

Elasticidade de P&D na Indústria Brasileira e Transbordamentos: Uma abordagem de fatores comuns não-observados

Resumo

Este trabalho apresenta estimativas para a elasticidade de P&D na indústria brasileira, com dados de grupos industriais das indústrias extrativa e de transformação brasileiras entre os anos de 2003 e 2017. Para isso é utilizado o estimador *augmented mean group* (AMG) de Eberhardt e Bond (2009), que controla para a presença de fatores comuns correlacionados e potencialmente não-estacionários. A estimativa para elasticidade de P&D obtida por meio do estimador AMG foi igual a 0,014, porém não foi significativa estatisticamente. A comparação com estimativas produzidas por estimadores para dados em painel que desconsideram a presença de fatores comuns correlacionados demonstra a ocorrência de um viés positivo nas estimativas produzidas pelos estimadores convencionais originário da desconsideração de fatores que podem representar os transbordamento de conhecimento entre os setores. Conclusão semelhante é obtida por Eberhardt, Halmers e Strauss (2013) em uma amostra de 12 países da União Europeia e OCDE. Adicionalmente, foram calculadas para os grupos de setores segundo a intensidade tecnológica medidas de produtividade total dos fatores (PTF) a partir do método proposto por Eberhardt e Teal (2010). Nossas estimativas apresentam comportamento cíclico heterogêneo segundo o nível de intensidade tecnológica dos setores industriais.

Palavras-chaves: Elasticidade de P&D, fatores comuns não-observados, transbordamento de conhecimento, produtividade total dos fatores.

Códigos JEL: D24, O32, O33, O47

Abstract

This work presents estimates for the elasticity of R&D in the Brazilian industry, by data from industrial groups in the extractive and manufacturing industries in Brazil between 2003 and 2017. For this, we employ the *augmented mean group* estimator (AMG) of Eberhardt e Bond (2009), which controls for the presence of correlated and potentially non-stationary common factors. The estimate for R&D elasticity obtained through the AMG estimator was equal to 0.014, but it was not statistically significant.

The comparison with estimates produced by estimators for panel data that disregard the presence of the common correlated factors demonstrates a positive bias in the estimates produced by conventional estimators originating from the disregard of factors that may represent the overflow of knowledge between sectors. The same conclusion had been obtained by Eberhardt for a sample of 12 OECD and EU countries. Additionally, measures of total factor productivity (TFP) were calculated for the groups of sectors according to the technological intensity using the method proposed by Eberhardt e Teal (2010). Our estimates show a pro-cyclical behavior and a positive correlation with technological intensity.

Keywords: R&D elasticity, unobserved common factors, knowledge spillovers, total factor productivity

JEL Codes: D24, O32, O33, O47

Introdução

Políticas que buscam incentivar o gasto em atividades de pesquisa e desenvolvimento (P&D) baseiam-se na suposição de que existe uma relação direta entre esforço inovativo e crescimento da produtividade. Para avaliar a relação entre inovação e produtividade, muitos trabalhos estimam a elasticidade do produto com relação a gastos em P&D (GRILICHES, 1979). A abordagem mais utilizada, proposta por Griliches (1979), consiste em estimar uma função de produção incluindo como fator de produção uma medida de estoque ou dispêndio interno em P&D, além dos fatores de produção tradicionais, como capital e trabalho. O caráter de bem público do conhecimento torna necessária a inclusão de uma variável para representar o conhecimento externo adquirido, para controlar o efeito do transbordamento de conhecimento, ou externalidade tecnológica.

Em razão de não serem diretamente observáveis ou mensuráveis, os transbordamentos de conhecimento são capturados pelo termo de erro quando este fator é omitido da regressão. Devido ao investimento privado em P&D estar correlacionado com o transbordamento de conhecimento, a presença deste efeito no erro deve levar à endogeneidade da variável de P&D e inconsistência na estimação da elasticidade de P&D. Além disso, o estoque de conhecimento externo é uma variável comum a todos os setores, porém absorvido em diferentes intensidades de acordo com as diferentes capacidades e necessidades de absorção de conhecimento. Uma vez que o transbordamento está presente nos erros das equações para todos os setores, estes erros podem estar correlacionados entre si, propriedade conhecida como dependência *cross-section* (EBERHARDT; HELMERS; STRAUSS, 2013).

O presente estudo tem por objetivo estimar a elasticidade de P&D, controlando o efeito do transbordamento de conhecimento, e calcular a produtividade total dos fatores na indústria brasileira. Para isso, foi estimada uma função de produção a partir de um painel de dados com setores das indústrias extrativa e de transformação entre os anos de 2003 e 2017. Foi adotada a estrutura analítica proposta por Eberhardt e Teal (2010), que consiste em representar a função de produção por meio de um modelo de fatores comuns não-observados e utilizar o estimador *augmented mean group* (AMG) proposto por Eberhardt e Bond (2009).

Essa estrutura analítica permite a heterogeneidade dos parâmetros, bem como a não estacionariedade das variáveis de valor adicionado, de fatores de produção, e dos fatores

não observados. De acordo com Eberhardt e Teal (2010), é possível considerar vários cenários, como variáveis dependente e independentes não estacionárias e cointegradas, além da possibilidade de ao menos alguns fatores comuns serem não estacionários e cointegrados. Com isso, a estrutura empírica geral fornece flexibilidade máxima no que diz respeito às propriedades das séries temporais e em cross-section, tanto das variáveis, como dos fatores não observados. A adoção dessa representação constitui uma inovação para a literatura nacional na estimação de funções de produção

A fim de superar as limitações do modelo de Solow (1957), que supunha o progresso técnico como exógeno, os modelos de crescimento endógeno como Aghion e Howitt (1990), Romer (1990) e Grossman e Helpman (1991), buscam explicar como o investimento em P&D pode representar uma estratégia ótima das firmas para alcançar ganhos de produtividade. Nesses modelos, em que são consideradas as possibilidades de produção de conhecimento específico ao produto e de caráter geral à economia, fundamenta-se microeconomicamente a existência de externalidades do conhecimento¹. Duguet (2006), analisando dados da indústria francesa, indica que o sucesso da atividade inovativa depende não só do esforço interno de P&D, mas também do acesso ao conhecimento externo, através do uso de patentes ou licenças produzidas em outros setores, por exemplo. Higon (2007) também aponta para a importância do transbordamento para explicar o crescimento de setores da indústria britânica.

Wieser (2005) destaca que um dos problemas na estimação de retornos de P&D é a dificuldade de representar adequadamente o componente de conhecimentos externos. É usual encontrar na literatura variáveis proxies espelhadas na proximidade produtiva, ou tecnológica, entre os setores da economia. Hall, Mairesse e Mohnen (2010) notam que, em geral, as medidas propostas na literatura são muito imprecisas, enquanto que Griliches (1991) argumenta que é difícil saber exatamente como o conhecimento se espalha pela economia.

Eberhardt, Helmers e Strauss (2013) concluem que a representação do conhecimento por meio de variáveis proxy não é uma solução adequada para controlar a dependência *cross-section* na função de produção. Em primeiro lugar, por essa estratégia implicar em assumir hipóteses muito fortes sobre esta variável, que invariavelmente estariam sujeitas a críticas. Além disso, a inclusão de uma *proxy* para o conhecimento externo implica assumir que esta é a única fonte de dependência *cross-section* nos dados. Os autores argumentam que esta propriedade dos erros também pode ser gerada pela reação dos setores a choques econômicos comuns, como crises financeiras globais, bem como pela interação entre os setores por meio de relações comerciais, institucionais *etc.* Assim, a *proxy* para o transbordamento tecnológico não capturaria totalmente a dependência *cross-section*. Mais ainda, o investimento privado em P&D pode ser influenciado por choques econômicos ou de produtividade e relações intersetoriais, por exemplo, uma vez que tais fatores podem afetar a eficiência da pesquisa e a disponibilidade de recursos para este tipo de investimento. Neste caso, a *proxy* também não resolveria o problema de endogeneidade da variável de P&D.

Chudik, Pesaran e Tosetti (2011) mostram que a dependência *cross-section* em modelos de dados em painel pode ser representada por um modelo de fatores comuns, ou seja, efeitos temporais comuns que podem afetar não uniformemente todas as unidades do painel. Em razão das críticas ao uso de variáveis proxy para o conhecimento externo, Eberhardt, Helmers e Strauss (2013) argumentam que a representação da função de pro-

¹Uma outra possibilidade de incorporação do progresso tecnológico endógeno pode ser encontrada na literatura evolucionária ou neo-schumpeteriana (NELSON, 1985; DOSI; NELSON, 1994)

dução como um modelo de fatores comuns não-observados é mais adequada para estimar a elasticidade de P&D. Neste caso, assume-se que o erro da regressão é composto por uma combinação de fatores comuns, que capturam não apenas o transbordamento de conhecimento, mas também poderiam captar choques de oferta e demanda, entre outros efeitos temporais comuns. Eberhardt e Teal (2010) mostram, ainda, como o modelo de fatores comuns não-observados configura uma forma simples de se modelar as relações estruturais implícitas à função de produção, como a relação entre a demanda por insumos e a produtividade das firmas, bem como a possível dependência temporal das variáveis do modelo.

Com relação às estimativas obtidas na literatura, Hall, Mairesse e Mohnen (2010) e Wieser (2005) analisam uma série de estudos com dados de firmas e indústrias. Seus resultados apontam valores para a elasticidade de P&D em torno de 0,1.² Eberhardt, Helmers e Strauss (2013) estimam a elasticidade de P&D em um painel com setores de diferentes países da OCDE aplicando a abordagem de fatores comuns latentes na função de produção. Seus resultados indicam que, quando a presença de fatores comuns é controlada, a elasticidade de P&D passa a ser não-significante estatisticamente. Segundo os autores, isso evidencia a dificuldade de se separar o retorno privado de P&D do efeito de transbordamento.

No caso brasileiro, Cavalcante, Jacinto e Negri (2015) utilizam dados da Pesquisa Nacional de Inovação Tecnológica (Pintec) para estimar a relação entre esforço inovativo e produtividade do trabalho na indústria brasileira. Os autores reportam valores entre 0,06 e 0,1 para a elasticidade de P&D na indústria brasileira entre 2000 e 2008, sendo os setores de alta intensidade tecnológica os que apresentam os maiores valores. Este resultado indica que a diferença na intensidade tecnológica dos setores pode levar a uma heterogeneidade nas elasticidades de P&D. Já Zucoloto et al. (2017), em um estudo que busca avaliar a efetividade da Lei do Bem, estimam uma elasticidade do P&D igual a 0,06 para o período de 2006 a 2013. Esse valor estaria próximo à mediana dos estudos internacionais, conforme apontado por Wieser (2005). Ambos trabalhos têm o mérito de produzirem estimativas a partir de microdados de empresas industriais. No entanto, ambos desconsideram o componente de transbordamento tecnológico em suas estimações. Neste sentido, Steingraber e Gonçalves (2010) apontam indícios de que a produtividade e o esforço inovativo na indústria brasileira apresentam forte componente setorial, sendo a variância destas variáveis menor entre firmas do mesmo setor que entre firmas de setores diferentes. Os autores encontram ainda evidências de que os diferenciais de produtividade entre firmas da indústria brasileira estão associados à cooperação com outras empresas e centros de pesquisa, indicando que o transbordamento de tecnologia entre firmas de um mesmo setor é um componente relevante da relação entre pesquisa e produtividade no Brasil.

A discussão acima justifica a preocupação deste estudo em utilizar um estimador que considere uma estrutura de fatores comuns não-observados na função de produção. Neste sentido, este trabalho contribui para a literatura econômica brasileira com a produção de estimativas atualizadas da elasticidade de P&D por meio de metodologia empírica robusta, além da investigação da relação entre o transbordamento de conhecimento, gastos em P&D e produtividade. Além disso, os resultados aqui apresentados ajudam a entender a relação entre gastos em P&D e a estagnação da produtividade industrial no Brasil, podendo subsidiar a formulação de políticas mais eficientes de crescimento e fomento à inovação.

²A elasticidade varia entre -0,015 e 0,38 nos trabalhos avaliados por esses autores

A elasticidade de P&D obtida pelo estimador AMG foi de 0,014, porém não foi estatisticamente significativa. A comparação com outros estimadores indica que não controlar a dependência *cross-section* nos dados produz viés positivo na estimativa da elasticidade de P&D, além de, em alguns casos, produzir resultados estatisticamente significantes³. Este resultado corrobora as evidências fornecidas por Eberhardt, Helmers e Strauss (2013) para um painel de 12 países contendo dados de 12 indústrias manufatureiras cobrindo o período de 1980-2005. Como discutido por esses autores, isso evidencia a dificuldade de separar o retorno do investimento em P&D do efeito do transbordamento de conhecimento e outros choques econômicos comuns.

Os resultados do AMG foram utilizados para calcular a produtividade total dos fatores (PTF) dos setores analisados como proposto por Eberhardt e Teal (2010). Observa-se uma correlação positiva da intensidade tecnológica dos setores com o nível da PTF, mas não tanto com seu crescimento. Além das externalidades produzidas pelos transbordamentos de conhecimento, é demonstrado que a evolução da produtividade da indústria tem relação heterogênea com o componente cíclico do produto. Esse resultados estão em concordância com Nelson (1988) que argumenta sobre a dificuldade de se estabelecer uma relação de causalidade entre investimentos privados em P&D e a produtividade nas indústrias, destacando para isso o papel dos transbordamentos de conhecimento.

O restante deste trabalho está organizado da seguinte forma. Na seção 1, é descrita a estrutura analítica utilizada para modelar a função de produção. Na seção 2 é apresentado o estimador AMG e suas hipóteses subjacentes. A seção 3 contém uma descrição da base de dados e das variáveis utilizadas nas estimações. Na seção 4 são apresentados e discutidos os resultados obtidos. A última seção conclui sobre as evidências obtidas e discute as implicações destes resultados para a formulação de políticas públicas.

1 Estrutura Analítica

Segundo Gunday et al. (2011), a inovação tecnológica pode afetar a produção das firmas de várias formas. Do ponto de vista da oferta, por exemplo, inovações organizacionais e de processo levam as firmas a serem mais eficientes, podendo produzir mais a custos menores. Além disso, inovações de produto, seja na forma de melhoria na qualidade de produtos existentes, seja por meio da introdução de novas mercadorias, podem aumentar a demanda das firmas, o que também tende a causar um aumento na produção. Tais fatores representam uma série de incentivos para as firmas investirem em P&D visando a produção de inovações.

Por outro lado, Arrow (1972) argumenta que as firmas não conseguem se apropriar plenamente do conhecimento que produzem, uma vez que não é possível controlar a difusão do conhecimento pela economia. É possível aprender parte do conhecimento produzido por outras firmas por meio da imitação, troca de informação entre trabalhadores, limitações do sistema de patentes *etc.* Este efeito de transbordamento do conhecimento indica que, além do gasto próprio em P&D, a produção externa de conhecimento também pode afetar a produtividade das firmas.

Além disso, o transbordamento também pode afetar a decisão de investimento em P&D das firmas. Arrow (1972) nota que o fato do conhecimento ser parcialmente apropriável, bem como a incerteza associada à produção de inovação, podem representar um desincentivo à alocação de recursos para pesquisa. Por outro lado, Cohen e Levinthal

³Os estimadores empregados neste estudo são descritos com mais detalhe na seção 2

(1989) argumentam que as firmas podem investir em P&D para aumentar sua habilidade de se apropriar do conhecimento produzido externamente. Neste caso, o transbordamento pode representar um incentivo para o gasto com inovação. Seria razoável, então, supôr que a produção de conhecimento interno e o transbordamento estejam correlacionados.

Segundo Hall, Mairesse e Mohnen (2010), a forma padrão na literatura de se modelar a elasticidade de P&D é por meio da função de produção. Tal abordagem consiste em relacionar o produto de determinado setor com seu estoque de conhecimento interno e outros fatores de produção como capital e trabalho. Inclui-se na análise, ainda, um estoque de conhecimento externo comum aos setores como forma de levar em conta o impacto dos transbordamentos de conhecimento. Considere a seguinte função de produção *Cobb-Douglas*:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} R_{it}^{\gamma} [R_{it}^0]^{\phi} e^{u_{it}}, \quad (1)$$

em que Y_{it} representa o produto do setor i no período t , K_{it} representa seu estoque de capital, L_{it} representa o trabalho, R_{it} e R_{it}^0 representam o estoque de conhecimento interno e externo, respectivamente, A_{it} representa a produtividade total dos fatores (PTF) e u_{it} é um termo de erro aleatório. Os parâmetros α , β e γ capturam as elasticidades do capital, trabalho e P&D. Aplicando logaritmo sobre (1), obtém-se

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma r_{it} + \phi r_{it}^0 + \eta_i + \theta_t + u_{it}, \quad (2)$$

em que as letras minúsculas indicam valores em logaritmo e assume-se, como em Hall, Mairesse e Mohnen (2010), que $\log(A_{it}) = \eta_i + \theta_t$. Ou seja, assume-se que a PTF apresenta diferenças de nível entre setores, capturadas pelo efeito fixo η_i , e uma evolução comum, ditada por θ_t .

Eberhardt e Teal (2011) revisam uma série de trabalhos que estimam funções de produção, ressaltando como cada estudo modela a PTF. Esta análise indica que uma das formas mais comuns de se modelar a evolução temporal da produtividade é incluir uma *dummy* temporal ou tendência determinística. Os autores notam, porém, que tais soluções apresentam problemas. A inclusão da *dummy* implica assumir que a evolução da PTF é a mesma para todos os setores. Se este não for o caso, Eberhardt e Teal (2011) argumentam que a *dummy* capturaria apenas uma evolução média da PTF, não corrigindo completamente para a correlação deste componente com os fatores de produção. Como alguns dos fatores que determinam a PTF também influenciam a escolha de fatores de produção das firmas, incluindo o investimento em P&D, e a *dummy* temporal captura apenas parcialmente a evolução da PTF, esta variável não seria suficiente para controlar a endogeneidade dos fatores de produção.⁴

Como solução para estes problemas, Eberhardt, Helmers e Strauss (2013) sugerem que o transbordamento pode ser visto como um estoque de conhecimento que impacta de modo heterogêneo os setores e que a consequente evolução da PTF seja determinada por uma combinação de choques econômicos comuns. Nesse caso, a produtividade dos setores difere por causa de suas características individuais e de diferenças no impacto dos choques comuns. A abordagem de fatores comuns considera a dependência da seção transversal como resultado de variáveis comuns omitidas não observadas e variáveis

⁴Exemplos de possíveis determinantes da produtividade associados à escolha de insumos incluem fatores específicos dos setores como qualidade da mão-de-obra, bem como choques econômicos comuns como crises financeiras ou políticas econômicas (FAGERBERG, 1994; GARCIA, 2003)

temporais ou choques comuns que afetam cada unidade da seção transversal de maneira diferente. Com base nessas considerações, a equação (2) pode ser reescrita como

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma r_{it} + \eta_i + \Lambda'_i \mathbf{f}_t + u_{it}, \quad (3)$$

em que $\Lambda'_i \mathbf{f}_t = \sum_{j=1}^M \lambda_{ji} f_{jt}$ e M é o número de fatores comuns. O termo \mathbf{f}_t representa um vetor de fatores comuns associados à PTF, bem como transbordamentos de conhecimento. Já Λ'_i é o vetor de parâmetros associados aos fatores latentes. Este termo indica que estes fatores podem ter impacto diferente entre os setores.

A presença de uma estrutura fatorial comum latente leva à correlação entre os erros dos setores, fenômeno conhecido como dependência *cross-section*. Pesaran e Tosetti (2009) notam que, caso haja correlação entre os regressores e alguns dos fatores em \mathbf{f}_t , estimadores que ignoram a estrutura fatorial do erro serão inconsistentes. Como discutido, é razoável supor que existe correlação entre a variável de P&D e o estoque de conhecimento comum e entre a escolha dos insumos e fatores comuns associados à produtividade. Eberhardt e Teal (2010) notam, ainda, que a forte persistência temporal observada empiricamente nas variáveis da função de produção pode ser consequência de choques comuns persistentes de produtividade. Assim, o termo \mathbf{f}_t permite controlar a dependência *cross-section*, a endogeneidade e a persistência temporal das variáveis em (3).

Especificamente no que tange à persistência temporal é necessário destacar que a persistência de choques de produtividade é uma suposição usualmente adotada na literatura econômica de ciclos econômicos reais. Nessa literatura, as inovações em tecnologia são consideradas altamente persistentes ou então permanentes. Desde os modelos básicos de Kydland e Prescott (1982), que ressalta a importância da suposição de requerimento de “tempo para construir”, e de King, Plosser e Rebelo (1988), ou ainda em suas extensões, como o trabalho de Christiano e Eichenbaum (1992), que permite que choques de consumo do governo influenciem a dinâmica do mercado de trabalho, a suposição de choques de produtividade persistentes, ou permanentes, é necessária para a reprodução de ciclos econômicos. Entretanto, a capacidade de reprodução de fatos estilizados apresenta limitações. Uma delas é a dificuldade de dar conta das propriedades dinâmicas produto, em razão dos mecanismos de propagação dos choques serem geralmente fracos. Cogley e Nason (1995) testam adequação de diversos modelos presentes na literatura e concluem que os achados mais promissores, ainda que incompletos, estão nos modelos com acúmulo de trabalho. Ainda assim, essa literatura representa importante contribuição à teoria econômica, como também do ponto de vista metodológico ao privilegiar o uso de simulações para o teste de teoria. Para mais detalhes sobre essa discussão ver Stadler (1994) e Plosser (1989). Sendo assim, a representação adotada para a evolução da PTF é condizente também com esse campo da teoria econômica.

Outra hipótese importante assumida em (3) é a homogeneidade dos parâmetros, que implica assumir a mesma função de produção para cada setor. No caso da análise de produtividade entre países, Harberger (1987) argumenta que é pouco razoável assumir uma tecnologia de produção comum entre diferentes países, com instituições, economia e geografia diversas. Mesmo no caso de setores, Hall, Mairesse e Mohnen (2010) argumentam, por exemplo, que o impacto de P&D na produção depende de vários fatores específicos dos setores e firmas analisados, de modo que possivelmente tal impacto não seria único entre diferentes atividades econômicas. A imposição equivocada de homo-

geneidade dos parâmetros implicaria na inconsistência nos parâmetros estimados.⁵

Com base nessas considerações, Eberhardt e Teal (2010) propõem o seguinte modelo analítico para a função de produção:

$$y_{it} = \mathbf{B}'_i \mathbf{x}_{it} + \eta_i + \mathbf{\Lambda}'_i \mathbf{f}_t + u_{it}, \quad (4)$$

$$\mathbf{x}_{it} = \mathbf{\Pi}'_i \mathbf{g}_t + \mathbf{\Psi}'_i \mathbf{f}_t + v_{it}, \quad (5)$$

$$\mathbf{f}_t = \mathbf{P}' \mathbf{f}_{t-1} + \epsilon_{it}, \quad (6)$$

$$\mathbf{g}_t = \mathbf{K}' \mathbf{g}_{t-1} + \varepsilon_{it}, \quad (7)$$

em que \mathbf{x}_{it} é o vetor de insumos, \mathbf{B}_i é a matriz de elasticidades dos insumos, \mathbf{g}_t é o vetor de fatores comuns específicos às variáveis, $\mathbf{\Pi}_i$ captura o impacto desses fatores específicos, \mathbf{f}_t é o vetor de fatores comuns a todos os setores e $\mathbf{\Psi}_i$ mede a relação entre \mathbf{f}_t e as variáveis. Os vetores de parâmetros \mathbf{P} e \mathbf{K} capturam a persistência temporal dos fatores comuns e v_{it} , ϵ_{it} e ε_{it} são erros idiossincráticos.

As características mais importantes desta configuração são a potencial não estacionariedade das variáveis observáveis e dos componentes não observáveis, e a heterogeneidade potencial no impacto das variáveis observáveis e dos componentes não observáveis. O modelo (4)-(7) permite controlar a dependência *cross-section* gerada por choques comuns de produtividade e transbordamentos de conhecimento. Adicionalmente, atende à necessidade de evitar, ao menos parcialmente, a endogeneidade das variáveis explicativas gerada pela presença de fatores comuns. A principal vantagem desta abordagem é considerar todos estes efeitos de forma simples, uma vez que os fatores comuns capturam boa parte da estrutura dos dados. Assim, é possível evitar os problemas econométricos implicados por eles de forma mais parcimoniosa.

A utilização de uma estrutura de painel dinâmico e do modelo de fatores comuns representa ainda uma contribuição importante à literatura brasileira de análise de produtividade e inovação. Enquanto muitos trabalhos que analisam a evolução da PTF focam em setores específicos e na estrutura de autocorrelação de suas séries, sem muita atenção para as relações entre setores, outros avaliam o impacto das interações entre setores na PTF e inovação sem muito foco na dinâmica destas variáveis, com dados em *cross-section* ou painéis curtos. Esta pesquisa une estes dois componentes, capturando a dinâmica das séries individuais, as correlações entre os setores e a evolução destas relações ao longo do tempo.

2 Metodologia

Segundo Eberhardt e Teal (2011), as três formas mais recorrentes de se controlar a presença de fatores comuns não-observados em modelos de dados em painel são por meio da inclusão de *dummies* de tempo na regressão, da modelagem da dependência *cross-section* gerada pela estrutura fatorial ou da utilização de estimadores que controlem explicitamente os fatores comuns. Coakley, Fuertes e Smith (2006) notam que a primeira solução implica assumir que o impacto dos fatores comuns é homogêneo. Como discutido acima, tal hipótese pode levar a um problema de regressão espúria caso as variáveis analisadas sejam não-estacionárias. Para a aplicação da segunda solução, Eberhardt,

⁵Note que, sob a hipótese de heterogeneidade dos parâmetros, o erro da equação (3) seria dado por $u_{it} = (\alpha - \alpha_i)k_{it} + (\beta - \beta_i)l_{it} + (\gamma - \gamma_i)r_{it}$, o que implicaria em endogeneidade das variáveis explicativas

Helmers e Strauss (2013) notam que é necessário entender como se dão as relações entre os setores e saber quais são os fatores associados a esta dependência, o que, em geral, não é possível. Por fim, Eberhardt e Teal (2011) argumentam que estimadores que controlam explicitamente os fatores comuns evitam estes problemas permitindo formas mais flexíveis para a estrutura fatorial latente.

Estimadores de painel com fatores comuns latentes, em geral, se dividem em dois grupos. O primeiro grupo consiste em estimar conjuntamente os parâmetros da regressão e os fatores comuns, como proposto por Bai (2009), entre outros. Segundo Chudik e Pesaran (2013), esta estratégia é limitada pois é preciso determinar o número de fatores a ser estimado, o que introduz incerteza na estimação. Já o segundo grupo consiste em aumentar a regressão original com alguma *proxy* dos fatores comuns. Pesaran (2006) propõe o estimador *common correlated effects* (CCE), que consiste em aproximar os fatores comuns por médias *cross-section* da variável dependente e dos regressores. Eberhardt e Teal (2008) ressaltam que o CCE é intensivo no uso de dados porque a quantidade de médias *cross-section* para se incluir na regressão aumenta com o número de regressores, tornando a estimação menos precisa. Eberhardt, Helmers e Strauss (2013) notam ainda que a estratégia de aproximar os fatores comuns pelas médias *cross-section* das variáveis não permite identificar os fatores e seus parâmetros.

O estimador AMG de Eberhardt e Bond (2009) considera a dependência em *cross section* por meio da inclusão de um processo dinâmico comum (CDP) na regressão. O estimador consiste em um procedimento de dois estágios. No primeiro, estima-se a função de produção com um modelo em primeiras diferenças incluindo $T - 1$ *dummies* de ano, também em primeiras diferenças:

$$\Delta y_{it} = \mathbf{B}' \Delta \mathbf{x}_{it} + \sum_{t=2}^T c_t \Delta D_t + \varepsilon_{it}, \quad (8)$$

em que \mathbf{x}_{it} é o vetor de insumos (capital, trabalho e P&D), ΔD_t são as *dummies* de tempo em primeira diferença e c_t são os coeficientes associados a elas. Segundo Eberhardt e Bond (2009), os valores estimados dos coeficientes das *dummies*, \hat{c}_t , podem ser interpretados como estimativas do processo dinâmico comum, que captaria a evolução média dos fatores comuns e são renomeados como $\hat{\mu}_t^*$.

No segundo estágio, $\hat{\mu}_t^*$ é incluído como um regressor adicional na função de produção:

$$y_{it} = \mathbf{B}' \mathbf{x}_{it} + \delta_i \hat{\mu}_t^* + d_i t + \varepsilon_{it}, \quad (9)$$

em que t é uma tendência determinística, incluída para controlar possíveis efeitos temporais não captados pelo processo dinâmico comum. Estima-se então (9) pelo estimador *mean group* (MG) de Pesaran e Smith (1995), que assume parâmetros heterogêneos. Como só adiciona uma variável à estimação, o AMG evita a perda de graus de liberdade causada pelo CCE.

Segundo Eberhardt e Bond (2009), como o processo comum é estimado por um modelo em primeiras diferenças, obtém-se uma estimativa consistente da evolução dos fatores comuns mesmo que estes sejam não-estacionários. Posteriormente, a estimativa do processo dinâmico comum passa a fazer parte da relação de cointegração heterogênea específica a cada grupo. Desta forma, a inclusão do processo comum na regressão original também permite estimar relações de cointegração específicas a cada unidade do painel. Como o AMG controla a presença de fatores comuns no erro e sua possível

não-estacionariedade, espera-se que o estimador produza uma estimativa consistente da elasticidade de P&D.

Eberhardt e Teal (2010) mostram ainda como as estimativas produzidas do processo dinâmico comum possibilitam um cálculo da PTF mais robusto que os métodos mais tradicionais para dados agregados.⁶ Para entender este ponto, considere uma função de produção *Cobb-Douglas* log-linear (3):

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma r_{it} + a_{it} + u_{it}, \quad (10)$$

em que $a_{it} = \eta_i + \Delta_t \mathbf{f}_t$ seria a representação para a produtividade total dos fatores. Segundo Beveren (2012), como a PTF não é observável, a forma usual de estimá-la consiste em calcular o resíduo da estimação da função de produção:

$$\hat{a}_{it} = y_{it} - \hat{\alpha} k_{it} - \hat{\beta} l_{it} - \hat{\gamma} r_{it}, \quad (11)$$

em que $\hat{a}_{it} = \hat{\eta}_i + \widehat{\Delta_t \mathbf{f}_t}$, os parâmetros $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ são as elasticidades estimadas e $\hat{\eta}_i$ é o efeito fixo individual estimado, que pode ser visto como um intercepto específico do setor i .

O autor nota que uma das maiores preocupações ao se estimar funções de produção é a simultaneidade entre a escolha dos fatores de produção e a produtividade. Quando não controlado, este efeito leva à endogeneidade dos fatores na regressão, produzindo viés nas estimativas. Da equação (11) é possível notar que estimativas viesadas dos parâmetros implicam também viés no cálculo da PTF.

Blundell e Bond (1998) propõem controlar a endogeneidade por meio de uma estimação de GMM em que as defasagens e primeiras diferenças dos insumos são utilizadas como instrumentos, estimador conhecido como *system*-GMM. A principal vantagem deste método é que as primeiras diferenças são instrumentos fortes mesmo sob não-estacionariedade das variáveis. Por outro lado, ainda se impõe homogeneidade dos parâmetros, o que implica assumir que a evolução da PTF é comum entre os setores.

Eberhardt e Teal (2010) argumentam que o AMG permite um cálculo mais robusto da PTF por dois motivos. Primeiro, o estimador controla a endogeneidade causada pela dependência *cross-section* dos dados, bem como a não-estacionariedade das variáveis e a heterogeneidade dos parâmetros. Assim, as elasticidades estimadas pelo AMG – e cálculos da PTF baseados nelas – seriam mais confiáveis. Além disso, os autores notam que a tendência comum estimada no primeiro estágio do AMG pode ser vista como uma *proxy* da evolução dos fatores comuns associados à PTF. Assim, tem-se uma estimativa da evolução da PTF, ditada pela dinâmica dos fatores comuns latentes.

A partir da estimação de (9), Eberhardt e Teal (2010) propõem calcular a evolução da PTF de cada setor da seguinte forma:

$$a_{i,0} = \hat{\eta}_i + \hat{\alpha}_i k_{i,0} + \hat{\beta}_i l_{i,0} + \hat{\gamma}_i r_{i,0}, \quad (12)$$

$$a_{i,\tau} = a_{i,0} + \hat{\delta}_i \hat{\mu}_\tau^* + \hat{d}_i \tau, \quad (13)$$

em que a é o logaritmo da PTF, $\hat{\eta}_i$ é o intercepto de cada setor, $\tau = 1, \dots, T$ é o período em questão, $\hat{\mu}_\tau^*$ é a tendência comum estimada no primeiro estágio do AMG, $\hat{\delta}_i$ é o coeficiente estimado para a tendência comum no segundo estágio e \hat{d}_i é a estimativa

⁶Uma discussão mais detalhada sobre os métodos de cálculo da PTF pode ser vista, por exemplo, em Beveren (2012)

do parâmetro da tendência determinística incluída no segundo estágio.

O cálculo proposto difere do tradicional, descrito pela equação (11), de duas formas. Primeiro, é usada a tendência comum estimada pelo AMG para calcular a evolução da PTF. Além disso, o método proposto também difere com relação ao cálculo do nível inicial da PTF. Neste caso, o cálculo a partir do AMG leva em conta, além do intercepto individual, as elasticidades individuais de cada setor. Assim, é possível capturar a heterogeneidade nos níveis da PTF dos setores de forma mais completa do que a inclusão de apenas um efeito fixo individual permitiria⁷.

2.1 Estimações Adicionais

Para avaliar comparativamente a validade das hipóteses assumidas no AMG sobre o processo gerador de dados, a função de produção será estimada também por meio de outros estimadores que não assumem uma estrutura multifatorial heterogênea do erro. Serão considerados o estimador de MQO aumentado com *dummies* de tempo, efeitos fixos *two-way* (EF), que adiciona um efeito fixo individual e um efeito temporal comum na regressão, primeiras diferenças (PD), também aumentado com *dummies* de tempo, *system*-GMM de Blundell e Bond (1998) e *mean group* (MG) de Pesaran e Smith (1995).⁸

Estes estimadores assumem impacto homogêneo do fator comum entre os setores. Caso o processo verdadeiro apresente impactos heterogêneos dos fatores comuns, os estimadores não controlarão a dependência *cross-section* gerada pelos fatores mesmo com a inclusão da *dummy* de tempo, como explicado na seção 1. Além disso, MQO, EF e MG podem estar sujeitos ao problema de regressão espúria no caso de regressores não-estacionários. Espera-se que apenas o estimador AMG produza resíduos sem dependência *cross-section* e estacionários.

Eberhardt e Bond (2009) comparam, por meio de uma série de simulações de Monte Carlo, os estimadores de MQO, efeitos fixos, primeiras diferenças e AMG. Os resultados indicam que os dois primeiros mostram vies quando os dados apresentam dependência *cross-section*, heterogeneidade e não-estacionariedade. O estimador de primeiras diferenças, quando incluída *dummy* de tempo, não apresenta vies. O AMG é o que apresenta estimativas mais próximas dos parâmetros verdadeiros em geral, apesar do estimador de primeiras diferenças ter uma performance similar ao AMG. Finalmente, os autores apresentam evidência de que o AMG tem uma variância menor que os outros estimadores.

Para avaliar a presença de dependência *cross-section* nos resíduos de cada estimação, será aplicado o teste CD de Pesaran (2004), cuja hipótese nula é de ausência de dependência *cross-section*. O autor apresenta este teste como uma alternativa a outros testes de correlação *cross-section*, como Breusch e Pagan (1980), que, em geral, são mais apropriados para dados em que N é fixo e $T \rightarrow \infty$. Neste sentido, o teste CD é mais apropriado para casos em que N é grande relativamente a T , como é o caso dos dados utilizados neste trabalho. A estatística de teste é:

⁷Vale notar que a PTF calculada desta forma depende das elasticidades estimadas apenas para o cálculo no período inicial, uma vez que sua evolução nos períodos seguintes é determinada totalmente pelo componente temporal comum estimado pelo AMG. Ou seja, a inclusão das elasticidades estimadas e dos fatores de produção afeta o nível das produtividades estimadas, mas não suas trajetórias. Privilegiamos a qualidade dos resíduos estimados, de modo que todos os insumos foram utilizadas no cálculo da PTF inicial, mesmo quando não foram estatisticamente significantes.

⁸O MG consiste em estimar regressões individuais para cada setor por MQO e obter uma estimativa da elasticidade média dos insumos por meio da média das elasticidades de cada setor

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij}, \quad (14)$$

em que $\hat{\rho}_{ij}$ é o coeficiente de correlação entre os resíduos das observações.

Já para avaliar a não-estacionariedade dos resíduos, será aplicado o teste CIPS de Pesaran (2007), cuja hipótese nula é que a variável tem raiz unitária.⁹ O procedimento consiste em estimar regressões ADF para cada unidade do painel, aumentadas pela média *cross-section* da variável analisada como *proxies* dos fatores comuns latentes:

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + \bar{y}_{i,t-1} + d_i \Delta \bar{y}_t + e_{it}. \quad (15)$$

A estatística de teste é a média das estatísticas t individuais do parâmetro b_i . Assim como no teste ADF, a hipótese nula é que a variável tem raiz unitária. A especificação do teste CIPS é interessante no contexto deste trabalho por controlar a dependência *cross-section* por meio da inclusão de uma *proxy* para a combinação dos fatores comuns, o que vai ao encontro dos argumentos apresentados anteriormente em favor deste tipo de estratégia, como é o caso do AMG. Espera-se que os modelos corretamente especificados apresentem resíduos estacionários e sem dependência *cross-section*.

3 Base de Dados

Este trabalho utiliza os dados de grupos das indústrias extrativa e de transformação brasileiras entre os anos de 2003 e 2017. Foram considerados dados da Pesquisa Industrial Anual (PIA) do IBGE e da Relação Anual de Informações Sociais (RAIS). Da PIA, foram utilizadas as variáveis de Valor de Transformação Industrial (VTI) para a construção da variável de produto, pessoal ocupado (PO) para a variável de trabalho e as variáveis de aquisições, melhorias e baixas de ativos para a construção do estoque de Capital. Como as informações de investimento não estão disponíveis para firmas com menos de 5 pessoas ocupadas, consideramos os dados da PIA para empresas com 5 ou mais pessoas ocupadas, informação disponibilizada a um nível de agregação de 3 dígitos da Classificação Nacional de Atividade Econômica (CNAE). Foi adotado o conceito de firma representativa da mesma forma que Garcia (2003): os valores das variáveis de cada setor foram divididos pelo número de empresas por setor, de forma que todas as análises neste trabalho dizem respeito a uma firma média de cada setor.¹⁰ As variáveis nominais da PIA foram deflacionadas utilizando o Índice de Preços por Atacado (IPA) setorial da Fundação Getúlio Vargas (FGV).

Como discutido por Gusso (2006), a mão-de-obra técnico-científica – pesquisadores, engenheiros e gestores de P&D – constitui elemento fundamental para o crescimento do estoque de conhecimento das firmas, bem como para a aplicação deste conhecimento acumulado de forma produtiva. Neste sentido, a presença de profissionais científicos nas empresas torna possível que o conhecimento produzido por cooperações com outras

⁹O autor mostra que este teste é válido na presença de fatores comuns não-observados

¹⁰O uso da firma média pode produzir resultados sensíveis a *outliers*. Porém, os dados da PIA são divulgados em nível mais desagregado (4 dígitos da CNAE) apenas para as empresas com 30 ou mais funcionários, de modo que a firma média, ou mesmo mediana, para este nível de desagregação não seria muito representativa dos setores. Dessa forma, optou-se por trabalhar com a CNAE a 3 dígitos por entender-se que as empresas incluídas nessa amostra seriam mais representativas

empresas e universidades e o conhecimento produzido externamente possam ser absorvidos internamente pelas empresas. Corroborando este ponto, Araújo, Cavalcante e Alves (2009) mostram que os gastos em mão-de-obra especializada constituem a maior parte dos gastos em pesquisa na indústria brasileira. Desta forma, a variável de pessoal em ocupações técnico-científicas representa uma boa aproximação do esforço inovativo das empresas. Assim, utilizamos os dados de ocupação dos trabalhadores da RAIS, classificados pela Classificação Brasileira de Ocupações (CBO), para construir a variável de pessoal ocupado em pesquisa por setor, como *proxy* para o gasto em P&D. Para estes dados também foi considerado o conceito de firma média: o número de pessoas ocupadas em pesquisa foi dividido pelo total de empresas que entregou a declaração da RAIS em cada setor, agregado novamente a 3 dígitos da CNAE para compatibilização com a PIA.

Ao trabalhar com os dados da PIA, foram encontrados dois problemas. O primeiro diz respeito à classificação dos grupos industriais, que mudou a partir de 2007. Para poder utilizar os dados anteriores a 2007, foi construído um dicionário de compatibilização entre a versão antiga e a nova da Classificação Nacional de Atividades Econômicas (CNAE).¹¹ O segundo problema é a quantidade de *missing* nos dados. Caso a presença de *missing* esteja associada à produtividade dos setores é possível que isso gere viés na estimação. Por exemplo, se apenas setores mais produtivos apresentam informação completa, os resultados das estimações serão mais representativos deste grupo de setores do que da indústria como um todo. Uma análise da relação entre as características dos setores e a presença de *missing*, porém, indicou que não existe tal associação.¹² Assim, foram considerados apenas os setores com informação completa. Feitas estas considerações, a amostra final contém informações de 96 grupos industriais entre 2003 e 2017, totalizando 1396 observações.¹³

3.1 Construção das Variáveis

Garcia (2003) nota que, enquanto dados de trabalho no Brasil, em geral, são disponibilizados em termos de pessoas ocupadas, a medida mais usada para variáveis de trabalho na literatura internacional é o número de horas trabalhadas. Assim, como sugerido por este autor, o número de pessoas ocupadas foi multiplicado por uma jornada média anual de trabalho de 2112 horas/homem. Dessa forma, a variável representativa do fator trabalho (PO-Total) foi calculada como:

$$PO - Total_{it} = \frac{PO_{it}^{PIA} \times 2112}{N_{it}^{PIA}}, \quad (16)$$

em que PO_{it}^{PIA} e N_{it}^{PIA} são o pessoal ocupado e o número de empresas do setor i amostradas na PIA no período t , respectivamente.

Dados sobre inovação, incluindo gastos em P&D, são coletados no Brasil pela Pesquisa de Inovação Tecnológica (Pintec). Esta pesquisa, porém, é realizada a cada três anos. Assim, a utilização da variável de dispêndio em P&D da Pintec acarretaria em uma redução considerável da amostra disponível para este trabalho. Para superar tal limitação, Araújo, Cavalcante e Alves (2009) propõem a construção da variável de pessoal em ocupações técnico-científicas (Potec) a partir das informações da RAIS, o que permite construir uma medida anual para o investimento em P&D dos setores. Os autores

¹¹O dicionário de compatibilização da CNAE está disponível sob requisição

¹²Os resultados desta análise são apresentados em anexo.

¹³A tabela 6 em anexo apresenta uma descrição de todos os setores utilizados

argumentam em favor da validade dessa variável *proxy* para o gasto em P&D em razão da alta proporção dos gastos de pessoal na composição dos gastos em P&D, sendo, portanto, um indicativo do esforço inovativo das firmas. Avaliando a correlação entre a variável de Potec e dados da Pintec de gastos em inovação, Araújo, Cavalcante e Alves (2009) encontram evidências de que a variável captura de forma satisfatória os padrões de inovação dos setores.

A construção da variável de Potec consiste na soma do número médio de pesquisadores, engenheiros, diretores e gerentes de P&D e profissionais científicos por setor, como definido pela Classificação Brasileira de Ocupações (CBO).¹⁴ Finalmente, o número de pessoas incluídas no Potec por setor também foi multiplicado por uma jornada média anual de trabalho de 2112 horas/homem. Assim, a construção da variável pode ser descrita como:

$$Potec_{it} = \frac{R_{it} \times 2112}{N_{it}^{RAIS}}, \quad (17)$$

em que R_{it} é o número de pesquisadores, engenheiros, diretores e gerentes de P&D e profissionais científicos e N_{it}^{RAIS} é o número de empresas do setor i no período t que declararam a RAIS.¹⁵

Da mesma forma que foi feito por Araújo, Cavalcante e Alves (2009), foi realizada uma comparação entre a variável de Potec e os dados da Pintec sobre gastos em inovação para os anos em que a pesquisa está disponível. A tabela 12 em anexo aponta para uma alta correlação do Potec com gastos em P&D. No caso do gasto total, por exemplo, a correlação foi de 0.84 levando em conta apenas a indústria de transformação e 0.68 para a amostra completa.¹⁶ Além disso, a figura 1 indica que o Potec adere bem à divisão dos setores com base na intensidade tecnológica, proposta pela OCDE e adaptada para a CNAE por Cavalcante (2014).¹⁷ A média do Potec em logaritmo foi de 8.64 para os setores de alta intensidade, 8.15 para os de média-alta, 7.28 para os de média-baixa e 6.6 para os de baixa, sendo que todas essas médias apresentaram diferenças estatisticamente significantes entre si. Estes resultados apontam para a qualidade do Potec como *proxy* do gasto em P&D.

A fim de evitar dupla contagem nas séries para o fator trabalho ($PO - Total_{it}$), a variável utilizada nos resultados principais desta pesquisa foi construída subtraindo o Potec das séries de pessoal total. Assim, a variável utilizada para o fator trabalho representa pessoal ocupado que não desempenha ocupações técnico-científicas, denotada por $NPotec$.

Para a série de capital, foi utilizado o método do inventário perpétuo, em que o estoque de capital em dado período é a soma do investimento corrente com o estoque de capital do período anterior, descontada uma taxa de depreciação. O investimento foi calculado como o somatório da média de aquisições e melhorias de ativos, subtraído da média das baixas. A taxa de depreciação utilizada, seguindo Garcia (2003), foi de 5%. Como sugerido por este autor, o cálculo do estoque inicial de capital foi feito da seguinte forma: com base no coeficiente do trabalho na distribuição funcional da renda,

¹⁴Uma descrição das categorias contempladas pelo Potec, identificadas pela CBO, é apresentada em anexo

¹⁵É importante ressaltar que este cálculo não considera funcionários que trabalham em tempo parcial nas atividades de P&D, o que pode levar a um erro de aproximação

¹⁶Tal diferenciação foi considerada porque os dados da indústria extrativa só são divulgados de forma agregada na Pintec, sem distinção de setor

¹⁷Alta (A), Média-Alta (MA), Média-Baixa (MB), Baixa (B)

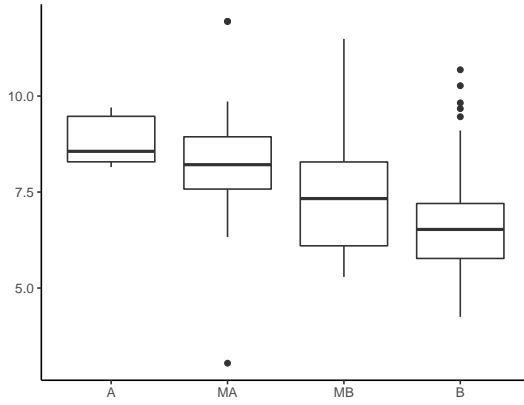


Figura 1: Potec por Intensidade Tecnológica da OCDE

$\alpha_i^l = W_{i0}/Y_{i0}$, em que W_{i0} é o total de salários do setor i no período inicial – no caso deste trabalho, o ano de 1996¹⁸ –, e no coeficiente do capital na distribuição funcional da renda, $\alpha_i^k = 1 - \alpha_i^l$, o estoque inicial de capital é calculado como:

$$K_{i0} = \frac{(Y_{i0})^{\frac{1}{\alpha_i^k}}}{(W_{i0})^{\frac{\alpha_i^l}{\alpha_i^k}}}. \quad (18)$$

Por fim, para a série de produto, foi utilizada a variável de Valor de Transformação Industrial (VTI) da PIA. Esta variável é calculada pela diferença entre o valor bruto da produção industrial e o custo das operações industriais. Assim, o VTI pode ser visto como uma medida de valor adicionado.

3.2 Análise Descritiva

A fim de avaliar a pertinência das hipóteses feitas neste trabalho sobre a presença de dependência *cross-section* e raiz unitária nas variáveis, esta seção apresenta uma análise descritiva dos dados utilizados. A tabela 1 contém estatísticas descritivas das variáveis utilizadas. A decomposição da variância entre seu componente temporal (*within*) e setorial (*between*) mostra que a maior parte da variação nos dados vem de diferenças de nível entre os setores, enquanto se observa pouca variação em torno de suas tendências.¹⁹ Isso sugere uma heterogeneidade nas funções de produção dos setores.

A figura 2 mostra a comparação entre setores – agregados nos dois primeiros dígitos da CNAE para facilitar a visualização – do logaritmo da razão do produto, capital e Potec por hora trabalhada (Y/L , K/L e R/L , respectivamente). É possível notar que os setores com maior produtividade do trabalho não são necessariamente aqueles com maior capital ou P&D por trabalhador. Por exemplo, o setor 07 (Extração de Minerais Metálicos) apresenta níveis maiores de capital e Potec por trabalhador que o setor 19 (Fabricação de Coque, de Produtos Derivados do Petróleo e de Biocombustíveis). Neste caso, sob homogeneidade das elasticidades e da PTF, espera-se que o setor 07 também apresentasse

¹⁸Assim, como as estimações utilizam os dados a partir de 2003, o efeito do cálculo do estoque inicial de capital é pequeno

¹⁹Componente *within* calculado como $x_{it} - \bar{x}_i$ e componente *between* calculado como $\bar{x}_i - \bar{x}$, em que \bar{x}_i é a média da variável x por setor e \bar{x} é sua média global

Tabela 1: Estatísticas Descritivas

	Média	Var (Within)	Var (Between)
VTI (R\$)	47.528.566	20%	80%
NPotec (Horas)	380.592.416	21%	79%
Capital (R\$)	172.838.728	7%	93%
Potec (Horas)	7.164.389	22%	78%

Fonte: Elaboração própria

maior produto por trabalhador, o que não acontece.²⁰ Isto pode ser visto como indício da heterogeneidade das elasticidades ou da PTF.²¹

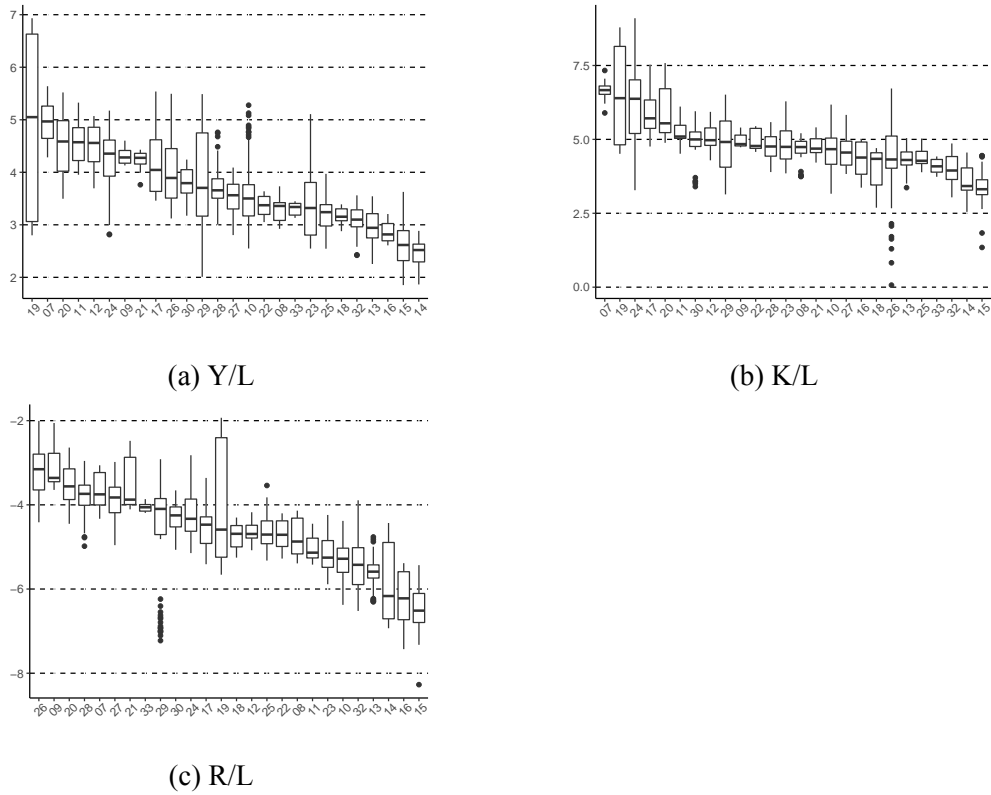


Figura 2: Produto e Insumos por Hora de Trabalho

A figura 3 apresenta a evolução das variáveis por setor. A fim de facilitar a visualização, os setores foram agrupados por proximidade com base em sua divisão da CNAE. Por exemplo, setores 13 a 15 da CNAE representam a fabricação de diversos produtos têxteis, setores 17 a 18 representam a fabricação de papel e derivados e impressão *etc.*²² As médias anuais de cada grupo são reportadas. Além disso, as séries são apresentadas como proporção do valor máximo da série, o que evita distorções na escala do gráfico.²³

²⁰A mesma comparação pode ser feita entre os setores 14 e 15 ou 23 e 10 por exemplo

²¹Para ilustrar este ponto, considere por simplicidade a função de produção com capital, trabalho e P&D e retornos constantes de escala. Neste caso, tem-se $y_{it} = a + \alpha k_{it} + \gamma r_{it}$, em que $y_{it} = \log(Y/L)_{it}$, $k_{it} = \log(K/L)_{it}$, $r_{it} = \log(R/L)_{it}$ e $a = \log(A)$ é o logaritmo da PTF. Assim, setores com maiores valores de k_{it} e r_{it} deveriam ter valores maiores também de y_{it}

²²Grupos construídos com base em Eberhardt, Helmers e Strauss (2013)

²³Para cada setor, tem-se $y_t^* = y_t / \max(y_t)$; são reportadas as médias de y_t^* entre os grupos de setores

A linha preta tracejada no gráfico representa uma aproximação local desta tendência enquanto a faixa cinza é o intervalo de confiança, construído a um nível de significância de 95%.

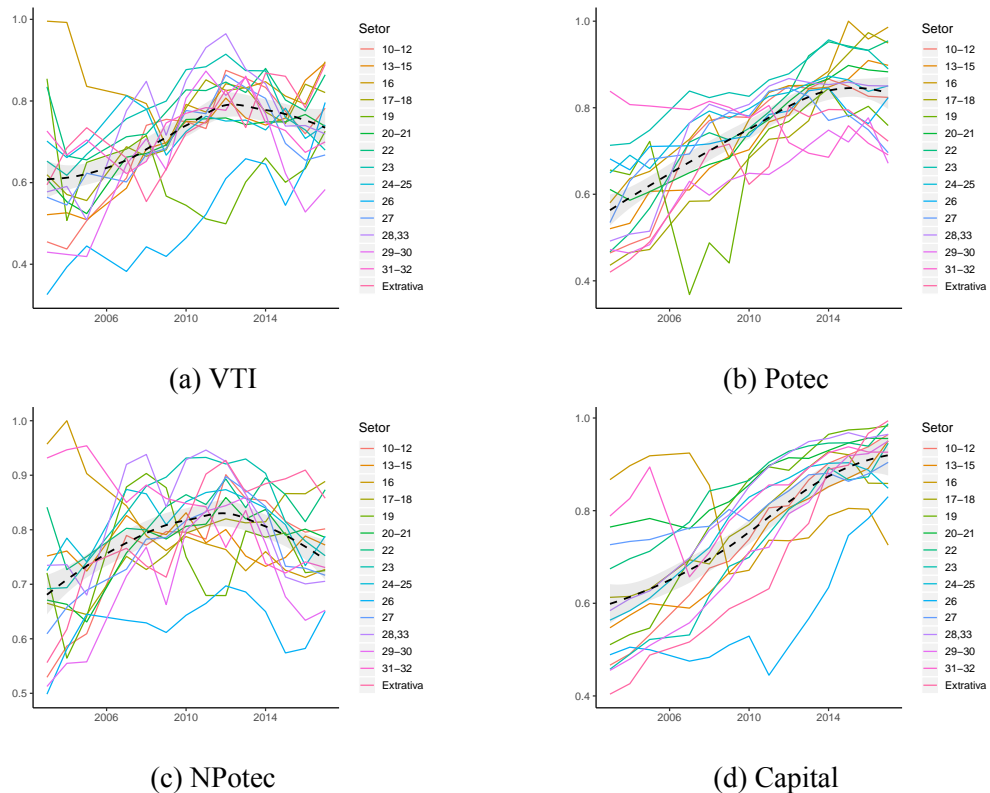


Figura 3: Evolução Média das Variáveis por Setor

Analisando a figura 3a, é possível observar uma tendência temporal comum às séries do produto de crescimento até 2012, e relativa estagnação desde então. Os únicos setores que desviam desse padrão são 16 (Fabricação de Produtos de Madeira), 19 (Fabricação de Coque, de Produtos Derivados do Petróleo e de Biocombustíveis) e 26 (Fabricação de Equipamentos de Informática, Produtos Eletrônicos e Ópticos). A figura 3b indica também um padrão temporal comum para o P&D, porém diferente da tendência observada nas séries de produto. As séries de P&D apresentam tendência de crescimento constante até 2014, com certa desaceleração no crescimento a partir deste ano, além de variarem menos em torno da evolução média. Novamente, o setor 19 desvia em alguns anos dessa tendência. A divergência nas tendências temporais comuns do produto e do P&D indica haver pouca correlação entre o comportamento das duas séries. Assim, é esperado que o esforço inovativo dos setores tenha pouca influência sobre sua produção. Além disso, a presença de uma evolução comum entre os setores para cada variável pode ser vista como indício da influência de fatores comuns não-observados nas séries individuais para cada variável.

As figuras 3c e 3d apresentam a evolução média das séries de trabalho e capital respectivamente. Conforme esperado, o estoque de trabalho responde mais à variação cíclica do produto, enquanto que o estoque de capital apresenta evolução mais constante, com tendência à desaceleração ao final da amostra, demonstrando a redução de investimentos com a fase recessiva do ciclo econômico.

Por fim, foram aplicados os testes CIPS e CD, descritos na seção 2, sobre as variáveis

construídas. A tabela 2 apresenta os resultados dos testes. Na primeira coluna, estão as estatísticas do teste CD. A segunda coluna traz os resultados do teste CIPS sem tendência enquanto a terceira coluna traz os resultados do teste incluindo tendência. Os resultados do teste CIPS com 0 e 1 defasagem são basicamente os mesmos. Assim, são reportados apenas os resultados do teste com 1 defasagem.

Tabela 2: Testes Preliminares

	CD	CIPS	
		Sem Tendência	Com Tendência
VTI	38,13***	-1,86	-2,6*
Capital	89,21***	-1,92	-1,77
Potec	63,09***	-1,4	-1,67
NPotec	27,5***	-1,71	-2,4

Nota: *p<0,1; **p<0,05; ***p<0,01.

Fonte: Elaboração própria

A hipótese de independência *cross-section* é rejeitada para todas as variáveis. Já a hipótese de raiz unitária não é rejeitada em todos os casos exceto para o VTI quando é incluída tendência no teste. Os testes indicam que não-estacionariedade e dependência *cross-section* são características presentes nos dados. Assim, justifica-se a estimação da função de produção com um modelo que leve em conta tais características, como o proposto na seção 1.

4 Resultados

4.1 Estimação Sem P&D

Conforme mencionado na seção 2, além do estimador AMG, a função de produção foi estimada com os estimadores de MQO, Efeitos Fixos, primeiras diferenças, *system-GMM* e MG. Serão reportados primeiramente os resultados das estimações da função de produção convencional, ou seja, com o pessoal total e sem o Potec. Em outras palavras, assume-se nesta primeira estimação que a elasticidade do trabalho é a mesma para pessoal ocupado em pesquisa (Potec) e em outras atividades (NPotec). Assim, é possível avaliar a importância de diferenciar os dois tipos de mão-de-obra por meio da qualidade desta especificação. A tabela 3 apresenta os resultados desta primeira estimação.²⁴ Como forma complementar de avaliação, são reportados também a estatística de teste de retornos constantes de escala (CRS)²⁵, a raiz do erro quadrático médio (RMSE) e o critério de informação de Akaike (AIC) para cada estimador.

As estimativas da elasticidade do trabalho apresentaram uma média de 1,043. Em geral, os valores obtidos foram próximos entre os estimadores. As exceções foram MQO, em torno de 1,6 desvios-padrão abaixo da média, e *system-GMM*, cerca de 1 desvio-padrão acima. Os valores da elasticidade do capital variaram entre 0,048 e 0,436, com média de 0,197. Novamente, a estimativa de MQO foi a que mais se distanciou das

²⁴Para o *system-GMM*, são apresentados os coeficientes de longo-prazo. Além disso, o teste Comfac, proposto por Blundell e Bond (2000), indica que a representação assumida pelo *system-GMM* é válida. Assim, são reportados os coeficientes restritos

²⁵Teste F da hipótese $\beta_K + \beta_L = 1$, em que β_L é o coeficiente do trabalho e β_K , o do capital

Tabela 3: Estimação com Pessoal Total

	MQO	EF	PD	Sys-GMM	MG	AMG
Capital	0.436 (5.42)***	0.16 (2.73)***	0.048 (2.07)**	0.188 (2.49)**	0.172 (2.44)**	0.179 (2.72)***
Pessoal	0.834 (6.38)***	0.983 (19.29)***	1.048 (29.28)***	1.198 (8.95)***	1.1 (26.37)***	1.095 (24.96)***
Obs	1396	1396	1300	1300	1396	1311
CRS	18.23***	11.81***	7.67***	5.19**	11.05***	12.06***
CD	-1.73*	-2.29**	-1.78*	-1.72*	5.35***	-2.14**
CIPS	-2.04	-1.87	-2.98***	-3.35***	-3.09**	-3.18***
RMSE	0.46	0.19	0.14	0.15	0.1	0.08
AIC	1848.91	-712.66	-1227.19	-1166.21	-2312.55	-2710.11

Fonte: Elaboração própria

outras, estando cerca de 1,87 desvios-padrão acima da média. Em menor escala, o coeficiente de primeiras diferenças também se distanciou do restante, estando cerca de 1,16 desvios-padrão abaixo da média.

Analisando os coeficientes obtidos, o MQO parece produzir o resultado de maior divergência, com uma elasticidade do capital muito acima da média das estimativas e a do trabalho abaixo da média. Portanto, ao se controlar para efeitos fixos na estimação é verificada uma tendência para a redução do coeficiente associado à variável de capital e aumento no coeficiente associado à variável de trabalho.

Já *system*-GMM, MG e AMG produzem estimativas bem próximas da média para a elasticidade do capital. Quando são avaliados comparativamente os estimadores MG e AMG percebe-se que os coeficientes estimados são todos próximos entre si. Ou seja, a estimação com efeitos comuns homogêneos ou heterogêneos, dada a hipótese de heterogeneidade das elasticidades, não afeta muito o resultado. No entanto, com relação às estimativas produzidas pelo estimador Sys-GMM, os coeficientes estimados pelo AMG são um pouco inferiores para as duas variáveis explicativas. Finalmente, os resultados de todos os estimadores indicam a presença de retornos crescentes de escala. Com relação aos testes sobre os resíduos, apenas MQO e EF não produzem resíduos estacionários, indicando que os resultados destes estimadores possivelmente são espúrios. Além disso, para essa especificação sem a variável de P&D, nenhum dos estimadores foi capaz de eliminar a dependência *cross-section* dos resíduos.

4.2 Estimação com P&D

Tendo em vista o comportamento dos estimadores sem a inclusão da variável de P&D, procede-se agora com a estimação da elasticidade de P&D. A tabela 4 apresenta os resultados das estimações da função de produção incluindo a variável Potec e substituindo a variável de Pessoal por NPotec.

Após a inclusão do Potec, observa-se que o AIC de todos os estimadores caiu, indicando que esta variável aumenta o poder explicativo de todos os modelos, mesmo nos casos em que seu coeficiente não foi estatisticamente significativo. Isto aponta para a importância de incluir a variável de P&D na estimativa da função de produção setorial.

Com relação aos coeficientes estimados, a média das estimativas da elasticidade do trabalho caiu para 0,94. Novamente, os valores obtidos pelos estimadores foram próximos, com exceção do MQO, cuja estimativa está cerca de 1,76 desvios-padrão abaixo da média. Vale notar a maior variabilidade nos coeficientes do trabalho entre os estimadores

Tabela 4: Estimação com Potec

	MQO	EF	PD	Sys-GMM	MG	AMG
Capital	0.337 (5.56)***	0.16 (2.66)***	0.047 (2.02)**	0.159 (1.94)*	0.117 (1.50)	0.162 (2.12)**
NPotec	0.646 (6.73)***	0.967 (16.87)***	1.022 (26.46)***	0.856 (6.44)***	1.089 (22.09)***	1.069 (22.20)***
Potec	0.267 (6.25)***	0.015 (0.32)	0.035 (1.35)	0.255 (4.7)***	0.029 (0.75)	0.014 (0.40)
Obs	1396	1396	1300	1300	1396	1311
CRS	25.28***	12.6***	7.97***	2.25	5.55**	6.42**
CD	-2.08**	-2.29**	-1.76*	-2.25**	6.11***	-1.52
CIPS	-2.03	-1.74	-3.13***	-3.35***	-2.92**	-2.93**
RMSE	0.4	0.19	0.14	0.14	0.09	0.08
AIC	1441.56	-716.70	-1228.04	-1201.42	-2580.60	-2855.58

Fonte: Elaboração própria

após a inclusão do Potec. As estimativas da elasticidade do capital também caíram. Os novos valores variaram entre 0,047 e 0,337, com média de 0,163. A estimativa de MQO continua sendo a que mais se distancia das outras, estando cerca de 1,8 desvios-padrão acima da média.

Quando são comparados os dois estimadores que admitem heterogeneidade dos coeficientes por grupo setorial (MG e AMG), percebe-se que a maior diferença nas estimativas se deve ao coeficiente associado à variável de Capital produzida pelo estimador MG, que é menor e deixa de ser estatisticamente significativa. Ou seja, essa diferença deve-se basicamente ao controle por efeitos comuns homogêneos ou heterogêneos.

Vale notar que, com a nova especificação, o modelo estimado por *system*-GMM passa a apresentar uma diferença maior em relação ao AMG. A estimativa produzida pelo modelo Sys-GMM apresenta um coeficiente para NPotec igual a 0,856, em oposição a 1,069 para o modelo AMG, e 0,255 para o coeficiente associado a Potec, em oposição a 0,014 para o modelo AMG. Ou seja, a estimativa produzida pelo modelo Sys-GMM indica uma alta elasticidade para o P&D, sendo próxima somente à produzida pelo estimador MQO.

O restante das estimativas para a elasticidade de P&D encontram-se abaixo dos valores reportados por Cavalcante, Jacinto e Negri (2015), entre 0,06 e 0,1, e Zucoloto et al. (2017) com estimativa consistente de 0,06. Para os estimadores EF, PD, MG e AMG não foram obtidas estimativas estatisticamente significantes em um nível de significância de 5%, sendo os menores coeficientes obtidos pelos estimadores EF (0,015) e AMG (0,014).

Com exceção do AMG, os resultados sobre as propriedades dos resíduos são os mesmos. Os resíduos do AMG, além de estacionários como antes, não apresentaram dependência *cross-section*. Comparando com a estimativa produzida pelo mesmo estimador sem a inclusão da variável de P&D é possível afirmar que a correção desse problema se deveu à modificação na especificação do modelo. Por fim, o AMG foi o estimador que apresentou menor RMSE e AIC entre todos os modelos, nas regressões com e sem o Potec, indicando melhor ajuste aos dados e maior poder explicativo. A análise dos resíduos corrobora o argumento feito neste trabalho de que o AMG é o estimador mais adequado estatisticamente dentre os analisados para estimar a função de produção.

O conjunto de resultados aqui apresentados indicam que estimativas de elasticidade estatisticamente significante do P&D podem ser consequência de problemas de espe-

cificação na função de produção. Questões relacionadas à dependência *cross-section* e não-estacionariedade das variáveis em painel devem ser consideradas. Resultado semelhante foi obtido por Eberhardt, Helmers e Strauss (2013), que estimam a elasticidade do P&D com o estimador CCE. Os autores argumentam que esse resultado evidencia a dificuldade de separar os retornos individuais de P&D do efeito de transbordamentos de conhecimento. Eles notam também que a melhor performance do modelo de fatores comuns latentes indica que o conhecimento externo pode não ser a única fonte de endogeneidade nos dados, uma vez que, nesta abordagem, não se fazem hipóteses sobre o que está presente nestes fatores.²⁶

Os resultados acima também sugerem que políticas de inovação estritamente setoriais podem não ser tão eficientes quanto se espera. Particularmente, o fato de que boa parte dos retornos de P&D vem de um componente comum entre setores indica que o transbordamento é um fator importante para explicar a relação entre pesquisa e produtividade. Assim, políticas que busquem facilitar a disseminação do conhecimento pela indústria e a interação entre setores devem produzir melhores resultados.

4.3 Análise da PTF

Eberhardt e Teal (2010) propõem o cálculo da PTF utilizando as estimativas fornecidas pelo estimador AMG conforme expresso pelas equações (12) e (13). Este procedimento foi realizado para obter a evolução da PTF de cada grupo industrial na amostra.²⁷ Como a PTF depende das elasticidades e dos coeficientes associados ao componente dinâmico comum estimados pelo AMG, é possível avaliar a qualidade dessas estimativas a partir da análise da PTF calculada.

A figura 4 apresenta a evolução da PTF por grupos de setores, da mesma forma que a figura 3. A comparação com as séries de produto e P&D mostra que a evolução da PTF tem um comportamento mais próximo à média da evolução do produto que à da evolução dos gastos em P&D. Ou seja, o declínio da produtividade tende a ocorrer após 2013 para a maior parte dos grupos industriais. Por outro lado, as séries da PTF seguem uma evolução mais suavizada que a do produto, apresentando um crescimento menos pronunciado e menor variação em torno da média. Esse comportamento da PTF estimada sugere que os componentes de transbordamentos de conhecimento e outros choques comuns são mais relevantes do que o gasto em P&D privado de cada grupo industrial para explicar a evolução na produtividade dos grupos industriais.

A tabela 5 apresenta a média e variância da PTF, além do componente temporal (*within*) e setorial (*between*) da variância. É possível notar que a maior parte da variância da PTF – cerca de 95% – vem da variância entre setores e não do componente temporal. Isso confirma o padrão observado na figura 4 de que as séries dos setores variam pouco em torno da tendência comum, mas que há bastante variação de nível entre os setores. Além disso, este padrão de composição da variância se observa para todas as variáveis. Isto é, para o VTI e os fatores de produção a maior parte da variância também vem da variância entre setores e não da variação temporal. Essa análise evidencia

²⁶A mesma comparação entre os estimadores, com e sem Potec, foi feita considerando apenas dados a partir de 2007 para garantir que os resultados reportados não refletem problemas na compatibilização da CNAE 1.0 (2003 a 2007) com a CNAE 2.0. Novamente, estimadores que produzem uma estimativa estatisticamente significativa para o Potec não controlam a dependência *cross-section*, sendo a única exceção o PD. Os resultados estão disponíveis sob requisição

²⁷Foram considerados os resultados produzidos pelo AMG com a inclusão do Potec, uma vez que este foi o melhor modelo tanto pelo critério do RMSE quanto pelo AIC

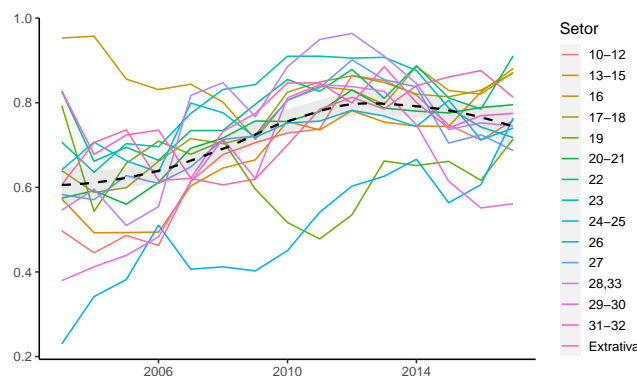


Figura 4: Evolução média da PTF por Setor

a heterogeneidade da PTF entre os setores, como discutido ao longo deste trabalho.

Tabela 5: Estatísticas Descritivas da PTF

Média	Variância	Var (Within)	Var (Between)
8.71	3.15	0.15	3

Partindo dos modelos mais tradicionais da literatura de inovação e produtividade, Nelson (1988) mostra que é possível que exista uma correlação entre P&D e PTF, mesmo que não haja um impacto direto da pesquisa na produção, como os resultados da seção anterior sugerem. Segundo o autor, a explicação para isso se baseia na ideia de oportunidade tecnológica. Em setores cuja produção depende de tecnologias de ponta, as firmas precisam investir mais em pesquisa para reposição do estoque de conhecimento, que se deprecia muito rápido. Por outro lado, o acesso a tecnologias melhores e constantemente atualizadas faz com que esses setores possam ser mais produtivos do que outros que dependam menos da tecnologia. Assim, é de se esperar que setores que investem mais em pesquisa tenham maior produtividade, mesmo que não se observe uma causalidade direta entre estas variáveis. Corroborando esta previsão, Hall, Mairesse e Mohnen (2010) notam que é comum na literatura encontrar uma correlação *cross-section* positiva entre intensidade tecnológica e a produtividade.

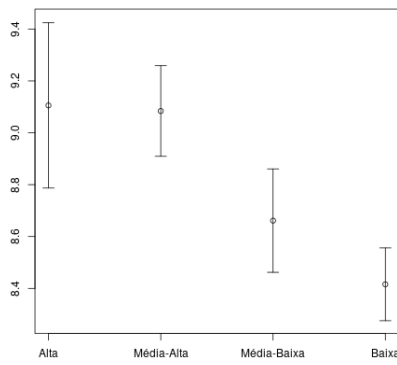
Porém, Nelson (1988) nota que diferenças setoriais no grau de apropriabilidade do conhecimento podem tornar essa correlação menos clara. Isso porque firmas inovadoras podem investir mais do que o ideal em pesquisa para se proteger de firmas imitadoras, enquanto as firmas imitadoras podem obter ganhos de produtividade por meio do conhecimento assimilado, mesmo que invistam pouco em pesquisa. Esses fatores podem fazer com que a correlação entre pesquisa e produtividade, gerada pela oportunidade tecnológica, seja menos perceptível nos dados.

Com base nestas considerações, as estimativas da PTF foram agregadas em setores segundo sua intensidade tecnológica, sendo investigada a relação entre produtividade e intensidade tecnológica. Para isso foram estimadas regressões da PTF – em nível e taxa de crescimento respectivamente – contra variáveis de *dummies* de intensidade tecnológica:

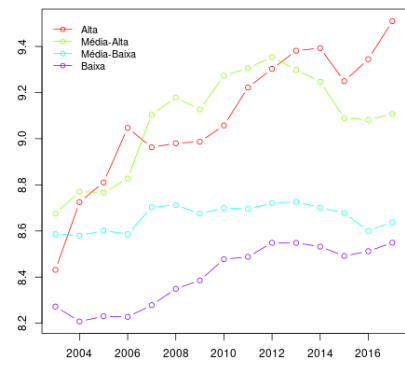
$$PTF_{it} = \delta_A \times IT_{Alta} + \delta_{MB} \times IT_{Média-Alta} + \delta_{MB} \times IT_{Média-Baixa} + \delta_B \times IT_{Baixa} + \varepsilon_{it} \quad (19)$$

$$\Delta PTF_{it} = \gamma_A \times IT_{Alta} + \gamma_{MB} \times IT_{Média-Alta} + \gamma_{MB} \times IT_{Média-Baixa} + \gamma_B \times IT_{Baixa} + \varepsilon_{it}. \quad (20)$$

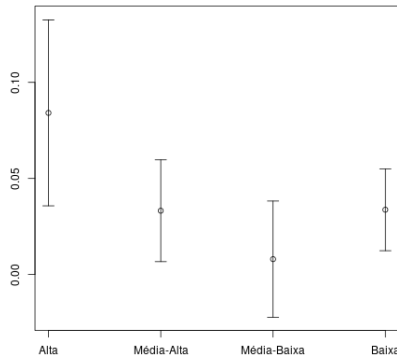
em que o coeficiente δ_j captura a média de PTF_{it} para setores do grupo j e de modo similar o coeficiente γ_j captura a média de ΔPTF_{it} para setores do grupo j . A figura 5 apresenta os resultados desta análise.²⁸



(a) PTF Média por Setor



(b) Evolução da PTF por Setor



(c) Crescimento Médio da PTF por Setor



(d) Variação Anual da PTF por Setor

Figura 5: Evolução da PTF por Grau de Intensidade Tecnológica

As figuras 5a e 5b apresentam respectivamente, os coeficientes estimados e seus intervalos de confiança de 95%, e as trajetórias das PTFs em nível por intensidade tecnológica. Inicialmente é possível notar que não há diferença estatisticamente significativa entre os setores de alta e média-alta intensidade tecnológica. Essa observação é confirmada pela semelhança na evolução da PTF para essas duas categorias de setores que, no entanto, apresentam os maiores níveis de produtividade da indústria.

²⁸Os coeficientes das regressões e os respectivos erros-padrão são apresentados em anexo

Por outro lado, nota-se uma diferença significativa entre os níveis de produtividade dos setores de média-baixa e baixa intensidade tecnológica, conforme pode ser visto nas figuras 5a e 5b, apesar de ainda haver uma pequena interseção entre os intervalos de confiança. Também nota-se que a evolução das PTFs desses dois grupos de setores foi distinta entre 2003 e 2017. Enquanto os setores de média-baixa apresentaram um comportamento de estagnação para a produtividade, os setores de baixa intensidade tecnológica apresentaram um aumento do nível, principalmente, entre os anos de 2007 e 2012.

Quanto ao crescimento da PTF, a figura 5c demonstra uma grande dispersão nas taxas de variação para os setores de alta intensidade tecnológica, ainda que com um nível médio mais elevado que os setores de média-alta intensidade. A figura indica também que os setores de alta intensidade tecnológica apresentam maior crescimento que os demais (0,08 % a.a.) , mas que entre os setores de média-alta e baixa intensidade tecnológica não há diferença significativa no crescimento (0,03 % a.a).

A observação das trajetórias de variação na figura 5d permite notar que as taxas de crescimento da PTF diferem entre os setores classificados segundo a intensidade tecnológica principalmente em momentos pontuais. Em primeiro lugar, observa-se que os setores de alta intensidade tecnológica apresentaram uma variação da PTF menos sensível ao ciclo econômico presente no início da amostra, enquanto foram extremamente sensíveis ao ciclo recessivo iniciado em 2014, com uma recuperação forte a partir de então. Já os demais setores industriais apresentaram um forte crescimento associado ao pico de crescimento econômico no ano de 2007. Por outro lado, os setores de média-baixa e baixa intensidade tecnológica não tiveram uma variação pronunciada da produtividade durante o ciclo recessivo de 2014. Somente os setores de média-alta intensidade tecnológica acompanharam o declínio observado para os setores de alta intensidade tecnológica.

De um modo geral, os resultados apresentados na figura 5 sugerem uma correlação positiva entre produtividade e intensidade tecnológica. No entanto, a diferenciação por meio de intensidades tecnológicas permite observar que a pequena diferença entre os níveis de investimento em P&D entre os setores de alta e média alta intensidade tecnológica se reflete em níveis ou taxas de crescimento de produtividade não muito diferentes. A maior percepção de diferença de comportamento da PTF vem dos setores de média-baixa intensidade tecnológica, que explicam grande parte do comportamento estagnado da produtividade da indústria no período.

Levando em conta o argumento de Nelson (1988), esse conjunto de evidências também sugere que a apropriabilidade – e, conseqüentemente, o transbordamento – do conhecimento é uma característica presente na indústria brasileira. Adicionalmente, outros fatores responsáveis por externalidades estão bastante presentes na composição dos efeitos comuns correlacionados que são determinantes na explicação da evolução da PTF na indústria brasileira. Isto corrobora a ideia de Nelson (1988) de que a correlação positiva observada entre pesquisa e produtividade não decorre de uma relação causal entre estas variáveis, mas sim do fato de que tanto o investimento ótimo em P&D quanto a produtividade são afetados por fatores comuns.

Conclusão

Este trabalho contribui à literatura econômica brasileira com a estimação da elasticidade produto para o P&D no Brasil considerando um modelo de fatores comuns não observados para a função de produção. Para isso, foi aplicado o estimador AMG de

Eberhardt e Bond (2009). Essas estimativas também permitiram o cálculo da PTF para os grupos industriais seguindo a proposição de Eberhardt e Teal (2010) para sua estimação.

Ainda que os estimadores para efeitos comuns correlacionados, com dados agregados, guardem relação com estimadores para função de produção que controlam para choques não observados de produtividade, em nível de microdados, não é possível assegurar plenamente que não exista viés nas estimativas para os parâmetros da função de produção. Ainda assim, duas observações ficaram claras nesta pesquisa, cujos dados estão em nível de grupo industrial. A primeira é que a especificação da função de produção desconsiderando a variável de P&D na função de produção implica em erro de especificação do modelo. A segunda observação foi de que a desconsideração de efeitos comuns correlacionados também pode produzir inconsistência nas estimativas dos coeficientes da função de produção. Considerando as estimativas dos coeficientes da função de produção e a qualidade dos resíduos, observou-se que a combinação de heterogeneidade dos parâmetros com a presença de efeitos comuns correlacionados foi a especificação que produziu os resultados estatisticamente mais adequados, a despeito de se lidar com dados agregados por grupo industrial e se enfrentar diversos problemas de compatibilização de dados entre diferentes amostras e pesquisas e a dimensão temporal relativamente restrita da amostra.

A elasticidade de P&D produzida pelo estimador AMG, de 0,014, não foi estatisticamente significativa. Comparado aos demais estimadores utilizados na pesquisa, esse foi o menor coeficiente obtido, embora alguns estimadores também tenham apresentado estimativas não significantes estatisticamente. Este é um indicativo de que a presença de fatores comuns na estimação pode induzir um viés positivo na estimativa da elasticidade. Isso aponta para a dificuldade de se separar o efeito individual do P&D do impacto de transbordamentos de conhecimento e demais efeitos comuns correlacionados. Nesse sentido, este trabalho corrobora a evidência fornecida por Eberhardt, Helmers e Strauss (2013) sobre a relação entre retornos de P&D e transbordamento de conhecimento.

A PTF dos grupos industriais foi calculada a partir das estimativas da função de produção obtidas pelo estimador AMG. Foi possível observar uma clara correlação positiva da intensidade tecnológica dos setores com o nível médio da PTF, mas não tanto com seu crescimento. A informação em nível sugere a importância do componente de transbordamento de conhecimento. Porém, além desse componente, a PTF também apresenta comportamento pró-cíclico. Esse comportamento não é homogêneo entre todas as categorias de setores segundo intensidades tecnológicas, indicando que os setores de mais alta intensidade tendem a responder mais fortemente às variações cíclicas do produto.

Em termos de formulação de políticas públicas, as evidências produzidas são úteis para se discutir a efetividade de políticas industriais estritamente setoriais. Os resultados aqui apresentados destacam a importância de externalidades que extrapolam os impactos esperados em nível setorial de políticas de estímulos a gastos em P&D. Isto tem implicação direta sobre a mensuração da efetividade das políticas setoriais de inovação tecnológica e a importância de políticas horizontais que estimulem a atividade produtiva e os incrementos de produtividade. Uma segunda observação é a de que é usual observar um componente pró-cíclico nos gastos em inovação por meio das informações públicas.²⁹ Foram observadas evidências de heterogeneidade no comportamento ao longo do ciclo econômico da produtividade dos grupos setoriais segundo sua intensidade tecnoló-

²⁹Ver PINTEC-IBGE em suas várias edições (<https://www.ibge.gov.br/estatisticas/multidominio/ciencia-tecnologia-e-inovacao/9141-pesquisa-de-inovacao.html>)

gica. Essas diferenças auxiliam na compreensão da relação entre gastos em inovação e a evolução da produtividade dos setores, inclusive, na sua relação com o ciclo econômico. Levar em conta tais diferenças permite melhor compreender a dinâmica da produtividade da indústria em nível intersetorial, o que, por sua vez, auxilia no desenho de políticas de crescimento econômico mais efetivas.

Referências

- AGHION, P.; HOWITT, P. *A model of growth through creative destruction*. [S.l.]: National Bureau of Economic Research Cambridge, Mass., USA, 1990.
- ALVES, P.; SILVA, A. M. Estimativa do estoque de capital das empresas industriais brasileiras. Instituto de Pesquisa Econômica Aplicada (Ipea), 2008.
- ARAÚJO, B. C.; CAVALCANTE, L. R.; ALVES, P. Variáveis proxy para os gastos empresariais em inovação com base no pessoal ocupado técnico-científico disponível na relação anual de informações sociais (rais). Instituto de Pesquisa Econômica Aplicada (Ipea), 2009.
- ARROW, K. J. Economic welfare and the allocation of resources for invention. In: *Readings in industrial economics*. [S.l.]: Springer, 1972. p. 219–236.
- BAI, J. Panel data models with interactive fixed effects. *Econometrica*, Wiley Online Library, v. 77, n. 4, p. 1229–1279, 2009.
- BEVEREN, I. V. Total factor productivity estimation: A practical review. *Journal of economic surveys*, Wiley Online Library, v. 26, n. 1, p. 98–128, 2012.
- BLUNDELL, R.; BOND, S. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, Elsevier, v. 87, n. 1, p. 115–143, 1998.
- BLUNDELL, R.; BOND, S. Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews*, Taylor & Francis, v. 19, n. 3, p. 321–340, 2000.
- BREUSCH, T. S.; PAGAN, A. R. The lagrange multiplier test and its applications to model specification in econometrics. *The review of economic studies*, JSTOR, v. 47, n. 1, p. 239–253, 1980.
- CAVALCANTE, L. R. Classificações tecnológicas: uma sistematização. Instituto de Pesquisa Econômica Aplicada (Ipea), 2014.
- CAVALCANTE, L. R.; JACINTO, P. d. A.; NEGRI, F. D. P&d, inovação e produtividade na indústria brasileira. *PRODUTIVIDADE NO BRASIL*, p. 43, 2015.
- CHRISTIANO, L. J.; EICHENBAUM, M. Current real-business-cycle theories and aggregate labor-market fluctuations. *The American Economic Review*, JSTOR, p. 430–450, 1992.
- CHUDIK, A.; PESARAN, M. H. Large panel data models with cross-sectional dependence: a survey. *CAFE Research Paper*, n. 13.15, 2013.
- CHUDIK, A.; PESARAN, M. H.; TOSETTI, E. *Weak and strong cross-section dependence and estimation of large panels*. [S.l.]: Oxford University Press Oxford, UK, 2011.
- COAKLEY, J.; FUERTES, A.-M.; SMITH, R. Unobserved heterogeneity in panel time series models. *Computational Statistics & Data Analysis*, Elsevier, v. 50, n. 9, p. 2361–2380, 2006.

- COGLEY, T.; NASON, J. M. Output dynamics in real-business-cycle models. *The American Economic Review*, JSTOR, p. 492–511, 1995.
- COHEN, W. M.; LEVINTHAL, D. A. Innovation and learning: the two faces of r & d. *The economic journal*, JSTOR, v. 99, n. 397, p. 569–596, 1989.
- DOSI, G.; NELSON, R. R. An introduction to evolutionary theories in economics. *Journal of evolutionary economics*, Springer, v. 4, n. 3, p. 153–172, 1994.
- DUGUET, E. Innovation height, spillovers and tfp growth at the firm level: Evidence from french manufacturing. *Economics of Innovation and New technology*, Taylor & Francis, v. 15, n. 4-5, p. 415–442, 2006.
- EBERHARDT, M.; BOND, S. Cross-section dependence in nonstationary panel models: a novel estimator. 2009.
- EBERHARDT, M.; HELMERS, C.; STRAUSS, H. Do spillovers matter when estimating private returns to r&d? *Review of Economics and Statistics*, MIT Press - Journals, v. 95, n. 2, p. 436–448, may 2013.
- EBERHARDT, M.; TEAL, F. Modeling technology and technological change in manufacturing: how do countries differ? 2008.
- EBERHARDT, M.; TEAL, F. Productivity analysis in global manufacturing production. 2010.
- EBERHARDT, M.; TEAL, F. Econometrics for grumblers: a new look at the literature on cross-country growth empirics. *Journal of Economic Surveys*, Wiley Online Library, v. 25, n. 1, p. 109–155, 2011.
- FAGERBERG, J. Technology and international differences in growth rates. *Journal of economic Literature*, JSTOR, v. 32, n. 3, p. 1147–1175, 1994.
- GARCIA, F. *A evolução da produtividade total de fatores na economia brasileira: uma análise do período pós-Real*. [S.l.]: EAESP/FGV/NPP, 2003.
- GRILICHES, Z. Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of economics*, v. 10, n. 1, p. 92–116, 1979.
- GRILICHES, Z. *The search for R&D spillovers*. [S.l.], 1991.
- GROSSMAN, G. M.; HELPMAN, E. Quality ladders and product cycles. *The Quarterly Journal of Economics*, MIT Press, v. 106, n. 2, p. 557–586, 1991.
- GUNDAY, G. et al. Effects of innovation types on firm performance. *International Journal of production economics*, Elsevier, v. 133, n. 2, p. 662–676, 2011.
- GUSSO, D. Agentes da inovação: quem os forma, quem os emprega. *Tecnologia, exportação e emprego*. Brasília: Ipea, p. 397–444, 2006.
- HALL, B. H.; MAIRESSE, J.; MOHNEN, P. Measuring the returns to r&d. In: *Handbook of the Economics of Innovation, Volume 2*. [S.l.]: Elsevier, 2010. p. 1033–1082.

- HARBERGER, A. C. The macroeconomics of successful development: What are the lessons?: Comment. *NBER macroeconomics annual*, MIT Press, v. 2, p. 255–258, 1987.
- HIGON, D. A. The impact of r&d spillovers on uk manufacturing tfp: A dynamic panel approach. *Research policy*, Elsevier, v. 36, n. 7, p. 964–979, 2007.
- KING, R. G.; PLOSSER, C. I.; REBELO, S. T. Production, growth and business cycles: Ii. new directions. *Journal of Monetary Economics*, Elsevier, v. 21, n. 2-3, p. 309–341, 1988.
- KYDLAND, F. E.; PRESCOTT, E. C. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 1345–1370, 1982.
- LUCINDA, C.; MEYER, L. Quão imperfeita é a competição na indústria brasileira?: estimativas de mark up setorial entre 1996 e 2007. *Estudos Econômicos (São Paulo)*, SciELO Brasil, v. 43, n. 4, p. 687–710, 2013.
- NELSON, R. R. *An evolutionary theory of economic change*. [S.l.]: harvard university press, 1985.
- NELSON, R. R. Modelling the connections in the cross section between technical progress and r&d intensity. *The Rand Journal of Economics*, JSTOR, p. 478–485, 1988.
- PESARAN, M. H. General diagnostic tests for cross section dependence in panels. CESifo working paper series, 2004.
- PESARAN, M. H. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, Wiley Online Library, v. 74, n. 4, p. 967–1012, 2006.
- PESARAN, M. H. A simple panel unit root test in the presence of cross-section dependence. *Journal of applied econometrics*, Wiley Online Library, v. 22, n. 2, p. 265–312, 2007.
- PESARAN, M. H.; SMITH, R. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics*, Elsevier, v. 68, n. 1, p. 79–113, 1995.
- PESARAN, M. H.; TOSETTI, E. Large panels with spatial correlations and common factors. *Journal of Econometrics*, Citeseer, v. 161, n. 2, p. 182–202, 2009.
- PLOSSER, C. I. Understanding real business cycles. *Journal of Economic Perspectives*, v. 3, n. 3, p. 51–77, 1989.
- ROMER, P. M. Endogenous technological change. *Journal of political Economy*, The University of Chicago Press, v. 98, n. 5, Part 2, p. S71–S102, 1990.
- SOLOW, R. M. Technical change and the aggregate production function. *The review of Economics and Statistics*, JSTOR, p. 312–320, 1957.
- STADLER, G. W. Real business cycles. *Journal of Economic Literature*, JSTOR, v. 32, n. 4, p. 1750–1783, 1994.

STEINGRABER, R.; GONÇALVES, F. d. O. Inovação, instituições e capital social na produtividade total dos fatores da indústria brasileira em 2005. *SALERNO, MS; NEGRI, JA de; TURCHI, LM; MORAIS, LM de. Inovação: estudos de jovens pesquisadores brasileiros*, v. 2, p. 119–147, 2010.

WIESER, R. Research and development productivity and spillovers: empirical evidence at the firm level. *Journal of Economic Surveys*, Wiley Online Library, v. 19, n. 4, p. 587–621, 2005.

ZUCOLOTO, G. et al. Lei do bem e produtividade das firmas industriais brasileiras. *Políticas de apoio à inovação tecnológica no Brasil: avanços recentes, limitações e propostas de ações*. Brasília: Ipea, 2017.

Apêndice

A Construção da Base de Dados

A.1 Dados da PIA

A.1.1 Setores Utilizados

A PIA contempla 111 grupos da CNAE, que representam setores agregados a 3 dígitos do código. Destes, conseguimos realizar a compatibilização entre o código antigo e novo da CNAE para 96 setores, sendo esse o número final de grupos na amostra. A tabela 6 apresenta os códigos e descrições dos setores utilizados.

A.1.2 Mudança da CNAE

A partir de 2007, a CNAE sofreu uma mudança na forma de classificação (CNAE 1.0 para 2.0), de modo que é necessário realizar uma compatibilização entre os dois códigos para que possam ser usados os dados antes de 2007. O IBGE disponibiliza uma tabela de correspondência entre as duas versões do código. Porém, alguns setores da CNAE 1.0 foram divididos na CNAE 2.0, enquanto outros foram agregados em um setor novo. Para que a correspondência possa ser feita, é preciso ter alguma medida de como se deu essa divisão ou união de setores. Neste trabalho, tal medida foi construída com base na informação do valor de produção dos produtos de cada setor, obtida da PIA-Produto de 2007, que foi disponibilizada nas duas versões do código.

Para ilustrar o procedimento, considere 3 setores: A, na CNAE 1.0, B e C na CNAE 2.0, todos agregados a um nível de 3 dígitos. Suponha que o setor A tenha 2 produtos A1 e A2 e que o primeiro represente 60% do valor da produção do setor, enquanto o segundo representa 40%. Assuma que, na CNAE 2.0, A1 pertença ao setor B e A2, ao setor C. Neste caso, divide-se o setor A da CNAE 1.0 entre os setores B e C da CNAE 2.0 utilizando as proporções citadas anteriormente como pesos, ou seja: 60% do setor A corresponde ao setor B e 40%, ao setor C.

Considere, por exemplo, o setor 160 da CNAE 1.0. Ele foi dividido de tal forma que alguns de seus produtos – correspondentes a 83% de seu valor de produção – foram associados ao setor 121 na CNAE 2.0, enquanto outros – correspondentes a 16% de seu valor de produção – foram associados ao setor 122 na CNAE 2.0. Por outro lado, o

Tabela 6: Setores utilizados nas estimações

Grupo	Denominação
071	Extração de minério de ferro
072	Extração de minerais metálicos não-ferrosos
081	Extração de pedra, areia e argila
089	Extração de outros minerais não-metálicos
091	Atividades de apoio à extração de petróleo e gás natural
101	Abate e fabricação de produtos de carne
102	Preservação do pescado e fabricação de produtos do pescado
103	Fabricação de conservas de frutas, legumes e outros vegetais
104	Fabricação de óleos e gorduras vegetais e animais
105	Laticínios
106	Moagem, fabricação de produtos amiláceos e de alimentos para animais
107	Fabricação e refino de açúcar
108	Torrefação e moagem de café
109	Fabricação de outros produtos alimentícios
111	Fabricação de bebidas alcoólicas
112	Fabricação de bebidas não-alcoólicas
121	Processamento industrial do fumo
122	Fabricação de produtos do fumo
131	Preparação e fiação de fibras têxteis
132	Tecelagem, exceto malha
133	Fabricação de tecidos de malha
134	Acabamentos em fios, tecidos e artefatos têxteis
135	Fabricação de artefatos têxteis, exceto vestuário
141	Confecção de artigos do vestuário e acessórios
142	Fabricação de artigos de malharia e tricotagem
151	Curtimento e outras preparações de couro
152	Fabricação de artigos para viagem e de artefatos diversos de couro
153	Fabricação de calçados
154	Fabricação de partes para calçados, de qualquer material
161	Desdobramento de madeira
162	Fabricação de produtos de madeira, cortiça e material trançado, exceto móveis
171	Fabricação de celulose e outras pastas para a fabricação de papel
172	Fabricação de papel, cartolina e papel-cartão
173	Fabricação de embalagens de papel, cartolina, papel-cartão e papelão ondulado
174	Fabricação de produtos diversos de papel, cartolina, papel-cartão e papelão ondulado
181	Atividade de impressão
182	Serviços de pré-impressão e acabamentos gráficos
192	Fabricação de produtos derivados do petróleo
193	Fabricação de biocombustíveis
201	Fabricação de produtos químicos inorgânicos
202	Fabricação de produtos químicos orgânicos
203	Fabricação de resinas e elastômeros
205	Fabricação de defensivos agrícolas e desinfestantes domissanitários
206	Fabricação de sabões, detergentes, produtos de limpeza, cosméticos, produtos de perfumaria e de higiene pessoal
207	Fabricação de tintas, vernizes, esmaltes, lacas e produtos afins
209	Fabricação de produtos e preparados químicos diversos
211	Fabricação de produtos farmoquímicos
212	Fabricação de produtos farmacêuticos
221	Fabricação de produtos de borracha
222	Fabricação de produtos de material plástico
231	Fabricação de vidro e de produtos do vidro
232	Fabricação de cimento
233	Fabricação de artefatos de concreto, cimento, fibrocimento, gesso e materiais semelhantes
234	Fabricação de produtos cerâmicos
239	Aparelhamento de pedras e fabricação de outros produtos de minerais não-metálicos
241	Produção de ferro-gusa e de ferroligas
242	Siderurgia
243	Produção de tubos de aço, exceto tubos sem costura
244	Metalurgia dos metais não-ferrosos
245	Fundição
251	Fabricação de estruturas metálicas e obras de caldeiraria pesada
252	Fabricação de tanques, reservatórios metálicos e caldeiras
253	Forjaria, estamparia, metalurgia do pó e serviços de tratamento de metais
254	Fabricação de artigos de cutelaria, de serralheria e ferramentas
259	Fabricação de produtos de metal não especificados anteriormente
261	Fabricação de componentes eletrônicos
262	Fabricação de equipamentos de informática e periféricos
263	Fabricação de equipamentos de comunicação
264	Fabricação de aparelhos de recepção, reprodução, gravação e amplificação de áudio e vídeo

265	Fabricação de aparelhos e instrumentos de medida, teste e controle; cronômetros e relógios
266	Fabricação de aparelhos eletromédicos e eletroterapêuticos e equipamentos de irradiação
271	Fabricação de geradores, transformadores e motores elétricos
272	Fabricação de pilhas, baterias e acumuladores elétricos
273	Fabricação de equipamentos para distribuição e controle de energia elétrica
274	Fabricação de lâmpadas e outros equipamentos de iluminação
275	Fabricação de eletrodomésticos
279	Fabricação de equipamentos e aparelhos elétricos não especificados anteriormente
281	Fabricação de motores, bombas, compressores e equipamentos de transmissão
282	Fabricação de máquinas e equipamentos de uso geral
283	Fabricação de tratores e de máquinas e equipamentos para a agricultura e pecuária
284	Fabricação de máquinas-ferramenta
285	Fabricação de máquinas e equipamentos de uso na extração mineral e na construção
286	Fabricação de máquinas e equipamentos de uso industrial específico
291	Fabricação de automóveis, camionetas e utilitários
292	Fabricação de caminhões e ônibus
293	Fabricação de cabines, carrocerias e reboques para veículos automotores
294	Fabricação de peças e acessórios para veículos automotores
295	Recondicionamento e recuperação de motores para veículos automotores
301	Construção de embarcações
309	Fabricação de equipamentos de transporte não especificados anteriormente
321	Fabricação de artigos de joalheria, bijuteria e semelhantes
323	Fabricação de artefatos para pesca e esporte
324	Fabricação de brinquedos e jogos recreativos
325	Fabricação de instrumentos e materiais para uso médico e odontológico e de artigos ópticos
329	Fabricação de produtos diversos
331	Manutenção e reparação de máquinas e equipamentos

setor 131 da CNAE 2.0. é formado por produtos que correspondem a 78% do valor de produção do setor 171 na CNAE 1.0 e por produtos que correspondem a 100% do valor de produção do setor 172 na CNAE 1.0. A construção dos setores 121, 122 e 131 da CNAE 2.0 deu-se então da seguinte forma:

$$VTI_{121} = 0.83 \times VTI_{160} \quad (21)$$

$$VTI_{122} = 0.16 \times VTI_{160} \quad (22)$$

$$VTI_{131} = 0.78 \times VTI_{171} + 1 \times VTI_{172}. \quad (23)$$

Desta forma, foi criado um dicionário para a compatibilização entre CNAE 1.0 e 2.0.

Para avaliar a qualidade da conversão, foram utilizados os dados da PIA de 2007, que, como dito acima, foram disponibilizados nas duas versões do código. Os dados de 2007 da CNAE 1.0 foram transformados para a CNAE 2.0 utilizando o dicionário e comparados com os valores verdadeiros disponibilizados na CNAE 2.0.

Primeiramente, foi realizado um teste de diferença na média global e local entre os valores calculados com o dicionário e os valores observados.³⁰ As tabelas 7 e 8 apresentam o p-valor destes testes. Os resultados indicam que não há diferença de média entre os valores reais e os calculados com o dicionário.³¹

Tabela 7: Teste de Diferença de Média

VTI	Pessoal	Aquisições	Melhorias	Baixas
0.52	0.44	0.76	0.82	0.56

Fonte: Elaboração própria

³⁰Os grupos utilizados no teste local foram definidos em Lucinda e Meyer (2013)

³¹P-Valores maiores que 0.99 arredondados para 1

Tabela 8: Teste de Diferença de Média Local

Grupo	VTI	Pessoal	Aquisições	Melhorias	Baixas
Extrativa	1.00	1.00	1.00	1.00	1.00
Alimentos	0.99	0.80	0.87	0.94	0.89
Têxtil	0.78	0.76	0.75	0.72	0.90
Florestal	0.98	0.97	0.99	0.95	0.82
Eletrônicos	0.95	0.87	0.72	0.83	0.77
Veículos	0.32	0.29	0.31	0.38	0.29
Metalurgia	0.87	0.69	0.92	1.00	0.95
Máquinas e Equipamentos	0.30	0.24	0.35	0.56	0.29
Química	0.92	0.78	0.83	0.67	0.70
Minerais Não-Metálicos	1.00	1.00	1.00	0.94	0.99
Outros	0.96	0.73	0.97	0.89	0.84

Fonte: Elaboração própria

Adicionalmente, foram comparadas a distribuição real das variáveis e a calculada pelo dicionário. Em geral, as distribuições são bastante próximas. A figura 6, apresenta esta comparação. Os resultados desta seção indicam que o dicionário realiza uma boa compatibilização entre CNAE 1.0 e 2.0.

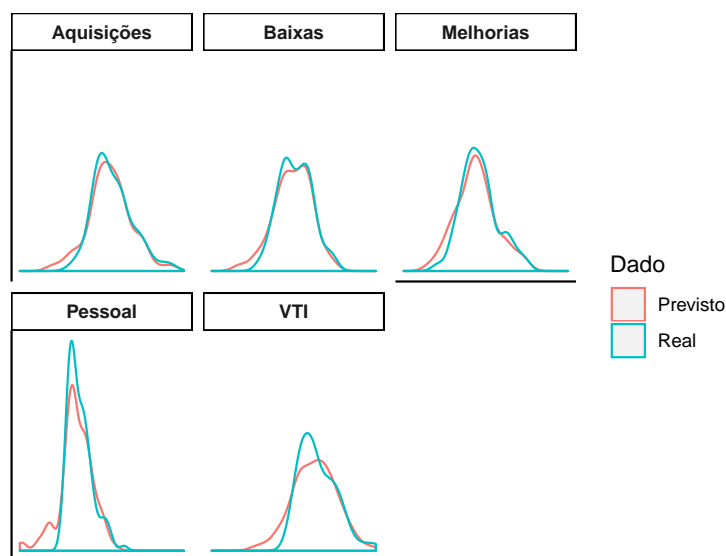


Figura 6: Distribuições reais e previstas pelo dicionário

A.1.3 Missing

Para alguns setores, a informação sobre as variáveis utilizadas não está completa, isto é, estes setores apresentam *missing* para algumas (ou todas) as variáveis em alguns períodos. Uma vez que os dados são agregados, isso levanta suspeita sobre a qualidade da informação dos setores que apresentam *missing*. Por exemplo, não parece razoável que um setor inteiro não apresente pessoal ocupado em determinado ano. Além disso, é necessária a série completa de investimento para construir o estoque de capital. Assim, cabe analisar se existe algum padrão entre os setores que apresentam *missing*.

A tabela 9 apresenta o número de *missing* por setor na CNAE 1.0 – dados entre 1996 e 2006 – e 2.0 – dados entre 2007 e 2017 – para cada variável que será utilizada da PIA.³²

Tabela 9: Ocorrência de Missing por Setor

(a) CNAE 1.0							(b) CNAE 2.0						
CNAE	VTI	Pessoal	Salarios	A	M	B	CNAE	VTI	Pessoal	Salarios	A	M	B
100	0	0	0	0	4	0	050	0	0	0	0	5	0
111	8	8	8	8	9	8	060	0	0	0	0	2	0
112	0	0	0	0	1	0	099	0	0	0	3	8	3
231	3	3	3	3	11	6	183	0	0	0	0	0	1
233	12	12	12	12	12	12	191	0	0	0	1	7	2
244	0	0	0	0	1	0	204	0	0	0	0	1	0
288	7	7	7	7	9	7	255	0	0	0	0	2	0
299	7	7	7	7	7	7	267	0	0	0	0	1	0
301	0	0	0	0	2	0	268	0	0	0	0	8	2
318	7	7	7	7	7	7	303	7	8	8	6	6	6
329	7	7	7	7	8	7	304	3	3	3	3	3	3
339	7	7	7	7	8	7	305	11	11	11	10	11	10
371	0	0	0	0	3	0	322	0	0	0	0	3	1
							332	0	0	0	0	1	0

Fonte: Elaboração Própria

Dos 111 setores da CNAE 2.0 a 3 dígitos, 14 apresentam *missing* em pelo menos uma variável. Os setores com a informação mais incompleta são 30.3 (Fabricação de veículos ferroviários) e 30.5 (Fabricação de veículos militares de combate). As variáveis associadas ao investimento são as que apresentam mais problema, em particular a de melhoria de ativos.

A tabela 10 apresenta o número de *missings* por ano para as variáveis. Um ponto interessante a notar é que o número de *missings* parece cair bastante a partir de 2003, o que coincide com a atualização da CNAE 1.0 implementada pelo IBGE ao final de 2002. Sendo assim, foram utilizados os dados da PIA apenas a partir de 2003 por considerar que a informação dos anos anteriores seria menos confiável.

A presença de *missing* dificulta o cálculo da variável de capital. Como o estoque de capital depende do fluxo de investimentos, seria necessário que a série de investimento de cada setor estivesse completa para construir o capital. Além disso, um número grande de *missings* pode indicar pouca confiabilidade dos dados de um setor. Assim, caso a presença de *missing* nos dados represente um erro de medida aleatório da pesquisa, seria ideal utilizar apenas setores com informação completa.

Porém, se houver alguma relação entre características econômicas do setor e o fato de sua informação estar incompleta, pode-se incorrer em um viés de seleção ao se excluir tais setores. Por exemplo, caso apenas setores de baixa intensidade tecnológica sejam excluídos, a elasticidade estimada para o P&D pode ser maior do que seu valor real. Para testar se existe relação entre a presença de *missing* e características econômicas, os setores com informação completa e incompleta foram comparados em termos do pessoal ocupado em taxa de crescimento e nível, como medidas da performance do setor e de seu tamanho, respectivamente. Estes setores foram também comparados em termos do seu nível de intensidade tecnológica.³³

As figuras 7 e 8 apresentam uma comparação gráfica do comportamento das 3 variáveis entre os setores com informação completa e incompleta. Não parece haver diferença

³²A, M e B são aquisições, melhorias e baixas de ativo respectivamente

³³A análise realizada segue, em partes, a metodologia proposta por Alves e Silva (2008)

Tabela 10: Ocorrência de Missing por Ano

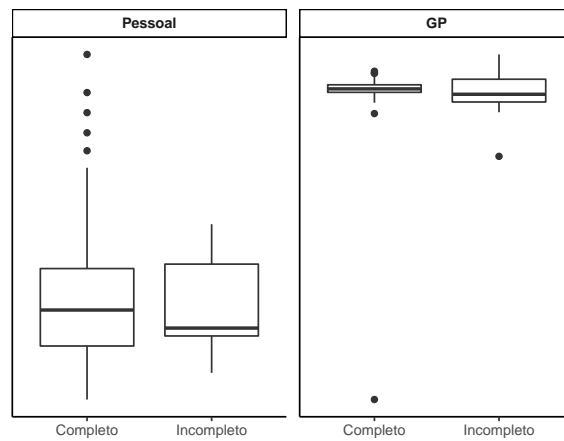
Ano	VTI	Pessoal	Salarios	A	M	B
1996	7	7	7	7	11	7
1997	7	7	7	7	9	7
1998	7	7	7	7	9	7
1999	7	7	7	7	10	7
2000	7	7	7	7	8	7
2001	7	7	7	7	8	8
2002	8	8	8	8	9	8
2003	2	2	2	2	4	2
2004	3	3	3	3	6	3
2005	1	1	1	1	2	1
2006	1	1	1	1	4	2
2007	2	2	2	2	4	3
2008	1	2	2	3	4	4
2009	2	2	2	1	5	1
2010	2	2	2	3	6	3
2011	2	2	2	2	5	2
2012	2	2	2	2	4	2
2013	2	2	2	2	5	2
2014	2	2	2	2	5	2
2015	2	2	2	2	6	2
2016	2	2	2	2	6	3
2017	2	2	2	2	8	4

Fonte: Elaboração própria

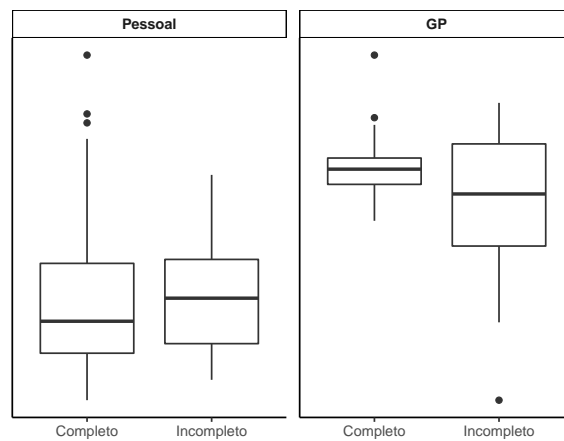
significativa entre a taxa de crescimento e o nível de pessoal ocupado entre os dois grupos de setores, tanto na CNAE 1.0 quanto na 2.0. Em relação à intensidade tecnológica, o grupo de setores com informação incompleta parece ser composto igualmente por setores de intensidade tecnológica alta, média-alta, média-baixa e baixa. Além disso, a amostra geral e a dos setores com informação completa parecem ter uma composição parecida dos grupos de intensidade tecnológica.

Para tornar a análise mais precisa, foi estimado um modelo logit usando uma *dummy* que assume 0 para setores completos como dependente e a taxa de crescimento e nível médios do pessoal do setor além do nível de intensidade tecnológica como explicativas. Esta estimação foi realizada tanto para os setores da CNAE 1.0 quanto 2.0. Os resultados, apresentados na tabela 11, indicam ausência de correlação entre características econômicas e tecnológicas dos setores e o fato de sua informação estar completa.³⁴

³⁴Erro-Padrão entre parênteses

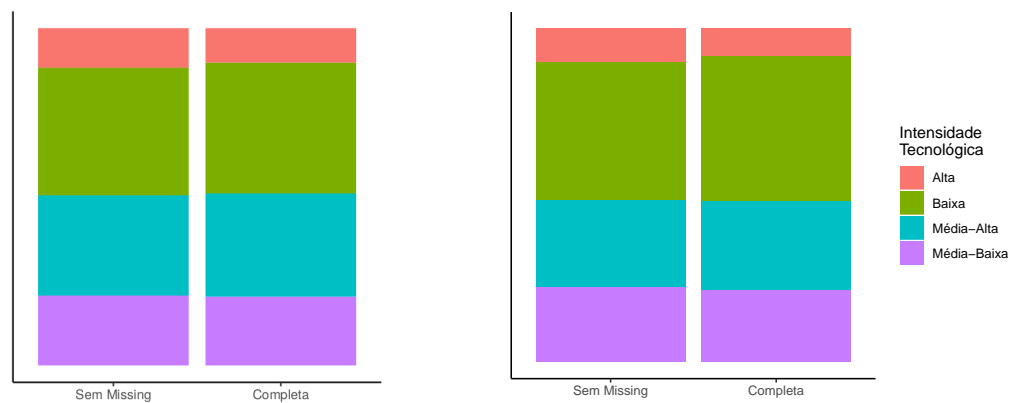


(a) 1996



(b) 2007

Figura 7: Características Econômicas dos Setores Completos e Incompletos



(a) Amostra completa x restrita - 1996

(b) Amostra completa x restrita - 2007

Figura 8: Quantidade de Missing por Intensidade Tecnológica

Tabela 11: Resultados - Logit

	CNAE 1.0	CNAE 2.0
log(Pessoal)	−0.192 (0.366)	−0.040 (0.279)
GP	−0.867 (6.359)	−13.703 (8.402)
Baixa	−1.046 (0.931)	−1.106 (0.917)
Média-Alta	−1.036 (0.973)	−2.145* (1.297)
Média-Baixa	−1.156 (1.121)	−0.911 (0.988)
Intercepto	−0.424 (1.774)	−1.246 (1.459)
Obs	110	109
Log Likelihood	−36.775	−32.188
Nota:	*p<0.1; **p<0.05; ***p<0.01	

A análise acima indica que a presença de *missing* nos dados não está associada a características econômicas, como baixo crescimento ou produtividade. Assim, a restrição da amostra a setores com informação completa não deve induzir um viés de seleção.

A.2 Descritivas Adicionais

A tabela 12 apresenta a correlação do Potec com cada variável da Pintec associada a gasto em P&D. A primeira coluna, (1), diz respeito apenas aos setores da indústria de transformação enquanto a segunda coluna, (2), contempla também valores agregados para a indústria extrativa.

Tabela 12: Comparação entre Potec e dados de gastos em inovação da Pintec

Gastos em Inovação	(1)	(2)
Total	0.84	0.68
Atividades internas de pesquisa e desenvolvimento	0.59	0.49
Aquisição externa de pesquisa e desenvolvimento	0.61	0.49
Aquisição de outros conhecimentos externos	0.87	0.75
Aquisição de software	0.45	0.36
Aquisição de máquinas e equipamentos	0.79	0.64
Treinamento	0.66	0.56
Introdução das inovações tecnológicas no mercado	0.85	0.79
Projeto industrial e outras preparações técnicas	0.83	0.76

Fonte: Elaboração própria

B Resultados da PTF

A tabela 13 apresenta os resultados das regressões 19 e 20. Como discutido na seção 4, os setores de alta e média-alta intensidade tecnológica apresentam um nível médio da PTF próximo entre si e significativamente acima dos setores de média-baixa e baixa

intensidade. Já com relação à variação da PTF, é possível notar que os setores de alta intensidade se destacam com um crescimento muito acima dos outros, enquanto os setores de média-baixa intensidade não apresentaram crescimento estatisticamente significativo e os setores de média-alta e baixa intensidade apresentaram crescimento baixo e próximo entre si.

Tabela 13: Regressões da PTF por Grau de Intensidade Tecnológica

	PTF	Δ PTF
Alta	9.105*** (0.163)	0.084*** (0.025)
Média-Alta	9.084*** (0.089)	0.033** (0.014)
Média-Baixa	8.661*** (0.102)	0.008 (0.015)
Baixa	8.416*** (0.072)	0.034*** (0.011)
Obs.	1,396	1,300
R ²	0.961	0.021
R ² Ajustado	0.961	0.018
Nota:	*p<0.1; **p<0.05; ***p<0.01	