

Uma proposta de construção da taxa de desocupação em Minas Gerais nas primeiras décadas do século XXI a partir de pesquisas domiciliares

Resumo

O objetivo deste trabalho é apresentar uma proposta metodológica para a construção de série histórica de 20 anos para a taxa de desocupação de Minas Gerais. Propõe-se retropolar as séries dos ocupados e desocupados no estado até 2002, utilizando a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), a Pesquisa Nacional por Amostra de Domicílios (PNAD), os dados da Pesquisa Mensal de Emprego (PME) e da Pesquisa de Emprego e Desemprego (PED) e das contas estaduais. Os resultados mostram a reconstrução da série histórica, com taxas de desocupação próximas entre 2002 e 2011, com diferença média de apenas 1%.

Palavras Chave: Taxa de Desocupação; Retropolação; Metodologia Estatística; Minas Gerais.

Classificação JEL: C88, J01, J64.

Abstract

This work presents a methodological proposal for the construction of a 20-year historical series for the unemployment rate in Minas Gerais. A retropolation of employed and unemployed in the state up to 2002 is proposed, using Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), the Pesquisa Nacional por Amostra de Domicílios (PNAD), the data from the Pesquisa Mensal de Emprego (PME) and from Pesquisa de Emprego e Desemprego (PED) and state accounts. The results reveal that the procedure allowed the reconstruction of the historical series, with similar unemployment rates between 2002 and 2011, a average difference of only 1%.

Key Words: Unemployment Rate; Retropolation; Statistical Methodology; Minas Gerais.

JEL Classification: C88, J01, J64.

1. Introdução

A compreensão do mercado de trabalho brasileiro e do mineiro, em particular, passa, necessariamente, por uma análise de médio e longo prazo de sua história, economia, política, sociedade e inserção no contexto global. Tal análise deve contar com séries históricas abrangentes, confiáveis e longas, capazes de retratar as nuances do mundo laboral em diferentes pontos no tempo. Em Minas Gerais, o acompanhamento de médio e longo prazo do mercado de trabalho tem sido uma tarefa desafiadora em virtude da ausência de bases de dados com horizonte temporal extenso.

As características estruturais e conjunturais do mercado de trabalho mineiro têm sido analisadas à luz das pesquisas amostrais e dos registros administrativos tradicionais existentes no país e no estado. No primeiro caso, destacam-se a Pesquisa Nacional por Amostra de Domicílios (PNAD), a Pesquisa Mensal de Emprego (PME) e a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), todas do Instituto Brasileiro de Geografia e Estatística (IBGE), a Pesquisa de Emprego e Desemprego (PED) do Departamento Intersindical de Estatística e Estudos Econômicos (DIEESE) e da Fundação Sistema Estadual de Análise de Dados (SEADE) e a Pesquisa por Amostra de Domicílios (PAD) de Minas Gerais da Fundação João Pinheiro. Já no segundo, tem-se o Cadastro Geral de Empregados e Desempregados (CAGED) e a Relação Anual de Informações Sociais (RAIS) ambos do Ministério da Economia e Previdência e o Cadastro Único, do Ministério da Cidadania.

À primeira vista, Minas Gerais parece estar munida de pesquisas suficientes para subsidiar suas políticas de emprego, fomentar estudos e análises baseadas em um diagnóstico situacional de longo prazo, contudo, um olhar mais cuidadoso para as bases de dados revela as limitações de cada uma delas, principalmente em relação à longevidade, o que impede uma análise histórica até os dias atuais do que tem sido e como tem-se comportado o mercado de trabalho mineiro.

Nesse sentido, o objetivo deste trabalho é apresentar uma proposta metodológica de retopolação da taxa trimestral de desocupação de Minas Gerais, calculada por meio dos dados da PNAD Contínua, para o período de 2002 a 2011, e calcular uma taxa de desocupação trimestral para 20 anos: 2002 a 2022. Para isso, serão realizadas quatro etapas: 1) utilização dos métodos de desagregação de séries temporais de Denton-Cholette e Chow-Lin para aumentar a frequência das séries históricas de ocupados e desocupados da PNAD no período de 2002 a 2015, tendo como referência a PME e a PED; 2) estimação de regressão univariada utilizando-se os dados de ocupados e desocupados da PNAD Contínua e PNAD para o intervalo de 2012 a 2015, quando as pesquisas coincidem; 3) retopolação dos ocupados e desocupados da PNAD Contínua por meio do coeficiente de regressão; e 4) cálculo da taxa de desocupação de Minas Gerais para o período de 2002 a 2022.

O artigo que ora se apresenta é um avanço da proposta apresentada por Marques *et al* (2022), em que foi realizada uma compatibilização conceitual dos ocupados e desocupados da PNAD Contínua na base de dados da PNAD e construída uma série histórica mais longa para a taxa de desocupação (2001-2021) para Minas Gerais. Ademais, contribui para a discussão metodológica sobre retopolação de indicadores, conforme IBGE (2002), Vaz e Barreira (2016), Guzmán e Villavicencio

(2019), Bacciotti e Marçal (2020), e sobre a importante tarefa de se construir séries temporais longas de indicadores socioeconômicos.

Na próxima seção, são descritas as principais pesquisas que trazem informações do mercado de trabalho de Minas Gerais, bem como suas temporalidades e abrangências, e uma revisão da literatura sobre a construção de séries temporais mais longas, com enfoque na retropolação. Em seguida, os ajustes necessários nas séries da PED e PME, utilizadas como referências para aumentar a frequência da PNAD e a metodologia de retropolação. Na quarta seção apresentam-se os resultados e finalmente na quinta, as principais conclusões do trabalho.

2. A retropolação como alternativa à descontinuidade de séries históricas

Um problema comumente enfrentado pelos estudiosos do mercado de trabalho brasileiro e mineiro, em particular, diz respeito à ausência de uma série histórica de longo prazo que forneça informações sobre o comportamento das variáveis laborais clássicas desde um passado longínquo. A temática conta com informações oriundas de pesquisas amostrais (PNAD e PNAD Contínua, PME e PED) e registros administrativos (RAIS, CAGED e CADÚNICO) cada qual com suas possibilidades e limitações que, no cômputo geral, se restringem a determinados grupos, localidades e não conseguem ligar o passado e o presente sob uma perspectiva continuada. Pelo contrário. Pelo fato de as bases possuírem diferentes metodologias, abrangências e temporalidades, a manipulação simples de seus dados permite análises compartimentadas, elaboração de hipóteses e apontamentos que, apesar de importantes, poderiam ser suprimidos pela existência de uma fonte de dados única e abrangente.

A Pesquisa Nacional por Amostra de Domicílios (PNAD), do IBGE, teve início em 1967, tendo seus resultados apresentados trimestralmente até 1970. A partir de 1971 os resultados passaram a ser anuais. A PNAD passou por diversas alterações ao longo do tempo, tanto do plano amostral, de sua abrangência e da forma de mensuração dos conceitos pesquisados. Mas a pesquisa básica visava investigar de forma contínua vários temas a fim de mensurar o nível e acompanhar as características socioeconômicas da população brasileira. Possuía abrangência nacional, com representatividade para Brasil, grandes regiões, estados e nove regiões metropolitanas (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba e Porto Alegre), e os dados eram coletados em uma única tomada no ano, geralmente setembro. Os principais indicadores do mercado de trabalho seguiam as recomendações internacionais, em que a força de trabalho era composta pelos ocupados na semana de referência da pesquisa e os desempregados classificados a partir da não realização de trabalho associada à procura em 30 dias anteriores à data de referência. A pesquisa foi descontinuada em 2015.

A Pesquisa Mensal de Emprego (PME), como o próprio nome sugere, era uma pesquisa mensal realizada pelo IBGE entre 1980 e 2016, com cobertura da área urbana de seis regiões metropolitanas (São Paulo, Rio de Janeiro, Belo Horizonte, Recife, Salvador e Porto Alegre). Nesse período, passou por algumas revisões: uma completa em 1982, duas parciais em 1988 e 1993, com ajustes do plano de amostragem, e por uma revisão metodológica em 2001. Nesta última, os indicadores

principais da força de trabalho ficaram mais próximos dos da PNAD. A mensuração do desemprego seguia a recomendação internacional de procura efetiva em 30 dias anteriores à entrevista, disponibilidade para trabalhar e a não realização de trabalho na semana de referência.

A Pesquisa de Emprego e Desemprego (PED) era uma pesquisa mensal contínua criada por meio de convênio entre a Fundação Seade, Departamento Intersindical e Estudos Socioeconômicos (Dieese) - desenvolvedores da metodologia, e realizada no Distrito Federal e nas regiões metropolitanas de São Paulo, Porto Alegre, Belo Horizonte, Salvador, Recife e Fortaleza. Na Região Metropolitana de Belo Horizonte foi realizada pela Fundação João Pinheiro e Secretaria de Estado do Trabalho e Desenvolvimento Social (Sedese) entre 1995 e 2014. Seu objetivo era investigar a estrutura e dinâmica do mercado de trabalho regional por meio de um levantamento sistemático sobre emprego, desemprego e renda. Tinha como propósito também a construção de indicadores compatíveis com as estatísticas internacionais, mas adotava como premissa as limitações das pesquisas oficiais na classificação dos indivíduos na força de trabalho, mais apropriadas para descrever mercados de trabalho homogêneos e mais estruturados, diferentes do brasileiro. Neste sentido, a PED utilizava procedimentos operacionais que mensuravam o desemprego aberto em 30 dias anteriores à pesquisa e outras manifestações do desemprego. Identificava as pessoas que estavam desalentadas, pois após procurarem trabalho por um longo período deixavam de fazê-lo por desestímulos do mercado, mesmo reconhecendo sua necessidade e disponibilidade de tempo para trabalhar. E considerava também como desempregadas aquelas pessoas que realizavam alguma atividade remunerada com ganhos avulsos, descontínuos e instáveis associados à procura de trabalho. Essa condição de inserção no mercado seria muito representativa das características do mercado de trabalho brasileiro, marcado por elevada heterogeneidade e concentração de renda e de um Sistema Público de Emprego que não amparava os trabalhadores em períodos longos de não trabalho. Assim, a PED identificava três situações de desemprego: desemprego aberto em 30 dias; desemprego oculto por desalento e desemprego oculto por trabalho precário que constituam o conjunto dos desempregados (MONTAGNER e HAGA, 2003).

Embora as três pesquisas se aproximassem da mensuração do desemprego aberto (30 dias) a partir das recomendações internacionais, a forma de acessar o fenômeno apresentava algumas distinções tanto no que tange a seus planos amostrais quanto aos instrumentos de coleta, gerando medidas com níveis distintos. Barros *et al* (1996) mostram como o fluxo do questionário da PED e da PME impactava no nível das taxas de desocupação, sendo a taxa estimada pela primeira mais elevada que a da segunda pesquisa. Segundo esses autores, a PME segue as recomendações internacionais quanto à prioridade da sequência das perguntas: emprego, desemprego e inatividade. Já a PED encontra-se em desacordo com tais recomendações, perguntando primeiro se o entrevistado “procurou trabalho”, para então questionar “se ele tinha trabalho”. Corseuil e Ramos (2014) corroboram os apontamentos de Barros *et al* (1996) ao investigarem os diferenciais de nível entre as taxas de desocupação da PNAD Contínua, PNAD e PME (ordem decrescente de valores). De acordo com os autores, as hipóteses de diferenças de cobertura e de amostra das pesquisas não são suficientes para explicar os diferenciais de nível entre

as taxas de desocupação, levando-os a considerar, por eliminação, as hipóteses de discrepâncias na ordem das perguntas, no fluxo do questionário, nas palavras utilizadas nas perguntas e a orientação dada aos entrevistadores como fatores responsáveis por tais diferenças.

A Relação Anual de Informações Sociais (Rais), o Cadastro Geral de Empregados e Desempregados (Caged) e o Cadastro Único (CadÚnico) fornecem medidas de outra ordem uma vez que são registros administrativos. A Rais é um instrumento de coleta de dados instituído pelo Decreto nº 76.900, de 23/12/1975, com o objetivo de fornecer informações para o controle da atividade trabalhista no Brasil, dados para a elaboração de estatísticas do trabalho para a população em geral e entidades governamentais. Declaram dados para a Rais todos os empregadores com ou sem funcionários, públicos e privados, empresas individuais, condomínios e sociedades civis. Relacionam-se os dados de todos os empregados contratados por pessoa jurídica ou física sob regime da Consolidação das Leis Trabalhistas (CLT), por prazo indeterminado ou determinado, assim como servidores da administração pública (direta, indireta ou fundações), trabalhadores avulsos, temporários, aprendizes, servidores públicos não efetivos, servidores e trabalhadores licenciados, e trabalhadores regidos pelo Estatuto do Trabalhador Rural. Os dados da Rais são disponibilizados anualmente pelo Ministério responsável pela pasta do Trabalho e Emprego, atualmente, o Ministério do Trabalho e Previdência.

O Caged, criado pela Lei 4.923 de 23/12/1965, visa acompanhar e fiscalizar o processo de admissão e dispensa dos trabalhadores regidos pela CLT. Desde 1986 é usado como suporte ao pagamento do seguro-desemprego e a partir de 1983, o Caged apresenta o índice mensal de emprego, a taxa de rotatividade e a flutuação da força de trabalho (admitidos/desligados). As desagregações possíveis são idênticas às da Rais em termos geográficos, setoriais e ocupacionais. No entanto, há diferenças no universo de cada registro quanto na forma e periodicidade de alimentação de cada um deles.

O Cadastro Único para Programas Sociais (CadÚnico) é um registro administrativo do governo federal que contém informações sobre as famílias brasileiras em condições de pobreza e extrema pobreza. Devem ser cadastradas as famílias que recebem até meio salário-mínimo per capita ou até três salários-mínimos de renda mensal total. O CadÚnico tornou-se importante instrumento base para seleção de famílias de diversos programas sociais. O Decreto nº 6.135 de 2007, que dispõe sobre o Cadastro Único, possibilita o registro de famílias com renda superior às de famílias pobres ou extremamente pobres, desde que sua inclusão esteja associada à seleção ou acompanhamento de algum programa social. No CadÚnico há informações sobre a situação no mercado de trabalho de cada membro da família, no entanto, as informações se referem à categorização básica do trabalho (ocupação e renda) ou a inexistência dele.

Assim, as pesquisas domiciliares amostrais (PNAD, PED e PME) tem uma longa série histórica, passaram por algumas modificações ao longo do tempo, mas mantiveram a comparabilidade relativa em suas trajetórias. Como mediam aspectos distintos da realidade, pode-se dizer, que de certa forma, as pesquisas se

complementavam, a despeito das divergências entre elas. Enquanto a PNAD retratava condições mais estruturais do mercado de trabalho, a PME apresentava as evoluções sazonais além das estruturais. Já a PED permitia acompanhar a sazonalidade e outras manifestações do desemprego nas regiões metropolitanas cobertas por ela. Com a interrupção da PED e substituição da PNAD e PME pela PNAD Contínua, acompanhar a trajetória do mercado de trabalho no país tornou-se tarefa mais desafiadora, a despeito dos grandes avanços propiciados por ela. Um desses desafios é a falta de uma série histórica que só será construída com o passar dos anos.

A fim de contornar o problema da comparação intertemporal de longo prazo, diversos autores têm lançado mão de metodologias de interpolação e retropolação para criar séries temporais mais longas de indicadores específicos. Friedman (1962) destaca que a interpolação é um método específico para estimar valores intermediários e difere-se, inclusive, de técnicas que se propõem aumentar a frequência de séries temporais de baixa frequência. Essas últimas, tratadas por Chow e Lin (1971) e Denton (1971) consistem em utilizar a sazonalidade de séries de referência para ajustar valores mensais e quadrimestrais, por exemplo, de séries anuais, prática bastante utilizada por economistas. Chow e Lin (1971) propõem um método de regressão que relaciona as variáveis de baixa e alta frequência, enquanto Denton, (1971), preocupado explicitamente com a preservação da sazonalidade da série a ser desagregada, apresenta uma abordagem geral para lidar com a não uniformidade do movimento da série de referência a fim de evitar quebras e descontinuidades entre o último período de um ano e o primeiro do ano seguinte da nova série.

A retropolação, por sua vez, é uma técnica comumente utilizada pelas contas nacionais para lidar com a ruptura de séries temporais em virtude de mudanças conceituais e/ou metodológicas. Ela caracteriza-se pelo cálculo de uma nova série para anos pretéritos considerando tais mudanças. Essa série do passado é encadeada com a nova fornecendo uma série de longo prazo para o indicador desejado (BACCIOTTI, 2017).

Um dos pioneiros na utilização de metodologias de retropolação no Brasil foi IBGE (2002) que, para o Sistema de Contas Nacionais, retropolou o Produto Interno Bruto (PIB) e seus componentes para o período de 1995 a 2001 com o propósito de manter uma série integrada das Contas Regionais. Em 2002, o IBGE realizou aperfeiçoamentos metodológicos no cálculo do PIB em relação à série anterior, referenciada no ano de 1985, e com isso interrompeu uma série histórica de 20 anos (1985 a 2006). A fim de minimizar as perdas provenientes da descontinuidade da série, IBGE (2002) retropolou o PIB e seus componentes até 1995, ano escolhido devido à homogeneidade da moeda brasileira. No processo, foram compatibilizados e reponderados os setores de atividade econômica de ambas as metodologias e a nova série passou a contar com um horizonte temporal de 10 anos (1995 a 2006).

Feu e Alvim (2007) comparam os resultados obtidos em IBGE (2002) com os do cálculo direto (ano de referência 1985) e acham uma pequena diferença de 3,2 pontos percentuais entre o PIB a preços constantes, entre o início e o final do período, e de 0,8 pontos percentuais nos investimentos (Formação Bruta de Capital Fixo – FBKF). No entanto, destacam as mudanças na composição do FBKF e de sua na participação relativa no total do PIB, esta última de 19,3% na série antiga para 16,6%

na nova. Adicionalmente, os autores realizam uma retropolação dos componentes do PIB pós 1995 com o objetivo de comparar seus resultados com os do IBGE (2002). Como a variação real do PIB, da FBKF e do deflator implícito nos períodos coincidentes (1995-2006) da série retropolada por IBGE (2002) e do cálculo direto foi muito parecida, e as diferenças deveram-se às alterações no nível dos preços relativos entre os bens e serviços, para a retropolação Feu e Alvim (2007) assumiram como pressuposto a manutenção das taxas de crescimento do preço e da quantidade dos componentes do PIB e do valor agregado para o intervalo de 1995 a 2006.

Assim como no Brasil, no México a série histórica do PIB nacional também foi interrompida em virtude de mudanças metodológicas e do ano-base dos cálculos. Com vistas a minimizar o a quebra da série, Guzmán e Villavicencio (2018) e INEGI (2019) propõem a retropolação trimestral do PIB mexicano até 1980, por meio da abordagem dos vetores autorregressivos univariados com variáveis auxiliares mais extensas temporalmente. A ideia é que as variáveis auxiliares/relacionadas sirvam de guia para as retropolações, fornecendo uma tendência temporal para o PIB e, no caso específico do México, restrições espaciais e temporais.

Corona e López-Pérez (2020) propõem a retropolação do Indicador Trimestral da Atividade Econômica Estadual (ITAE)¹ mexicano por grandes setores de atividade como alternativa ao método de Guzmán e Villavicencio (2018) e incorporaram, adicionalmente, restrições temporais para cada estado federativo, por meio da participação do PIB nominal para os anos de 1980 e 1985. Os autores mostram como a utilização de valores nominais para restringir as retropolações distorcem os resultados estaduais ao longo do tempo, uma vez que refletem a estrutura econômica do ano corrente e não do ano-base, e sugerem a utilização de sua proposta à de Guzmán e Villavicencio (2018) em virtude da utilização de uma quantidade maior de informações e, por conseguinte, não necessitar de aumentar a frequência de séries, conforme estes autores.

Na temática do mercado de trabalho, vários autores têm apresentado propostas de retropolação dos dados da PNAD Contínua a fim de criar uma série histórica dos principais indicadores, haja vista a contemporaneidade da pesquisa. Vaz e Barreira (2016), por exemplo, apresentam uma proposta de retropolação da população economicamente ativa, da população em idade ativa, ocupada, da taxa de desocupação e da renda do Brasil, entre 1992 e 2012, utilizando-se a base da PNAD Contínua. Os autores realizam uma compatibilização conceitual entre a PNAD e a PNAD contínua e restringem os níveis das variáveis geradas na *PNAD Ajustada* (conceitos da PNAD Contínua na base de dados da PNAD) no mês de setembro, mês de realização da pesquisa. De posse das séries com as restrições em setembro, calculam uma variação mensal dos indicadores pela PME e a aplicam na PNAD Ajustada, gerando, desta forma, indicadores mensais para a PNAD Contínua.

Bacciotti e Marçal (2020) também retropolam a taxa de desocupação brasileira calculada por meio da PNAD Contínua para um horizonte de 40 anos (1976-2016). Os autores utilizam um modelo estrutural de componentes não observados, utilizando as taxas de desocupação mensal da PED e anual da PNAD como bases para a estimação de nível e sazonalidade da nova série, denominadas por eles de *PNAD Reconstruída*. Os resultados são comparados com os de um modelo similar estimado

¹ Indicador estadual de conjuntura da situação e evolução macroeconômica do México com maior desagregação temporal.

com dados da PNAD a partir de 1984, e mostram-se muito próximos no período de 2012-2015, quando as pesquisas coexistem.

Carvalho (2016) utiliza os dados da PNAD e da PME para criar séries longas, de alta frequência e compatíveis com a PNAD Contínua, para a taxa de desocupação e de atividade do Brasil, no intervalo de setembro de 1992 a setembro de 2014. O autor utiliza o método de Denton para aumentar a frequência da PNAD por meio dos movimentos sazonais mensais da PME. Os resultados mostram um encadeamento adequado da taxa de desocupação da nova série da PNAD com a da PNAD Contínua, mas uma pequena discrepância na taxa de atividade explicada pela mudança na classificação da população não economicamente ativa na PME após dezembro de 1994, gerando um pico na taxa de atividade quando ela se encontrava em queda na PNAD.

Marques *et al* (2022) apresentam uma proposta de compatibilização conceitual dos ocupados e desocupados da PNAD e da PNAD Contínua para calcular a taxa de desocupação de Minas Gerais para o período de 2001-2021. Os autores mostram a evolução do conceito de desocupação no Brasil desde a década de 1980 e as formas de sua mensuração pelas principais pesquisas. Comparam os conceitos de ocupados e desocupados adotados pela PNAD e PNAD Contínua, destacando suas semelhanças e divergências e, a partir daí, reconstroem os dois indicadores para Minas Gerais na base da PNAD, conforme as definições adotadas pela PNAD Contínua, e o denominam de *PNAD Ajustada*. De acordo com Marques *et al* (2022), a maior diferença média (entre 2012 e 2015) entre a PNAD Contínua e a *PNAD Ajustada* ocorreu na categoria dos ocupados, com a primeira registrando 204 mil pessoas a menos que a segunda, em virtude, principalmente, das pessoas que desempenhavam atividades para o próprio consumo que passaram a não ser consideradas ocupadas na PNAD Contínua. Quanto aos desocupados, a diferença entre as pesquisas foi somente de 33 mil pessoas, uma vez que o conceito de desocupação adotado pela PNAD Contínua é um pouco mais abrangente do que o considerado na PNAD. A taxa de desocupação da *PNAD Ajustada* manteve praticamente o padrão da taxa de desocupação fornecida pela PNAD, com um nível mais elevado devido aos ocupados que impactaram no denominador das taxas e, por conseguinte, no resultado.

3. Metodologia

A retropolação dos ocupados e desocupados da PNAD Contínua para, em seguida, calcular a taxa de desocupação de Minas Gerais para o período de 2002 a 2011 foi realizada por meio de regressão linear univariada entre suas séries e as da PNAD, para o intervalo coincidente das pesquisas: 2012 a 2015. Contudo, antes da retropolação em si, foram necessários alguns ajustes nas séries de referência e na própria PNAD, conforme se segue:

- 1) Aumento da frequência dos ocupados e desocupados da PNAD anual de 2002 a 2015 para trimestral, conforme frequência da PNAD Contínua, por meio de regressão. Para isso foram tomadas como referência as séries temporais trimestrais de ocupados e desocupados da PME, PED e do PIB de Minas Gerais, este último calculado e publicado pela Fundação João Pinheiro. A ideia foi obter séries trimestrais da PNAD com referência em cada uma dessas pesquisas, de modo que a desagregação captasse o movimento (isto é, a

estrutura da curva) dessas outras séries que mensuram trabalho e produção, para então comparar os resultados e selecionar, dentre eles, o que alcançasse o melhor ajuste em relação aos dados originais anuais da PNAD;

2) Complementação das séries de ocupados e desocupados da PED para o terceiro e quarto trimestres de 2014 e da PME para o primeiro trimestre de 2002. Como para desagregar a PNAD anual para trimestral era preciso que houvesse multiplicidade exata entre ela e a frequência das séries de referência ou “indicadoras” (PED, PME e PIB), foi necessário estimar alguns registros para PME e PED. Para isso foram feitas regressões lineares entre o PIB de Minas Gerais (variável independente) e os ocupados e desocupados dessas pesquisas com vistas estimar os valores faltantes;

3) Complementação dos ocupados e desocupados da PNAD para o ano de 2010. Além da PME e da PED, a própria PNAD anual também precisou ser complementada, já que não existe dado para 2010, ano de realização do Censo Demográfico. Também nesse caso o método adotado foi uma regressão linear entre os valores do PIB entre 2002 e 2015 (excluído o ano de 2010) e da PNAD;

Para cada série acima mencionada, foram compostos vetores numéricos no *software* RStudio, de modo a realizar as regressões e desagregações utilizando pacotes desenvolvidos em linguagem R. Para as regressões e operações matemáticas simples foram utilizadas as funções estatísticas dos pacotes básicos do próprio RStudio. Já para as desagregações foi usado o pacote “tempdisagg”², que operacionaliza um conjunto de métodos já consolidados nesse âmbito, cada um promovendo a desagregação das séries temporais segundo técnicas estatísticas distintas. Neste trabalho, testou-se os métodos de Denton-Cholette e de Chow-Lin, optando-se pelo segundo, mas utilizando também o primeiro para a regressão entre PNAD e PNAD Contínua na série dos desocupados. Conforme Sax e Steiner (2013), criadores do pacote, o primeiro tem como principal característica a preservação do movimento da série temporal cuja frequência será aumentada, gerando curva similar à da série indicadora, ainda que as séries não tenham correlação entre si. Já o método de Chow-Lin opera uma regressão pelo método dos mínimos quadrados generalizados sob o pressuposto de que a relação linear entre as séries de baixa e de alta frequência também deveria se manter caso ambas tivessem alta frequência – além disso, diferentemente do método de Denton-Cholette, o de Chow-Lin admite o uso de mais de uma série de referência para a desagregação (SAX e STEINER, 2013).

Com base nestes procedimentos, foram geradas três séries desagregadas para ocupados e outras três para desocupados a partir dos dados da PNAD entre 2002 e 2015, combinando cada uma das três séries de referência (PME, PED e PIB), pelo método de Chow-Lin. Também foram testadas desagregações combinando as três ou duas das três séries de referência, mas em nenhuma dessas combinações os coeficientes estimados pelo método foram, todos eles, estatisticamente significativos para intervalos de confiança aceitáveis. Quando, mesmo se limitando a uma série de referência, os coeficientes estimados não se mostraram estatisticamente significativos para os desocupados. Posto isto, mostrou-se necessária a transformação dos valores das séries em seus logaritmos naturais, obtendo-se, assim, resultados mais

² As funções do pacote são descritas em <https://cran.r-project.org/web/packages/tempdisagg/tempdisagg.pdf>. Acesso em 13 jun. 2022.

consistentes e, em seguida revertidos os valores para o nível original por meio da operação inversa. Também foram feitas desagregações das séries da PNAD pelo método de Denton-Cholette com e sem referência na PED, PME e no PIB.

As séries da PNAD desagregadas em trimestres serviram, então, de variável independente na regressão linear contra as séries trimestrais originais de ocupados e de desocupados da PNAD Contínua, para o período entre 2012 e 2015. Utilizando as séries desagregadas pelo método de Chow-Lin, foram feitas seis regressões que tiveram como ajuste a definição do intercepto igual a zero, de modo que a linha de regressão passasse necessariamente pela origem do plano cartesiano. Noutras palavras, partiu-se do pressuposto de que ambas as pesquisas – PNAD e PNAD Contínua – teriam a mesma capacidade de registrar zero trabalhadores ocupados quando não houvesse de fato nenhum trabalhador ocupado (da mesma forma que, com o PIB igual a zero, não haveria produto e, conseqüentemente, nenhum trabalhador ocupado). No caso da série de desocupados, a regressão também foi feita com a PNAD desagregada sem série de referência.

Por fim, os coeficientes obtidos nessas regressões foram usados para gerar valores preditos para a PNAD Contínua entre 2002 e 2011, obtendo séries retropoladas. Essas séries foram comparadas para identificar, dentre elas, aquela cuja curva caminhasse mais próxima dos dados originais da PNAD (tanto no nível quanto na estrutura) e cujo dado para o último trimestre de 2011 fosse o mais próximo do dado original da PNAD Contínua para o primeiro trimestre de 2012.

4. Resultados

O Quadro 1 apresenta os valores dos coeficientes, o erro padrão e a estatística t das desagregações dos ocupados da PNAD anual por meio do método de Chow e Lin (1971) e indica significância estatística para as três desagregações.

Quadro 1. Coeficientes, erro padrão, estatística t e nível de significância

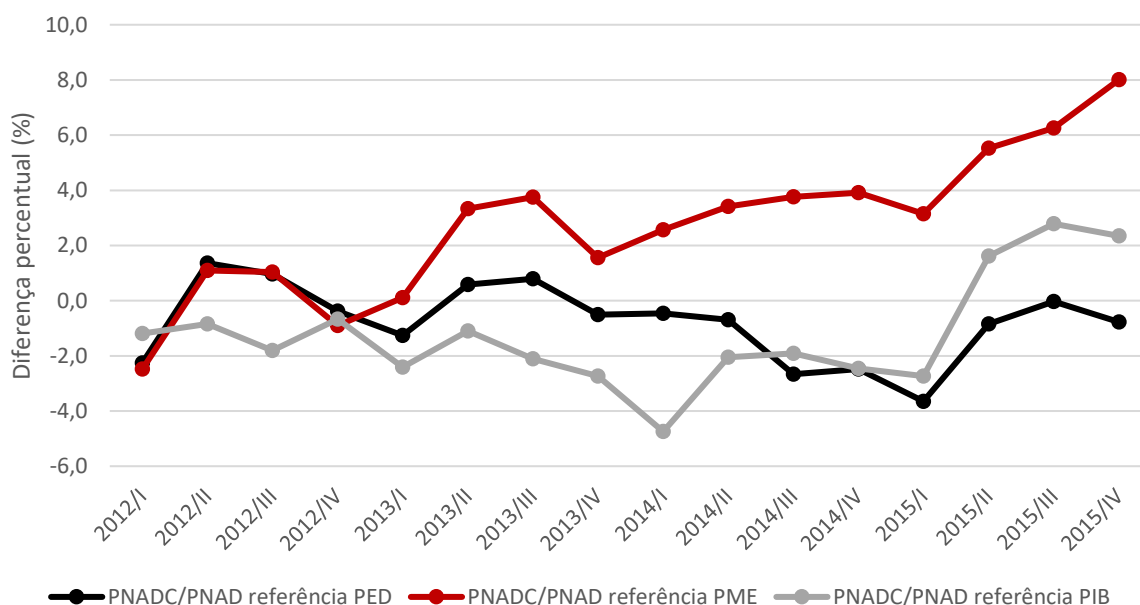
Variáveis dependentes	Coeficiente	Erro padrão	t valor	Pr (> t)
Ocupados - PNAD desagregada - ref. PED	6,1903	0,1548	39,98	$5,4 \cdot 10^{-15}$ (***)
Ocupados - PNAD desagregada - ref. PME	4,1007	0,3763	10,9	$6,59 \cdot 10^{-8}$ (***)
Ocupados - PNAD desagregada - ref. PIB	73,852	6,278	11,76	$2,65 \cdot 10^{-8}$ (***)

Fonte: elaboração própria. Nota: níveis de significância de 0 '****' 0.001 '***' 0.01 '**'.

O Gráfico 1 mostra as diferenças relativas entre as séries de ocupados das PNADs desagregadas, utilizando-se como referência os dados da PED, PME e do PIB, e da PNAD Contínua para o período de 2012 a 2015. Observa-se que na média do intervalo, a PNAD desagregada via PED registrou a menor diferença entre as

desagregações na comparação das pesquisas: -0,8%, com pico de divergência no primeiro trimestre de 2015 e mesmo valor no terceiro trimestre deste mesmo ano. O maior diferencial ocorreu com a desagregação que utilizou como referência a PME: média de desvio de 2,8%, com a maior diferença (8,0%) no quarto trimestre de 2015 e a menor (0,1%) no primeiro trimestre de 2013. Já a desagregação com base no PIB registrou um desvio médio de -1,3% na série de ocupados, com maior discrepância no primeiro trimestre de 2014 (-4,77%) e menor no quarto trimestre de 2012 (-0,7%).

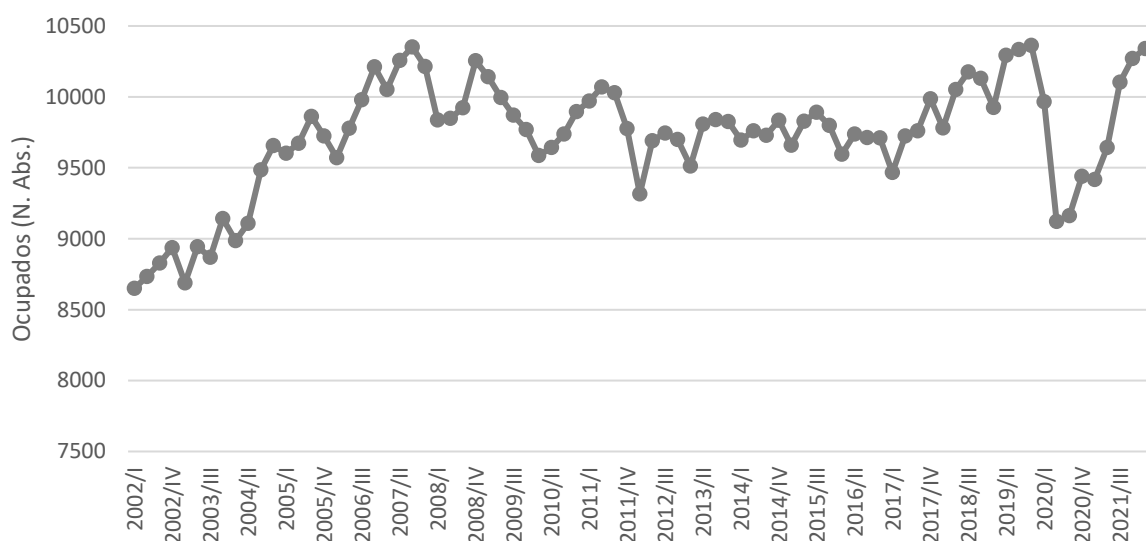
Gráfico 1. Diferenças relativas entre as séries de ocupados da PNAD Contínua e da PNAD desagregada pela PED, PME e PIB – % - 2012/I a 2015/IV



Fonte: elaboração própria a partir dos dados da PME, PED, PIB e PNAD Contínua.

Em virtude do menor diferencial médio, foi selecionada a série de ocupados da PNAD que utilizou a PED como referência para a estimação do coeficiente de regressão entre essa pesquisa e a PNAD Contínua. Tal coeficiente (0,9921), significativo ao nível de 0,1%, foi gerado por meio de regressão linear por mínimos quadrados ordinários (MQO) para o período coincidente das pesquisas (2012 a 2015) e aplicado à série retropolada da PNAD para estimar o contingente de ocupados no estado para o intervalo de 2002 a 2022 (Gráfico 2).

Gráfico 2. Série de ocupados da PNAD Contínua oficial (2012 a 2021) e retropolada com base na PNAD desagregada com referência na PED (2002 a 2011) – em mil



Fonte: elaboração própria até o quarto trimestre de 2011. A partir do primeiro trimestre de 2012, PNAD Contínua.

Pelo Quadro 2 observa-se os valores dos coeficientes, dos erros padrões e das estatísticas t das desagregações dos desocupados da PNAD anual por meio do método de Chow e Lin (1971), bem como a significância estatística para as três desagregações. Destaca-se que, diferentemente dos ocupados, as séries dos desocupados que utilizaram a PED e a PME como referência foram transformadas em logaritmo neperiano, uma vez que não houve significância estatística entre as séries da PNAD Contínua e das pesquisas supracitadas para os dados em nível. A série do PIB foi a única que gerou coeficiente significativo com os valores absolutos.

Quadro 2. Coeficientes, erro padrão, estatística t e nível de significância

Variáveis dependentes	Coeficiente	Erro padrão	Valor	Pr (> t)
Desocupados - PNAD desagregada - ref. PED	1,3186 (log)	0,0885	14,90	$1,5 \cdot 10^{-9}$ (***)
Desocupados - PNAD desagregada - ref. PME	1,2936 (log)	0,04399	29,41	$2,8 \cdot 10^{-13}$ (***)
Desocupados - PNAD desagregada - ref. PIB	6,8960	0,7732	8,92	$6,67 \cdot 10^{-7}$ (***)

Fonte: elaboração própria. Nota: níveis de significância de 0 '****' 0.001 '***' 0.01 '**'.

No caso da série de desocupados, as desagregações da PNAD com referência na PED, PME e PIB resultaram em desvios mais amplos em relação aos dados da PNAD Contínua. Na desagregação com referência na PED, a média dos desvios foi de -33,3%, sendo o maior deles igual a -52,6%, no primeiro trimestre de 2014, e

havendo apenas um trimestre, o quarto de 2015, em que o dado estimado foi igual ao da PNAD Contínua. A desagregação que considerou o PIB como série indicadora também apresentou desvios significativos: na média, eles foram de -27,1%, sendo o maior de 43,9%, no quarto trimestre de 2013, e o menor de -1,4%, no terceiro trimestre de 2015. Por sua vez, a desagregação com base na PME resultou em diferencial médio de 6,2%, com o maior deles no quarto trimestre de 2014, de 55,5%, e o menor no quarto trimestre de 2013, de -1,5%.

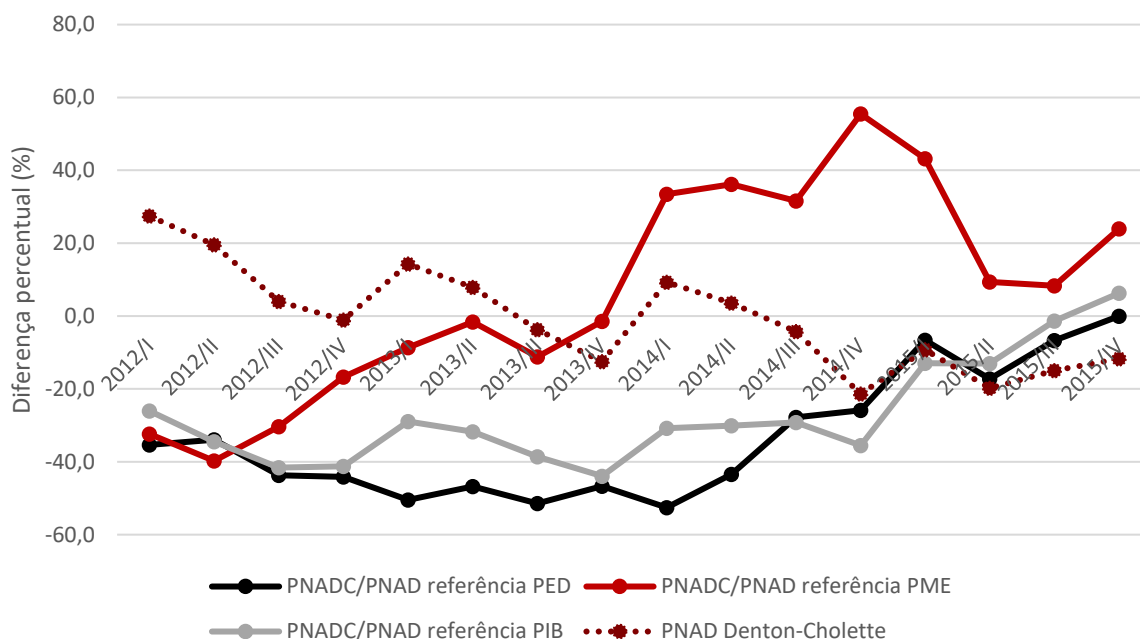
Destaca-se que os elevados diferenciais da PNAD desagregada utilizando-se a PED e a PME eram esperados, haja vista as discrepâncias de nível entre as taxas de desocupação dessas duas pesquisas e a PNAD Contínua, conforme mencionado anteriormente. Já em relação ao PIB, as divergências poderiam ser fruto da ausência de correspondência entre os dados do produto e os da desocupação. Embora haja convergência em alguns momentos entre 2002 e 2015 (o aumento do número de desocupados mensurados pela PNAD entre 2009 e 2010, coincidente com a queda do produto por decorrência da crise econômica global de 2008, bem como o aumento dos desocupados a partir de 2014, quando a série do produto interrompe a tendência de alta e entra em declínio), há também períodos nos quais esses dados não seguem tendências dentro do que se poderia esperar (entre 2002 e 2008, por exemplo, o PIB apresenta tendência de crescimento, mas os dados de desocupação permanecem estáveis entre 2002 e 2005, reduzindo apenas a partir de 2006). Divergências desse tipo são compreensíveis, pois é natural que haja alguma defasagem entre a queda do investimento e produção, de um lado, e a dinâmica do mercado de trabalho, de outro. Porém, o efeito estatístico dessas discrepâncias na mecânica dos métodos de desagregação pode ser de tal monta que resulte em séries desagregadas com estrutura e níveis distintos do que seria esperado.

Vale destacar, nesse sentido, as restrições inerentes a cada método de desagregação, bem como do próprio pacote “tempdisagg”, utilizado neste trabalho. Um exemplo disso é a necessidade de se indicar, entre os argumentos da função que opera a desagregação, o modo de conversão que, tal como descrito pelos criadores desse pacote (SAX e STEINER, 2013), “indica se os valores [da série] de baixa frequência são somas, médias, primeiros ou últimos valores dos valores [da série] de alta frequência”. Noutras palavras, o argumento em questão informa se, para cada valor da série original de baixa frequência, os valores da série de alta frequência gerados devam corresponder à divisão dos valores originais, à média desses valores ou, ainda, se os valores iniciais ou finais da série desagregada devam ser fixos. Neste trabalho, optou-se pela conversão que resultasse em valores médios da série original de modo a manter, na série desagregada, o mesmo nível da original. No entanto, como a função de desagregação não permite fixar valores em conjunto com o argumento, os resultados foram séries desagregadas que, qualquer que fosse o indicador ou método utilizado, não mantiveram os registros da PNAD original fixos, isto é, as curvas estimadas não passaram necessariamente pelos pontos originais. Trata-se, portanto, de restrição incontornáveis, que poderão ser superadas à medida em que o pacote de desagregação for atualizado ou novos pacotes surjam para esse tipo de operação.

O Gráfico 3 exibe as curvas geradas por essas desagregações, em conjunto com outra gerada por uma quarta desagregação, feita pelo método de Denton-

Cholette e sem referência em nenhuma série³. Nessa última desagregação, a média dos desvios foi bem menor, de -0,8%, sendo o maior deles no primeiro trimestre de 2012, de 27,4%, e o menor no quarto trimestre do mesmo ano, de -1,1%.

Gráfico 3. Diferenças relativas entre as séries de desocupados da PNAD Contínua e da PNAD desagregada pela PED, PME, PIB e por ela mesma – % - 2012/I a 2015/IV

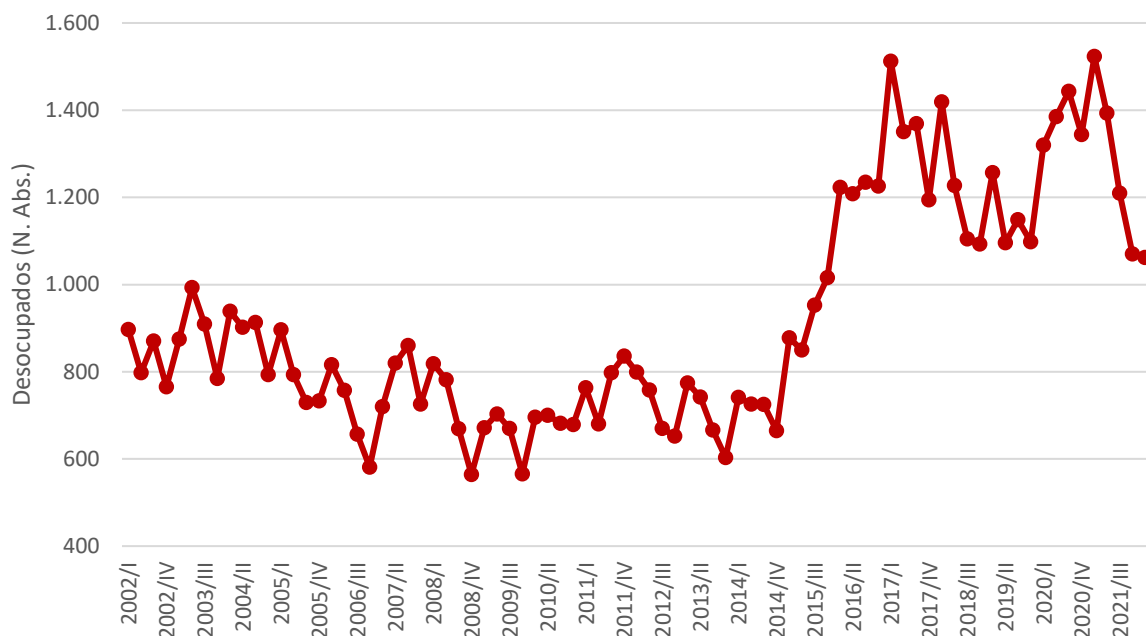


Fonte: elaboração própria a partir dos dados da PNAD, PME, PED, PIB e PNAD Contínua.

Mantendo-se o mesmo critério usado no caso da série de ocupados, optou-se por trabalhar, no caso dos desocupados, com duas séries desagregadas da PNAD. Para gerar a regressão linear com a PNAD Contínua e obter o estimador usado na retopolação, foi utilizada a série da PNAD desagregada pelo método de Denton-Cholette, em virtude do menor desvio médio em relação à série da PNAD Contínua entre 2012 e 2015. Já para a retopolação em si, optou-se pela série de desocupados desagregada com referência na PME, no intuito de não deixar de captar as oscilações no movimento da curva, que refletem as variações típicas do mercado de trabalho. O estimador de 0,9475 apresentou significância ao nível de 0,1% e gerou a série dos desocupados, conforme Gráfico 4.

³ Conforme Sax e Steiner (2013, p. 84), na prática, a desagregação sem nenhuma série indicadora como referência opera uma simples interpolação a partir dos dados originais da série de baixa frequência, levando em consideração a restrição do intervalo temporal da série.

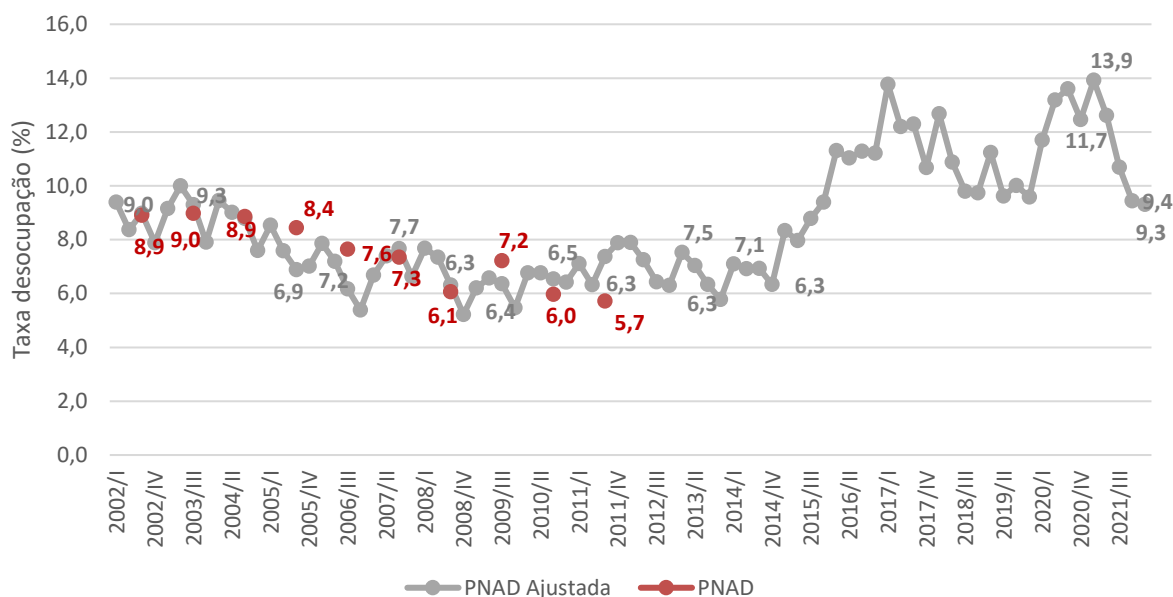
Gráfico 4. Série de desocupados da PNAD Contínua oficial (2012 a 2021) e retropolada com base na PNAD desagregada com referência na PME (2002 a 2011)



Fonte: elaboração própria até o quarto trimestre de 2011. A partir do primeiro trimestre de 2012, PNAD Contínua.

De posse das séries trimestrais de ocupados e desocupados para o período de 2002 a 2022, calculou-se a taxa de desocupação de Minas Gerais. O Gráfico 5 apresenta a taxa de desocupação para o estado de Minas Gerais para o horizonte de 20 anos e traz também a taxa de desocupação registrada pela PNAD entre 2002 e 2011 para comparação.

Gráfico 5. Taxas de desocupação de Minas Gerais, compatibilização conceitual e PNAD anual – 2002 a 2022



Fonte: elaboração própria da série compatibilizada. PNAD (2002 a 2011).

Verifica-se variação muito grande da taxa de desocupação de Minas Gerais ao longo do horizonte temporal analisado, com pico de 13,9% no primeiro trimestre de 2021 e menor valor no quarto trimestre de 2006. Observa-se também uma proximidade entre as séries, com diferença média entre as taxas anuais da PNAD e do terceiro trimestre da PNAD Ajustada⁴, no intervalo de 2002 a 2011, de 1,0%. A maior diferença entre as séries ocorreu em 2005, quando a PNAD registrou uma taxa de desocupação de 9,3%, enquanto a compatibilização indicou 6,9%.

5. Considerações finais

A interrupção da série histórica da PNAD tem representado um grande desafio para os estudiosos do mercado de trabalho no que diz respeito às séries longas de indicadores. Se por um lado, a PNAD Contínua, substituta da PME e da própria PNAD, trouxe aperfeiçoamentos conceituais e metodológicos, por outro é bastante recente e, por isso, sua amplitude temporal só será construída com o passar dos anos. Nesse sentido, este trabalho buscou contornar esta questão, propondo uma metodologia de retopolação estatística da taxa de desocupação para o estado de Minas Gerais, por meio de seus componentes (ocupados e desocupados) e de séries temporais de referência.

⁴ Comparou-se as taxas de desocupação compatibilizadas dos terceiros trimestres dos anos de 2002 à 2011 com as da PNAD anual como forma de aproximação, uma vez que a PNAD era realizada todo mês de setembro e é no terceiro trimestre de cada ano, da taxa compatibilizada, é que está setembro.

Embora não seja tarefa simples, o mesmo exercício pode ser replicado para outros indicadores calculados utilizando-se os dados da PNAD Contínua ou de outras pesquisas, seja por meio de retropolação ou extrapolação. Ressalta-se a importância dos pacotes estatísticos e o avanço dos softwares para a realização dos cálculos, democratizando o acesso das metodologias e permitindo a replicação dos resultados por diferentes pesquisadores.

6. Referências Bibliográficas

BACCIOTTI, R.; MARÇAL, E. F. Taxa de desemprego no Brasil em quatro décadas retropoladas da PNAD contínua de 1976 a 2016. *Estud.Econ.*, São Paulo, vol.50, n.3, p.513-534, 2020. Disponível em: <https://www.scielo.br/j/ee/a/YQcXMRc5xjPBgPHZztrnKDj/?lang=pt>. Acesso em: 3 abr. 2022.

CHOW, G.; LIN, An-loh. Best linear interpolation, distribution and extrapolation of time series by related series. *The Review of Economics and Statistics*, v.53, n.4, p.372-375, 1971. Disponível em: <https://www.jstor.org/stable/1928739>>. Acesso em: 20 maio 2022.

CORONA, F.; LÓPEZ-PÉREZ, J. Una evaluación econométrica de la retropolación de la actividad económica estatal de México. *Estudios Económicos (México, D.F.)*, v. 35, n. 2, p.193-212, 2020. El Colegio de México, A.C. Disponível em: http://www.scielo.org.mx/scielo.php?pid=S0186-72022020000200193&script=sci_abstract>. Acesso em: 11 abr. 2022.

CORSEUIL, C. H.; RAMOS L. Sobre a comparação da taxa de desocupação na PNAD Contínua e na PME. *Nota técnica* n. 18. Disponível em: http://repositorio.ipea.gov.br/bitstream/11058/5756/1/NT_n18_Sobre-comparacao-taxa-desocupacao_Disoc_2014-abr.pdf> Acesso em: 30 jun. 2022.

DENTON, F. 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, 1971. Disponível em: <https://www.oecd.org/sdd/21779760.pdf>>. Acesso em: 20 maio 2022.

FEU, A.; ALVIM, C. F. Retropolação" das Contas Nacionais até 1947: como compatibilizar os dados da nova série do Sistema de Contas Nacionais do IBGE com modelos de longo prazo, como o projetar_e. *Economia & Energia* Ano XI, n. 64, 2007 disponível em: <https://ecen.com/eee64/eee64p/eee64p.pdf>>. Acesso em: 9 abr. 2022.

FRIEDMAN, M. The Interpolation of Time Series by Related Series. *Journal of the American Statistical Association*, v.57, n.300, p.729-757, 1962. Disponível em: <https://www.jstor.org/stable/2281805>>. Acesso em: 20 maio 2022.

GERMAN-SOTO, V. 2005. Generación del producto interno bruto mexicano por entidad federativa, 1940-1992, *El Trimestre Económico*, 72(287): 617-653.

GUZMÁN, V. M. G.; VILLAVICENCIO, F. de J. C. Retropolación hasta 1980 del PIB trimestral de México por entidad federativa y gran actividad económica. Realidad, datos y espacio revista internacional de estadística y geografía, v.9, n.3, p.111-136. Disponível em: <https://rde.inegi.org.mx/index.php/2019/01/25/retropolacion-1980-del-pib-trimestral-mexico-entidad-federativa-gran-actividad-economica/>> Acesso em: 9 abr. 2022.

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA. Producto Interno Bruto trimestral, anual y anual por entidad federativa. Retropolación hasta 1980. Síntesis metodológica. Aguascalientes, México, INEGI, 2019. Disponível em: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825189167>>. Acesso em: 10 abr. 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Nota Metodológica nº 26

Revisão da série 1995-2001 (retropolação). Rio de Janeiro: IBGE, 2002. Disponível em: https://ftp.ibge.gov.br/Contas_Nacionais/Sistema_de_Contas_Nacionais/Notas_Metodologicas/26_RetropolacaoRegional.pdf>. Acesso em: 9 abr. 2022.

MARQUES, D. H. F.; *et al.* Evolução recente da taxa de desocupação em minas gerais nos últimos 20 anos: uma proposta de compatibilização entre a Pesquisa Nacional por Amostra de Domicílios (PNAD) e a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua)

MONTAGNER, P.; HAGA, A. Pesquisa de emprego e desemprego: sua importância como metodologia de pesquisa. São Paulo: São Paulo em Perspectiva, 17 (3-4), 2003. Disponível em: <https://www.scielo.br/j/spp/a/8R7Cwv3nCsSk5N5f9qTvZVn/?lang=pt>>. Acesso em: 15 jun. 2022.

SAX, C.; STEINER, P. Temporal disaggregation of time series. The R Journal, v. 5 (2): 80-87, 2013. Disponível em: <https://journal.r-project.org/archive/2013-2/sax-steiner.pdf>. Acesso em 13 jun. 2022.

VAZ, B. O. E.; BARREIRA, T. C. Metodologia de retropolação da pesquisa nacional por amostra de domicílios contínua de 1992 a 2012. Rio de Janeiro: FGV IBRE, 2016. Nota técnica. Disponível em: <https://ibre.fgv.br/observatorio-produtividade/artigos/metodologia-de-retropolacao-da-pesquisa-nacional-por-amostra-de>. Acesso em: 7 dez. 2021.