do conceito de trabalho, com a inclusão do trabalho sem remuneração para consumo próprio, houve mudança no período de referência, que deixou de ser o ano anterior e passou a ser a semana (Deddeca et al, 2004).

As pessoas desocupadas nos censos até 1991 eram classificadas como quem não trabalhou nos doze meses anteriores à data da entrevista, mas que tomou alguma providência para encontrar trabalho. A partir de 2000, as pessoas desocupadas eram aquelas que se encontravam sem trabalho na semana de referência (entre 23 a 29 de julho de 2000 e entre 25 a 31 de julho de 2010), mas que estavam disponíveis para assumir uma posição nessa semana e que tomou alguma providência efetiva para conseguir trabalho no período de 30 dias.

Vale dizer que a população-alvo na investigação do tema trabalho e rendimento não sofreu alterações, abrangendo as pessoas com 10 anos ou mais de idade.

Os dados de 1950 a 2010 para os Censos foram obtidos nas bases do IBGE e em Mattos (2011). Cada ponto (relativo a julho) foi encaixado no terceiro trimestre.

4.5 Pesquisa Mensal de Emprego

A PME, também realizada por meio de uma amostra probabilística de domicílios, permitia a análise conjuntural em termos mensais do mercado de trabalho restrita ao conjunto de seis regiões metropolitanas (Belo Horizonte, Porto Alegre, Recife, Rio de Janeiro, Salvador e São Paulo). Implantada em 1980, passou por uma grande revisão metodológica em 2002 e foi encerrada em março de 2016, com a divulgação dos dados referentes a fevereiro de 2016. O tamanho da amostra era de aproximadamente 40.000 domicílios distribuídos em 145 municípios.

Na metodologia adotada a partir de 2002, dentre outras mudanças, a população investigada para o tema trabalho passou a representar as pessoas com 10 anos ou mais de idade (frente à idade mínima de 15 anos na metodologia anterior).

A definição das pessoas ocupadas era bem próxima à PNAD Contínua, englobando as pessoas que na semana de referência, durante pelo menos uma hora, estavam envolvidas com trabalhos remunerados, independente ou sem remuneração direta, ajudando algum membro da família na produção de bens e serviços. A definição da população desocupada também, representando as pessoas que na semana de referência não tinham trabalho, mas estavam disponíveis e que haviam tomado algum tipo de providência no período de referência de 30 dias para conseguir emprego.

De janeiro de 1980 até dezembro de 2012 (na metodologia anterior), a taxa de desocupação representava a percentagem das pessoas desocupadas na semana de referência em relação às pessoas economicamente ativas (compostas por indivíduos com 15 anos ou mais), ao passo que na metodologia reformulada, passou a captar o mesmo percentual englobando pessoas com 10 anos ou mais de idade.

Como a taxa de desemprego representa uma variável de estoque, o dado de março foi encaixado no primeiro trimestre, o dado de junho no segundo e assim por diante. As séries históricas foram obtidas nas bases do Ipeadata.

4.6 Pesquisa de Emprego e Desemprego

A PED também é uma pesquisa domiciliar, desenvolvida pelo Dieese e pela Fundação Sistema Estadual de Análise de Dados – Seade. O tamanho da amostra é de aproximadamente 9.000 domicílios pertencentes aos municípios da Região Metropolitana de São Paulo¹⁴.

Na pesquisa, a população ocupada é definida como aquelas pessoas que possuem um trabalho remunerado em dinheiro ou não remunerado auxiliando nos negócios de familiares, ao passo que a população desocupada representa os indivíduos que se encontram em uma situação involuntária de não-trabalho por falta de oportunidade ou que estão em um trabalho irregular e almejam mudança.

A PED, ao contrário das demais pesquisas que captam o desemprego aberto, investigam também o desemprego oculto pelo trabalho precário e pelo desalento

¹⁴ Notas metodológicas disponíveis em <http://www.dieese.org.br/metodologia/metodologiaPed.html>. Acesso em 12/03/2017

(pessoas sem trabalho que foram desestimuladas a procurar emprego pelas condições do mercado). Para os efeitos das estimações, considerou-se a informação do desemprego aberto (pessoas sem trabalho na última semana que efetivamente procuraram uma colocação no mês anterior à entrevista).

Assim, a taxa de desemprego aberto ao longo de todo o período, se refere ao percentual de pessoas desocupadas no período de referência de 30 dias sobre a força de trabalho, composta por pessoas de 10 anos ou mais.

A evolução da taxa de desemprego na região metropolitana de São Paulo está disponível na frequência mensal, praticamente inalterada em termos metodológicos, desde dezembro de 1984. Os dados foram obtidos no Ipeadata compreendendo o período inicial até dezembro de 2016. A informação de março foi encaixada no primeiro trimestre, o dado de junho no segundo e assim por diante.

A tabela abaixo sintetiza as principais características das pesquisas.

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
Censo	IBGE	Decenal	Nacional	1950 a 2010	10 anos ou mais
PME	IBGE	Mensal	6 Regiões metropolitanas	1984 a 2016	10 anos ou mais**
PED	Dieese	Mensal	Região metropolitana de São Paulo	1984 a 2016	10 anos ou mais

* Com exceção de 1994 e dos anos censitários

** Até 2001 o universo era 15 anos ou mais

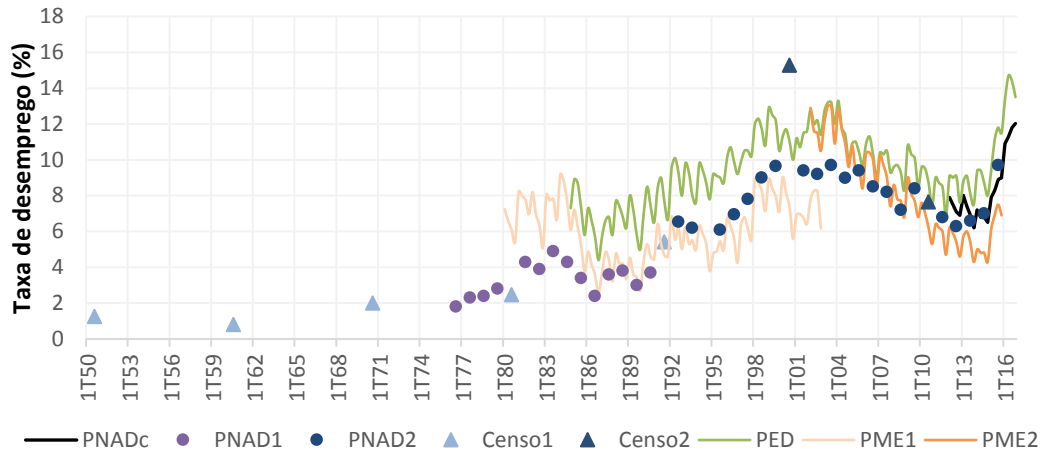
Fonte: IBGE e Dieese. Elaboração própria

4.7 Base de dados e estatísticas descritivas

As variáveis testadas nas estimações foram as taxas obtidas na PNAD, separada em PNAD 1 (1976 a 1991) e PNAD 2 (1992 a 2015); nos Censos Demográficos, separados em Censo 1 (1950, 1960, 1970, 1980 e 1991) e Censo 2 (2000 e 2010); na PME, separada em PME 1 (de 1980 a 2002) e PME 2 (de 2002 a

fevereiro de 2016) e na PED (desde dezembro de 1984). Elas estão dispostas no gráfico abaixo¹⁵.

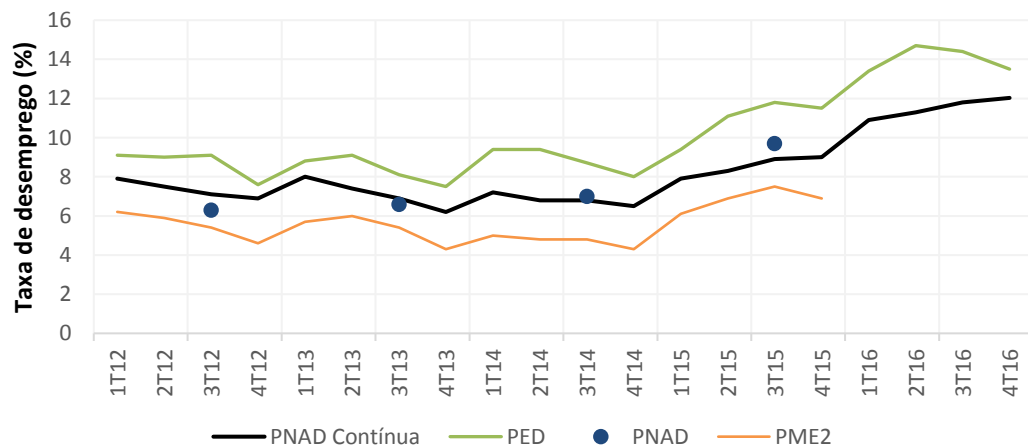
Figura 1. Taxa de desemprego (%) desde 1950



Fonte: Elaboração própria a partir dos dados obtidos no IBGE, na OIT, Dieese e em Mattos (2011).

Apesar da diferença de nível, é possível observar que a tendência e as oscilações na PNAD Contínua, na PED e na PME parecem comuns, o que fica visualmente mais claro na focalização do período em que elas coexistem, de 2012 a 2016 (gráfico 2). Nota-se também que o nível da PNAD Contínua está mais próximo do registrado pela PNAD anual. Essa comparação não é possível com os dados da PME (metodologia anterior) e do Censo Demográfico, pois os últimos registros constam de 2002 e 2010, anteriores ao início da PNAD Contínua.

Figura 2. Taxa de desemprego (%) desde 2012



Fonte: IBGE e Dieese. Elaboração própria

¹⁵ A distinção entre Censo 1 e Censo 2, PNAD 1 e PNAD 2 e PME 1 e PME 2 está associada às mudanças metodológicas promovidas pelo IBGE nessas pesquisas (no Censo de 2000, na PNAD de 1992 e na PME de 2002), já comentadas.

De modo geral, as potenciais fontes de divergência entre as taxas podem decorrer de uma série de fatores associados à cobertura e abrangência geográfica, à conceituação das pessoas ocupadas e desocupadas, ao recorte etário que é considerado na composição da população em idade de trabalhar, às variações nos períodos de referência (se a condição de atividade é medida na semana, no mês ou ao longo dos últimos doze meses) e ainda às diferentes formas de estruturação dos questionários levados a campo (Corseuil e Ramos, 2015).

Em alguns casos, como no ano de 2000 em que o Censo Demográfico mostrou que a taxa de desocupação atingiu 15% da força de trabalho, enquanto a PME registrava 7,3%, as razões das divergências são de difícil compreensão, intrigando especialistas como Neri (2002) e Pastore (2002).

Abaixo, são apresentadas as principais estatísticas descritivas para cada uma das fontes investigadas.

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 1	14	3.3	0.9	1.8	4.9
PNAD 2	21	8.0	1.3	6.1	9.7
PME 1	84	6.0	1.7	2.6	9.2
PME 2	56	8.1	2.5	4.3	13.0
Censo 1	5	2.4	1.8	0.8	5.4
Censo 2	2	11.5	5.4	7.6	15.3
PED	129	9.5	2.1	4.4	14.7

Fonte: Elaboração própria.

Vale mencionar que os patamares mais baixos da taxa de desocupação verificados, sobretudo, até o início da década de 1990, estiveram associados com a grande capacidade de absorção de mão-de-obra na economia brasileira, como argumentaram Sabóia (1986) e Amadeo *et al.* (1994). A diminuição do emprego formal decorrente da perda de dinamismo da atividade econômica foi em parte compensada com o fluxo de pessoas para o setor informal da economia, de modo que os efeitos sobre o mercado de trabalho apareceram menos na taxa de desocupação e mais na precarização das relações de trabalho.

Por outro lado, Neri *et al.* (2000) encontram evidências de que as mudanças da década seguinte, decorrentes da abertura comercial, absorção novas tecnologias e

alterações nos processos produtivos elevaram estruturalmente a taxa de desocupação.

5. Resultados

Com base na metodologia apresentada brevemente na seção 3, foram estimados dois modelos multivariados no módulo STAMP do programa OxMetrics para a taxa de desocupação da PNAD Contínua a partir da dinâmica da PNAD anual e da PED. A opção pela PED, alternativamente à PME, está associada ao maior número de informações disponíveis no período em que há coexistência com a PNAD Contínua (a PME foi descontinuada em fevereiro de 2016) e à manutenção metodológica da PED ao longo de toda a série histórica iniciada em 1984.

5.1 Retropolação iniciada em 1984

Na primeira estimação, a amostra de dados começa em 1984 e embute a PNAD anual apenas a partir de 1992, sendo parcimonioso no sentido de considerar os dados harmonizados pelo IBGE.

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 auto-regressivo de primeira ordem.

As três séries de desemprego foram consideradas no vetor de variáveis observadas, seguindo os modelos SUTSE. 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.

Dos resultados apresentados na tabela 3 é possível notar que os valores dos componentes do vetor de estado estimado no último instante de tempo (quarto trimestre de 2016) se mostraram estatisticamente significativos, no caso do AR (1) a 10%, sinalizando que a formulação do modelo com a presença desses componentes parece adequada.

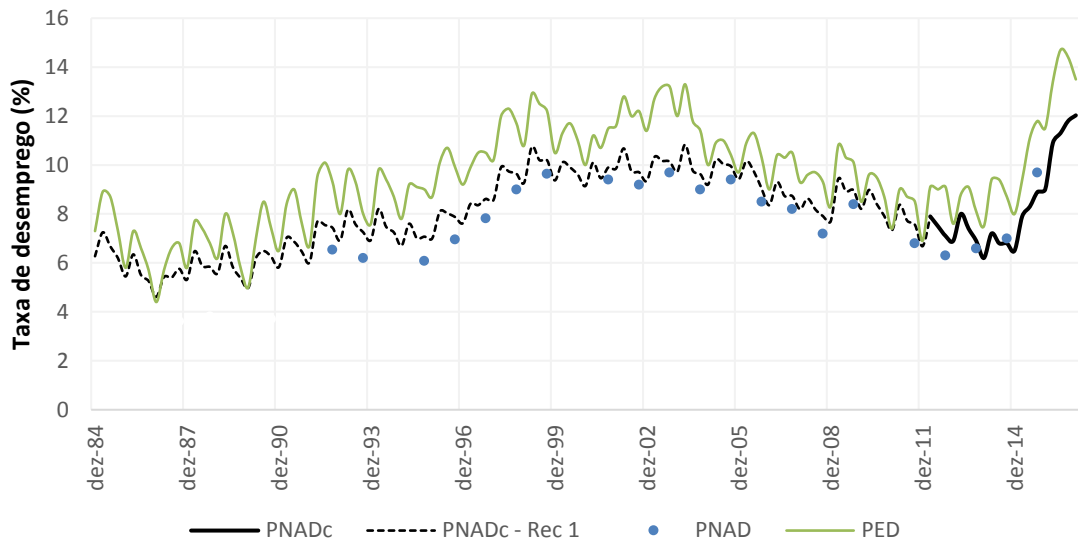
Tabela 3: Modelo multivariado PNAD Rec 1

Vetor Y: PNADc, PED, PNAD2			
Y = Nível + Sazonalidade + Irregular + AR(1)			
Seleção dos componentes: 0=out, 1=in, 2=dependent, 3=fix			
	PNADc	PED	PNAD2
Nível	2	1	1
Sazonalidade	2	1	0
AR(1)	2	1	1
Irregular	2	1	1
Variâncias dos choques na PED		Valor	razão - q
	Nível	0.1862	1.0000
	Sazonalidade	0.0004	0.0020
	AR(1)	0.1175	0.6312
	Irregular	0.0000	0.0000
Variâncias dos choques na PNAD2		Valor	razão - q
	Nível	0.0177	1.0000
	Sazonalidade	-	-
	AR(1)	0.0002	0.0105
	Irregular	0.0000	0.0000
Vetor de estado (4ºT/2016)		Valor	Prob
	Nível	12.4870	[0.00000]
	Sazonalidade 1	-0.3028	[0.00359]
	Sazonalidade 2	0.3143	[0.00231]
	Sazonalidade 3	-0.2288	[0.00016]
	AR(1)	0.0734	[0.08247]
	T	20	
	P	4	
	Rd ²	0.96626	

Fonte: elaboração própria.

Os dados faltantes obtidos pelo filtro de Kalman compreendidos no período do quatro trimestre de 1984 ao quarto trimestre de 2011 estão sintetizados na série PNADc – Rec 1, que pode ser visualizada no gráfico abaixo em conjunto com as variáveis relacionadas (PNAD anual e PED).

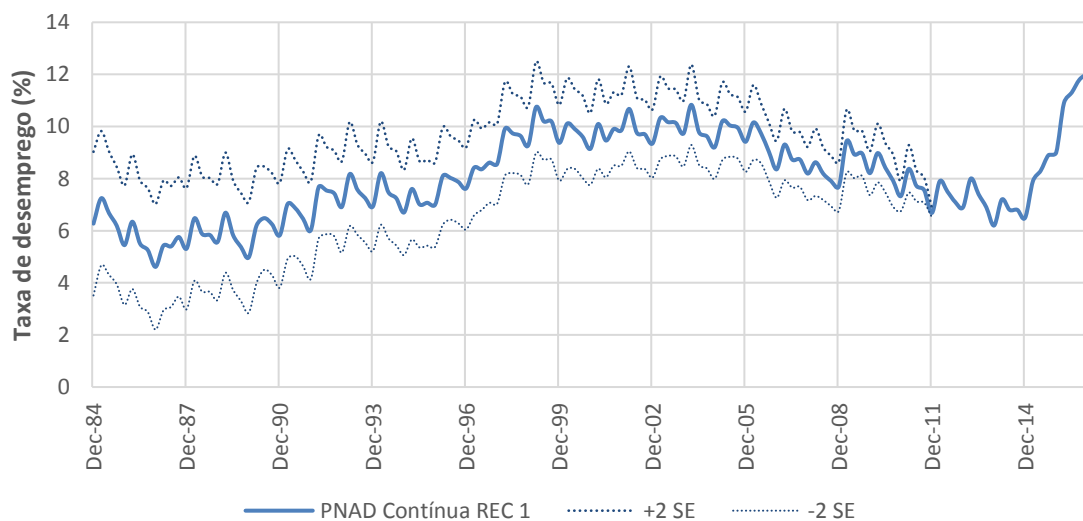
Figura 3. Retropolação iniciada em 1984



Fonte: Elaboração própria com dados do IBGE e Dieese.

Do exercício depreende-se, em uma perspectiva histórica, que o patamar da taxa de desemprego alcançado no quarto trimestre de 2016 (12,0%) é o mais elevado da série iniciada em 1984. Se adicionarmos os limites de dois desvios padrão da série reconstruída, considerando, portanto, a incerteza associada à estimação, o nível atual seria compatível em magnitude apenas com o limite superior do primeiro trimestre de 1999.

Figura 4. Intervalo de confiança da série reconstruída



Fonte: Elaboração própria.

5.2 Retropolação iniciada em 1976

Utilizando-se todos os dados da PNAD obtidos nas bases da OIT é possível estender a PNAD Contínua a partir de 1976. Apesar de o universo para a população em idade ativa estar ajustado (de 15 para 10 anos), o que, em alguma medida, contribui para diminuir as diferenças entre os pontos da PNAD antes e depois da revisão metodológica realizada pelo IBGE em 1992, restariam ainda outras questões relevantes, como a definição mais abrangente de trabalho que, idealmente, poderiam ser minimizadas através da manipulação criteriosa dos microdados.

Mesmo com as ressalvas, um modelo similar ao formulado anteriormente foi estimado para a amostra de dados iniciada em 1976. O vetor de estados no último instante de tempo da amostra encontra-se na tabela abaixo. A partir dele, com exceção do ciclo AR (1), os valores dos demais componentes são significativos e, portanto, diferentes de zero.

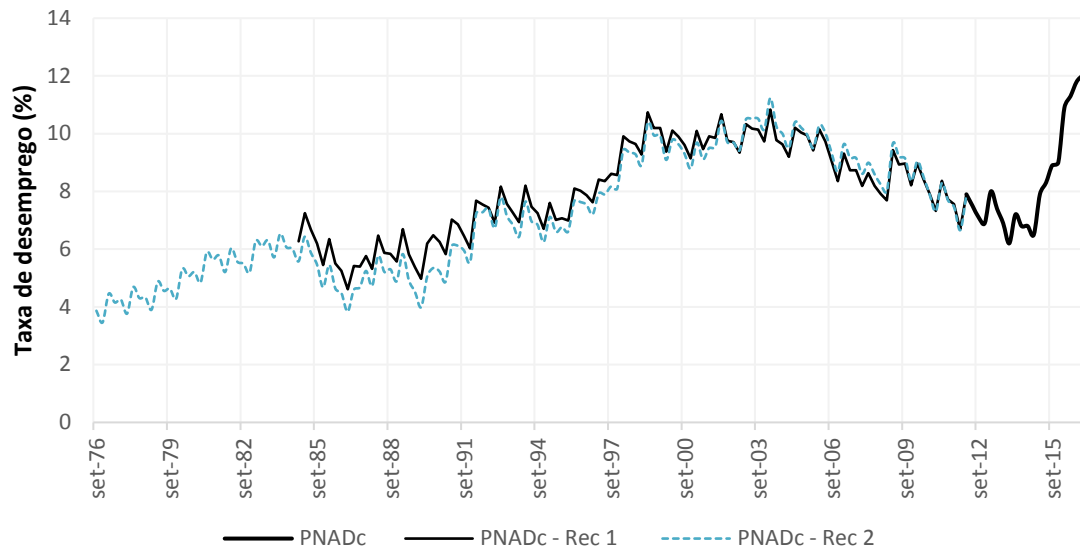
Tabela 4: Modelo multivariado PNAD Rec 2

Vetor Y : PNADc, PNAD, PED			
Y = Nível + Sazonalidade + Irregular + AR(1)			
Seleção dos componentes: 0=out, 1=in, 2=dependent, 3=fix			
	PNADc	PED	PNAD
Nível	2	1	1
Sazonalidade	2	1	0
AR(1)	2	1	1
Irregular	1	1	1
Variâncias dos choques na PED		Valor	razão - q
	Nível	0.0091	0.0511
	Sazonalidade	0.0003	0.0016
	AR(1)	0.1780	1.0000
	Irregular	0.0000	0.0000
Variâncias dos choques na PNAD		Valor	razão - q
	Nível	0.2442	1.0000
	Sazonalidade	-	-
	AR(1)	0.0000	0.0000
	Irregular	0.0000	0.0000
Vetor de estado (4 ^o T/2016)		Valor	Prob
	Nível	12.8356	[0.00000]
	Sazonalidade 1	-0.2894	[0.00178]
	Sazonalidade 2	0.3391	[0.00038]
	Sazonalidade 3	-0.2361	[0.00009]
	AR(1)	-0.2814	[0.67748]
	T	20	
	P	4	
	Rd ²	0.73935	

Fonte: elaboração própria.

Os dados faltantes recuperados com filtro de Kalman geraram uma nova série retropolada (PNADc – Rec 2) relativamente parecida, no tempo de coexistência, com a PNADc – Rec 1, embora com uma diferença de patamar mais persistente ao longo da década de 1980 e 1990. Na média, a PNADc Rec 2 é 0,3 ponto percentual inferior à PNADc Rec 1.

Figura 5. PNADc Rec 1 e PNADc Rec 2



Fonte: elaboração própria.

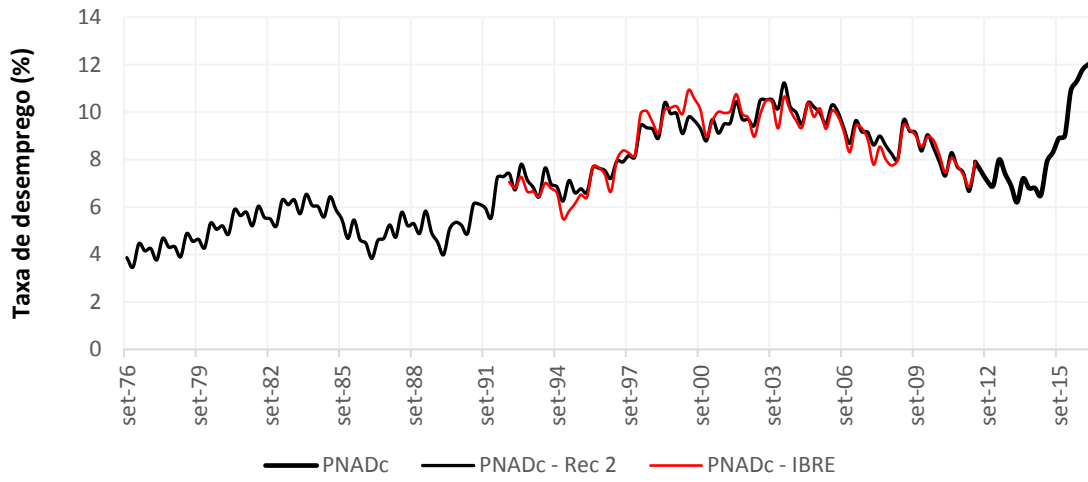
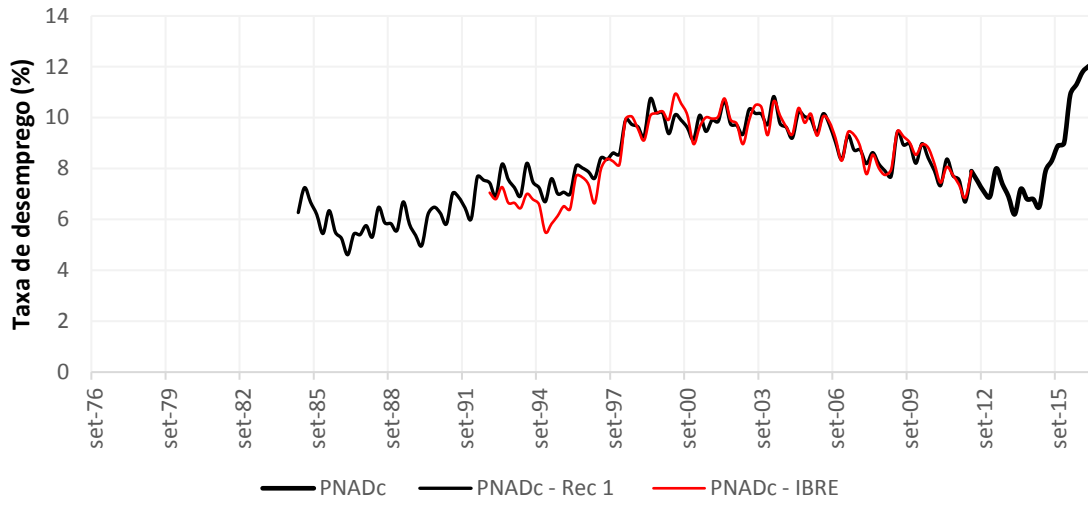
5.3 Comparações

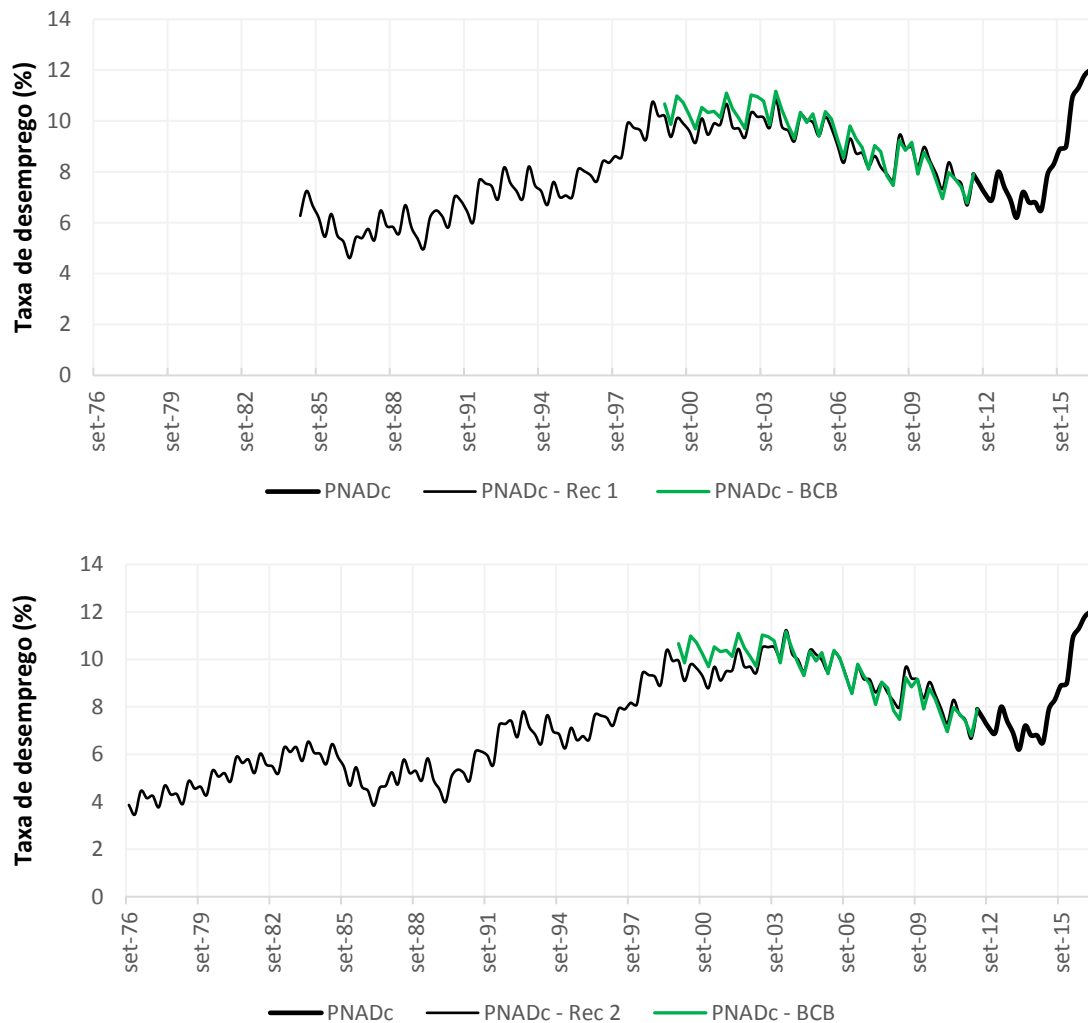
As séries PNADc Rec 1 e PNADc Rec 2 estimadas a partir da abordagem de modelos de componentes não observados apresentaram correlação elevada¹⁶ com as séries reconstruídas por Alves e Fasolo (2015) e por Vaz e Barreira (2016)¹⁷.

¹⁶ A correlação no período em comum da PNADc Rec 1 e da PNADc Rec 2 com as séries de Alves e Fasolo (2015) e Vaz e Barreira (2016) foram de 97% e 92% e de 96% e 95%, respectivamente.

¹⁷ 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).

Figura 6. Comparação com Alves e Fasolo (2015) e Vaz e Barreira (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)¹⁸.

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

¹⁸ Ver <http://portalibre.fgv.br/main.jsp?lumChannelId=4028808126B9BC4C0126BEA1755C6C93>. Acesso em 26/04/2017.

valores estimados na PNADc Rec 1, o pico anterior havia sido registrado em 1999 (10,1%) e 2003 (10,1%).

6. Possíveis extensões e limitações

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 segundo exercício (iniciado em 1976). Esse refinamento seria relevante no sentido de 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 se compatibilizar informações de diferentes períodos, tanto para as PNAD's quanto para os Censos Demográficos.

Além da possibilidade de trabalho sobre os microdados, outros passos podem ser elencados como sugestões de pesquisa e desdobramentos deste trabalho.

Um complemento para garantir robustez aos resultados apresentados diretamente para a taxa de desemprego poderia vir da comparação do resultado oriundo de reconstruções independentes das séries da população ocupada e desocupada¹⁹.

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. Como o agregador temporal da série é conhecido, seria possível aplicar a mesma metodologia na reconstrução da taxa de desemprego mensal.

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 também poderiam ser úteis na medida em que representam subsídios relevantes para o entendimento da dinâmica do consumo, crédito e arrecadação de certos tributos, por exemplo), o

¹⁹ O IBGE não disponibiliza as informações dos valores absolutos em termos harmonizados.

emprego por setores de atividades e por posição (com carteira assinada, sem carteira assinada, conta-própria etc.), além de suas aberturas por regiões do País.

7. 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 das séries históricas em alta frequência que só serão recuperados com o tempo.

Os dados faltantes, anteriores ao primeiro trimestre de 2012, foram reconstruídos a partir de modelos de componentes não observados multivariados (SUTSE). O exercício permite analisar a dinâmica histórica do mercado de trabalho de maneira mais apropriada e a série resultante pode ser testada em estimações como em uma Curva de Phillips, por exemplo.

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 Continua atingiu no final de 2016 o patamar mais elevado da série histórica retropolada desde 1976, superando em magnitude os resultados obtidos em outros períodos recessivos da história econômica do País.

8. Referências

ALVES, Sergio Afonso Lago e FASOLO, Angelo Marsiglia. *Not Just Another Mixed Frequency Paper*. Brasília: Banco Central do Brasil, 2015. (Working Series Paper, n. 400).

AMADEO, Edward et al. *A natureza e o funcionamento do mercado de trabalho brasileiro desde 1980*. Brasília: IPEA, 1994. (Texto para discussão, n. 353).

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 Censos (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, n. 2, p. 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, n. 32, Jul./Set. 2016. Nota técnica.

CHEN, Baoline. *An Empirical Comparison of Methods for Temporal disaggregation at the National Accounts*. Washington, DC: Bureau of Economic Analysis, 2007. (Working Paper, n. 4).

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, n. 4, p. 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.

CORSEUIL, Carlos H.; RAMOS, Lauro. Comparação da taxa de desocupação na PNAD Contínua e na PME. *Boletim Mercado de Trabalho – Conjuntura e Análise*, Brasília, DF, n. 58, p. 19-25, Abr. 2015.

DEDECCA, C. et al. O Censo Demográfico 2000 e a mensuração das formas de trabalho. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, 17., 2004, Caxambu. Anais... Caxambu: ABEP, 2004.

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, p. 99-102, Mar. 1971.

DI FONZO, Tommaso. *Temporal disaggregation of economic time series: towards a dynamic extension*. Luxemburgo: European Commission – Eurostat, 2003a. (Working Papers and Studies).

DI FONZO, Tommaso. Constrained reinterpolation of high-frequency data using related series: A simple dynamic model approach. *Statistical Methods & Applications*, v. 12, n. 1, p. 109-119, Fev. 2003b.

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.

FERNANDEZ, Roque. A methodological note on the estimation of time series. *The Review of Economics and Statistics*, v. 63, n. 3, p. 471-476, 1981.

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, n. 3, p. 303-309, 2000.

HARVEY, Andrew C.; PIERSE, Richard G. Estimating missing observations in economic time series. *Journal of the American Statistical Association*, v. 79, n. 385, p. 125-131, Mar.1984.

HARVEY, Andrew C.; KOOPMAN, Siem J.; PENZER, Jeremy. Messy time series: a unified approach. *Advances in Econometrics*, New York, v. 13, p. 103-143.

HARVEY, Andrew C.; VALLS PEREIRA, Pedro L. The Estimation of dynamic models with missing observation. *Brazilian Review of Econometrics*, v.5, n.2, p. 81-95, 1985.

HYPÓLITO, Elizabeth Belo. Pesquisa Nacional por Amostra de Domicílios: principais mudanças conceituais referentes ao tema trabalho. *Boletim Mercado de Trabalho – Conjuntura e Análise*, Brasília, DF, n. 43, p. 17-27, Mai. 2010.

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).

_____. *Pesquisa Nacional por Amostra de Domicílios*. Rio de Janeiro: IBGE, 2014. (Notas Metodológicas, v.1).

_____. *Sistema integrado de pesquisas domiciliares - SIPD*. Rio de Janeiro: IBGE, 2007. (Texto para Discussão, n. 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, p. 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, n. 2, p. 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, n. 1582).

MOAURO, Filippo. Monthly Employment Indicators of the Euro Area and Larger Member States: Real-Time Analysis of Indirect Estimates. *Journal of forecasting*, v. 33, n. 5, p. 339-349, 2014.

MOAURO, Filippo; SAVIO, Giovanni. Temporal disaggregation using multivariate structural time series models. *The Econometrics Journal*, v. 8, n. 2, p. 214-234, Jul. 2005.

NERI, Marcelo; CAMARGO, José M.; REIS, Maurício C. *Mercado de trabalho nos anos 90: fatos estilizados e interpretações*. Rio de Janeiro: IPEA, 2000. (Texto para discussão, n. 743).

NERI, Marcelo. O desemprego no censo: a febre e o termômetro. *Revista Conjuntura Econômica*, v. 56, n. 6, p. 54-55, Jun. 2002.

ORAIR, Rodrigo O.; SILVA, Wesley de J. *Investimentos dos governos subnacionais no Brasil: estimação e análise por modelos de espaço de estado*. Brasília: IPEA, 2012. (Texto para discussão, n. 1771).

PASTORE, José. O desemprego e a guerra de números. O Estado de São Paulo, 10 set. 2002.

PROIETTI, Tommaso. Temporal disaggregation by state space methods: Dynamic regression methods revisited. Luxemburgo: European Commission – Eurostat, 2005. (Working Papers and Studies).

SABÓIA, João Luiz Maurity. Transformações no mercado de trabalho no Brasil durante a crise: 1980-1983. *Revista de Economia Política*, São Paulo, SP, v. 6, n. 3, p. 82-106, jul./set. 1986.

SILVA, João S.; CARDOSO, Fátima. The Chow-Lin method using dynamic models. *Economic Modelling*, v. 18, n. 2, p. 269-280, Abr. 2001.

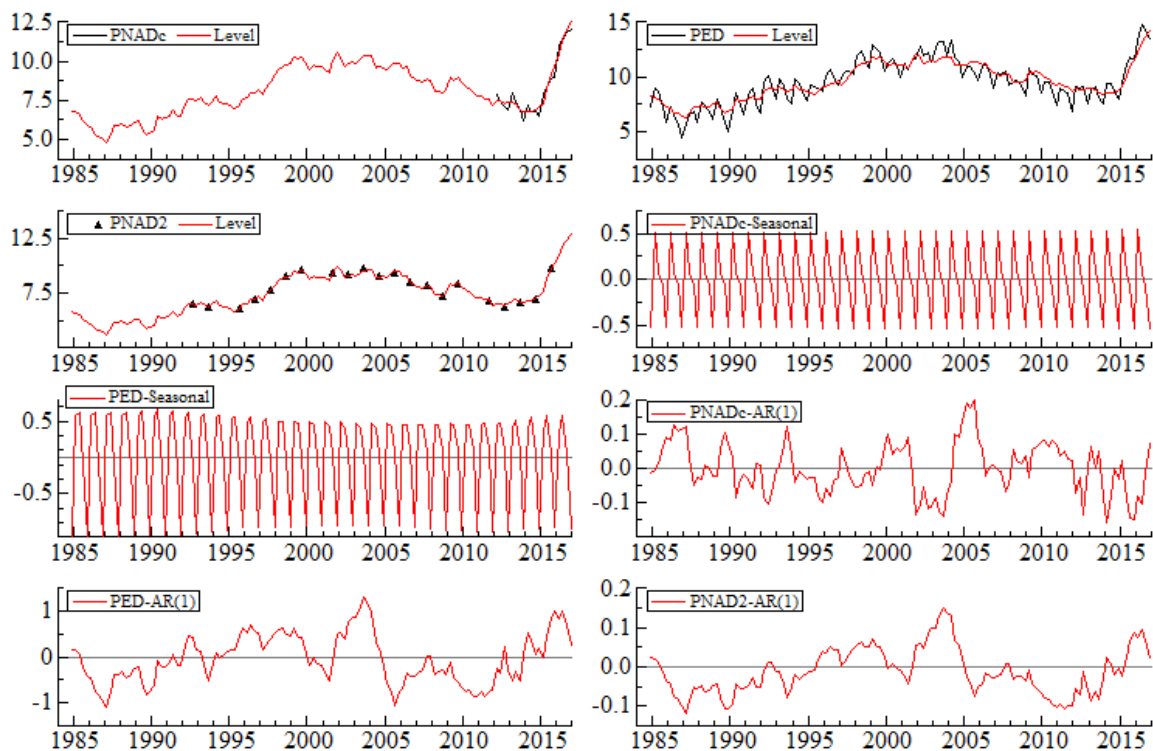
VARGAS, Francisco Eduardo Beckenkamp. *Formes et expériences de privation de travail au Brésil: La construction sociale du chômage dans la perspective d'une*

sociologie des rapports sociaux. 2008. 575 f. Tese (Doutorado em Sociologia) - Université de Versailles-Saint-Quentin-en-Yvelines, Versailles, 2008.

VAZ, Bruno Ottoni Eloy; BARREIRA, Thiago Cabral. *Metodologia de retropolação da Pesquisa Nacional por Amostra de Domicílios Contínua de 1992 a 2012*. Rio de Janeiro, RJ: FGV-IBRE, 2016. 13 p. (Nota técnica).

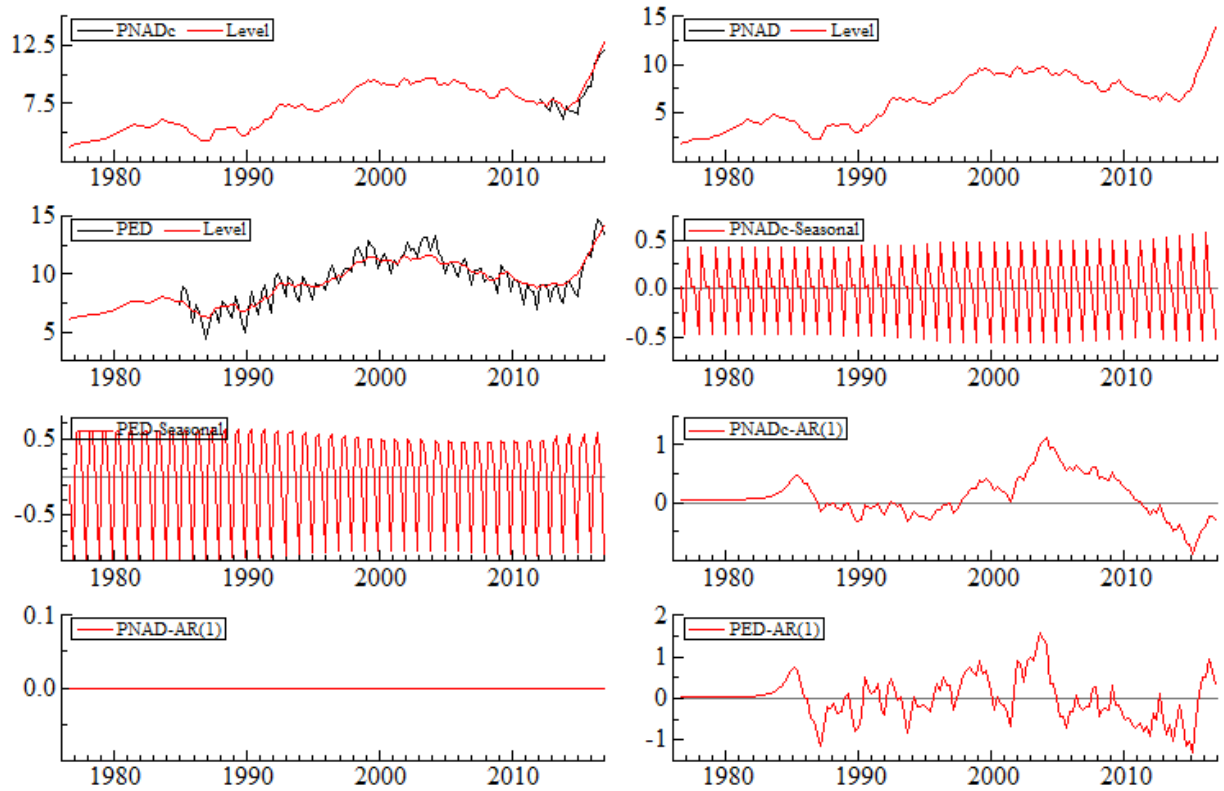
Anexos

1) Componentes estimados: PNADc - REC 1



Fonte: Elaboração própria.

2) Componentes estimados: PNADc - REC 2



Fonte: Elaboração própria.

3) Taxa de desemprego (%) – média anual

	PNADc	PNADc - Rec 1	PNADc - Rec 2	PNADc - IBRE	PNADc - BCB
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	-	6.39	5.61	-	-
1986	-	5.43	4.60	-	-
1987	-	5.47	4.81	-	-
1988	-	5.94	5.29	-	-
1989	-	5.72	4.82	-	-
1990	-	6.19	5.13	-	-
1991	-	6.59	5.94	-	-
1992	-	7.39	7.16	6.93	-
1993	-	7.48	7.06	6.76	-
1994	-	7.41	6.93	6.47	-
1995	-	7.17	6.77	6.22	-
1996	-	7.90	7.51	7.34	-
1997	-	8.49	8.02	8.19	-
1998	-	9.64	9.24	9.67	-
1999	-	10.13	9.84	10.11	-
2000	-	9.69	9.38	10.14	10.40
2001	-	9.83	9.46	9.91	10.34
2002	-	9.87	9.82	9.85	10.35
2003	-	10.09	10.42	10.03	10.65
2004	-	9.86	10.24	9.93	10.19
2005	-	9.90	10.02	9.91	9.99
2006	-	9.32	9.58	9.35	9.58
2007	-	8.74	9.14	8.85	9.04
2008	-	8.10	8.46	8.07	8.29
2009	-	8.88	9.09	9.07	8.79
2010	-	8.16	8.19	8.35	7.91
2011	-	7.58	7.52	7.50	7.47
2012	7.35	7.35	7.35	7.35	7.35
2013	7.13	7.13	7.13	7.13	7.13
2014	6.83	6.83	6.83	6.83	6.83
2015	8.53	8.53	8.53	8.53	8.53
2016	11.51	11.51	11.51	11.51	11.51

Fonte: IBGE, Vaz e Barreira (2016) e Alves e Fasolo (2015). Elaboração própria.