

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade de
Ribeirão Preto
Departamento de Economia
Programa de Pós-graduação em Economia - Área: Economia
Aplicada

Rafael Campello de Alcantara

Estimação da Elasticidade de P&D no Brasil: Uma Abordagem de
Fatores Comuns Não-Observados

Orientador: Prof. Dr. Sérgio Kannebley Júnior

Ribeirão Preto

2019

Prof. Dr. Vahan Agopyan
Reitor da Universidade de São Paulo

Prof. Dr. André Lucirton Costa
Diretor da Faculdade de Economia, Administração e Contabilidade de
Ribeirão Preto

Prof. Dr. Sérgio Kannebley Júnior
Chefe do Departamento de Economia

Prof. Dr. Sergio Naruhiko Sakurai
Coordenador do Programa de Pós-Graduação em Economia - Área:
Economia Aplicada

RAFAEL CAMPELLO DE ALCANTARA

Estimação da Elasticidade de P&D no Brasil: Uma Abordagem
de Fatores Comuns Não-Observados

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Sérgio Kannebley Júnior

Ribeirão Preto

2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Alcantara, Rafael Campello de
Estimação da Elasticidade de P&D no Brasil: Uma Abordagem de Fatores Comuns Não-Observados/ Rafael Campello de Alcantara; Orientador: Prof. Dr. Sérgio Kannebley Júnior
Ribeirão Preto, 2019- 24 p. : il.

Dissertação (Mestrado) – Universidade de São Paulo, 2019.

1. Elasticidade de P&D 2. Fatores Comuns Não-Observados 3. *Spillover* de Conhecimento 4. Produtividade I. Orientador: Prof. Dr. Sérgio Kannebley Júnior. II. Universidade De São Paulo - Campus Ribeirão Preto. III. Faculdade de Economia, Administração e Contabilidade. IV. Estimação da Elasticidade de P&D no Brasil: Uma Abordagem de Fatores Comuns Não-Observados.

CDU

Agradecimentos

Primeiramente, agradeço a meus pais, Antônio e Iara, por sempre estarem comigo, sem nunca duvidar de minhas escolhas e da minha capacidade, e por estarem sempre dispostos a me ajudar como possível. Este trabalho – bem como qualquer outro objetivo que eu tenha alcançado ou venha a alcançar na vida – só foi possível por causa desse apoio incondicional.

Ao meu orientador, Sérgio Kannebley, por guiar as escolhas feitas ao longo deste trabalho e por proporcionar meu amadurecimento como pesquisador. Aos membros da banca, por aceitarem meu convite e por sua valiosa avaliação deste trabalho. Aos professores do PPGE-FEARP/USP com quem tive contato por meio de aulas ou outras atividades acadêmicas e aos professores com quem tive contato durante a graduação na UnB, por sua contribuição em minha formação acadêmica.

Aos amigos que fiz no mestrado – companheiros de turma, calouros e veteranos – e que espero levar para a vida, por todo apoio acadêmico e, principalmente, pessoal. Vocês certamente fizeram este tempo em Ribeirão ser mais agradável e valer mais a pena.

Aos amigos de Brasília, que já me aturam desde os tempos de colégio ou faculdade. Por me entenderem sempre, me ouvirem quando preciso falar e sempre me estimularem a dar o meu melhor. E, principalmente, por sempre conseguirem arrancar uma risada mesmo nos momentos mais difíceis.

Ao IBRE/FGV pela cessão dos dados necessários tornando possível esta pesquisa.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -Brasil (CAPES) - Código de Financiamento 001.

Finalmente, agradeço a Deus por me fazer chegar até aqui com tantos motivos para agradecer.

Resumo

ALCANTARA, R. C. **Estimação da Elasticidade de P&D no Brasil: Uma Abordagem de Fatores Comuns Não-Observados**. 2019. Dissertação (Mestrado) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2019.

A existência de *spillovers* de conhecimento, choques de produtividade e relações entre setores econômicos justifica a presença de fatores comuns não-observados no contexto da estimação de funções de produção. Este trabalho tem por objetivo estimar a elasticidade de P&D, controlando a presença de fatores comuns não-observados. Utilizando dados das indústrias extrativa e de transformação no Brasil, a elasticidade de P&D foi estimada aplicando o estimador *augmented mean group* (AMG) de Eberhardt e Teal (2010), que considera explicitamente um modelo de fatores comuns. A elasticidade de P&D obtida pelo AMG foi de 0.014. Porém, tal estimativa não foi estatisticamente significativa. A comparação com outros estimadores que não controlam a presença de fatores comuns indica que ignorar esta propriedade causa um viés positivo na estimativa da elasticidade de P&D. Este resultado aponta para a dificuldade de separar o retorno de gastos em P&D do efeito de *spillovers* e outros choques econômicos. Adicionalmente, os resultados do AMG foram utilizados para calcular a produtividade total dos fatores (PTF) dos setores como proposto por Eberhardt e Teal (2010). A PTF apresenta uma evolução temporal que vai de acordo com outras estimativas da PTF no Brasil, apontando para a qualidade das estimativas do AMG. Além disso, a PTF apresenta uma evolução próxima à das variáveis utilizadas. Isso indica uma possível relação de cointegração entre a PTF, não-observada, e as variáveis da função de produção. Finalmente, altos níveis e taxas de crescimento da PTF estão positivamente associados à orientação exportadora dos setores e à alta intensidade tecnológica.

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

Abstract

ALCANTARA, R. C. **Estimation of the R&D Elasticity in Brazil: An Unobserved Common Factor Approach** 2019. Dissertation (Master Degree) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2019.

The existence of knowledge spillovers, productivity shocks and relations between economic sectors justifies the presence of unobserved common factors in the context of estimating production functions. The aim of this research is to estimate the R&D elasticity controlling for the presence of unobserved common factors. Using a panel of Brazilian industrial data, the R&D elasticity was estimated by the augmented mean group estimator (AMG) of Eberhardt e Teal (2010), which explicitly assumes this property. The R&D elasticity obtained by the AMG estimator was of 0.014. However, this estimate was not statistically significant. A comparison with other estimators that do not control for the presence of common factors indicates that ignoring this property causes an upward bias in the estimated R&D elasticity. This result suggests that it is difficult to separate the returns to R&D from the effect of knowledge spillovers. Additionally, the results obtained by the AMG estimator were used to calculate the total factor productivity (TFP) of each sector as proposed by Eberhardt e Teal (2010). The TFP shows a temporal evolution similar to other TFP measures calculated for Brazil, which highlights the quality of the AMG estimates. Furthermore, the TFP shows a temporal pattern close to the variables used in the estimation. This indicates a possible cointegration relationship between unobserved TFP and the variables in the production function. Finally, high levels and growth rates of TFP are positively related to export orientation and to high levels of technological intensity.

Key-words: R&D elasticity, unobserved common factors, knowledge spillovers, total factor productivity

Lista de ilustrações

Figura 1 – Produto e Insumos por Hora de Trabalho	52
Figura 2 – Evolução Média do Produto por Setor	53
Figura 3 – Evolução média da PTF por Setor	60
Figura 4 – Relação entre PTF e intensidade tecnológica	60
Figura 5 – Relação entre PTF e Exportações	61
Figura 6 – Distribuições reais e previstas pelo dicionário	73
Figura 7 – Quantidade de Missing por Intensidade Tecnológica	81
Figura 8 – Quantidade de Missing por Intensidade Tecnológica	82
Figura 9 – Potec por Intensidade Tecnológica da OCDE	84
Figura 10 – Evolução Média das Variáveis por Setor	85

Lista de tabelas

Tabela 1 – Ocorrência de Missing por Setor - CNAE 2.0	45
Tabela 2 – Ocorrência de Missing por Ano	46
Tabela 3 – Ocupações do Potec	49
Tabela 4 – Estatísticas Descritivas	51
Tabela 5 – Teste Preliminares	54
Tabela 6 – Estimação com Pessoal Total	56
Tabela 7 – Estimação com Potec	58
Tabela 8 – Estimação sem Potec com Dados a Partir de 2007	62
Tabela 9 – Estimação com Dados a Partir de 2007	63
Tabela 10 – Teste de Diferença de Média Global	72
Tabela 11 – Teste de Diferença de Média Local	72
Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0	73
Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0	74
Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0	75
Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0	76
Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0	77
Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0	78
Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0	79
Tabela 13 – Setores com Missing - CNAE 1.0	79
Tabela 14 – Resultados - Logit	80
Tabela 15 – Comparação entre Potec e dados de gastos em inovação da Pintec	83

Sumário

	Lista de ilustrações	7
	Lista de tabelas	8
	Sumário	9
1	INTRODUÇÃO	11
2	REVISÃO DA LITERATURA	16
2.1	Estrutura Analítica	16
2.2	Principais Resultados	22
2.2.1	Elasticidade de P&D	22
2.2.2	Spillovers e PTF	24
3	METODOLOGIA	26
3.1	Estimadores de Painel com Fatores Comuns Não-Observáveis	27
3.2	Estratégia de Estimação	32
3.2.1	Estimador AMG	32
3.2.2	Cálculo da PTF	33
3.2.3	Estimações Adicionais	36
3.2.4	Testes dos Resíduos	39
4	BASE DE DADOS	41
4.1	Fontes de Dados	41
4.1.1	Pesquisa Industrial Anual (PIA)	41
4.1.2	Relação Anual de Informações Sociais (RAIS)	47
4.1.3	Fundação Centro de Estudos do Comércio Exterior (Funcex)	47
4.2	Variáveis	48
4.2.1	Investimento em P&D	48
4.2.2	Trabalho	50
4.2.3	Capital	50
4.2.4	Produto	51
4.3	Estatísticas Descritivas	51
4.3.1	Testes Preliminares	54
5	RESULTADOS	56
5.1	Resultados Principais	56

5.1.1	Estimação Sem P&D	56
5.1.2	Estimação com P&D	57
5.2	Cálculo da PTF	59
5.3	Estimação com dados a partir de 2007	61
6	CONCLUSÃO	65
	REFERÊNCIAS	68
A	CONSTRUÇÃO DA BASE DE DADOS	72
A.1	PIA	72
A.1.1	Mudança da CNAE	72
A.1.2	Missing	79
B	POTEC X DADOS DE INOVAÇÃO	83
C	DESCRITIVAS ADICIONAIS	85

1 Introdução

A inovação tecnológica constitui um dos principais meios de que as empresas dispõem para aumentar sua produtividade. Para melhor entender a relação entre tecnologia e crescimento ou avaliar o impacto de políticas de incentivo à inovação, muitos trabalhos estimam o impacto de atividades de P&D na produtividade. A abordagem mais utilizada para tal, proposta por Griliches (1979), consiste em estimar uma função de produção incluindo como insumo alguma medida de estoque ou dispêndio interno em P&D, além dos insumos tradicionais, como capital e trabalho. Inclui-se, ainda, uma medida de conhecimento externo adquirido, para controlar o efeito de *spillovers*. Pode-se, então, estimar a elasticidade do produto com relação ao P&D, ou seja, a variação percentual na produção implicada por um aumento de 1% no investimento em pesquisa.

Como, teoricamente, *spillovers* não são diretamente observáveis ou mensuráveis, na estimação da função de produção como descrita acima, este efeito seria capturado pelo termo de erro. Eberhardt, Helmers e Strauss (2013) notam que, como o investimento privado em P&D tende a estar correlacionado com a geração de *spillovers*, a presença deste efeito no erro levaria à endogeneidade da variável de P&D. Consequentemente, a estimação da elasticidade de P&D sem levar em conta o efeito de *spillovers* seria inconsistente.

Além disso, o estoque de conhecimento externo é uma variável comum a todos os setores. Assim, uma vez que a variável de *spillover* está presente nos erros de todos os setores, tem-se que estes erros podem estar correlacionados entre si, propriedade conhecida como dependência *cross-section*.

Muitos trabalhos buscam modelar explicitamente a dependência *cross-section* gerada pelos *spillovers* por meio da construção de uma *proxy* para o estoque de conhecimento externo. Eberhardt, Helmers e Strauss (2013) notam, porém, que esta não é uma solução adequada para controlar a dependência *cross-section* na função de produção por dois motivos. Primeiro, como discutido por Griliches (1991), é difícil saber exatamente como o conhecimento se espalha pela economia. Eberhardt, Helmers e Strauss (2013) argumentam

que, por isso, qualquer estratégia para construir uma medida de *spillover* implica assumir hipóteses muito fortes sobre esta variável e estaria sujeita a críticas. Neste sentido, Hall, Mairesse e Mohnen (2010) notam que, em geral, as medidas propostas na literatura são muito imprecisas.

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. Eberhardt, Helmers e Strauss (2013) argumentam que esta propriedade 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* de *spillovers* 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 que afetam todas as unidades do painel. Por isso, Eberhardt, Helmers e Strauss (2013) argumentam que a representação da função de produção 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 capturariam *spillovers*, ciclos econômicos, choques de oferta ou 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.

No tocante aos valores estimados para a elasticidade de P&D na literatura internacional, Hall, Mairesse e Mohnen (2010) revisam uma série de estudos com dados de

firmas e indústrias. Os autores apresentam estimativas baseadas na variação *cross-section* e temporal dos dados separadamente¹. No primeiro caso, os valores variam entre 0.006 e 0.25, com uma média em torno de 0.098. Já no segundo, os valores se situam entre -0.015 e 0.32, com uma média de 0.084. Os autores notam, ainda, que as estimativas baseadas na variação temporal dos dados não são estatisticamente significantes em geral. Por fim, Hall, Mairesse e Mohnen (2010) apontam evidências de que a inclusão de *dummies* individuais, para controlar diferenças fixas entre os setores, leva a elasticidades de P&D menores.

Wieser (2005) realiza uma meta-análise de vários estudos, também sobre firmas e indústrias. O autor encontra estimativas da elasticidade de P&D entre -0.016 e 0.38, com uma mediana de 0.1. Também são encontradas evidências de que a inclusão de *dummies* individuais leva a estimativas menores. Finalmente, Wieser (2005) nota que a elasticidade de P&D tende a variar por setor, sendo mais alta, em geral, nos setores mais intensivos em tecnologia.

Os trabalhos discutidos não consideram a possível dependência *cross-section* na função de produção. Para preencher esta lacuna, 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 é menor e não-significante. Segundo os autores, isso evidencia a dificuldade de se separar o retorno privado de P&D do efeito de *spillovers*.

Existe pouca evidência sobre a elasticidade de P&D no Brasil. Analisando firmas das indústrias extrativa e de transformação entre os anos de 2000 e 2008, Cavalcante, Jacinto e Negri (2015) reportam valores entre 0.06 e 0.1, sendo este valor maior para os setores mais intensivos em tecnologia. Santana (2017) estima, para o período de 2006 a 2013, que a elasticidade se encontra entre 0.016 e 0.169, com média em torno de 0.07. Além disso, não foram encontrados estudos que apliquem o modelo de fatores comuns

¹ Para a variação individual, são avaliados dados de *cross-section* ou, no caso de dados em painel, são apresentados resultados para os dados em termos das médias individuais (transformação *between*). Já para a variação temporal, são considerados dados em termos da taxa de crescimento ou de desvios da média individual para cada firma (transformação *within*)

para a estimação da função de produção no Brasil.

O presente estudo busca contribuir com esta literatura por meio da estimação da elasticidade de P&D na indústria brasileira. Para isso, foi estimada uma função de produção com um painel de setores das indústrias extrativa e de transformação brasileiras entre os anos de 2003 e 2017. Os dados são agregados a um nível de 3 dígitos da Classificação Nacional de Atividade Econômica (CNAE).

Dada a discussão apresentada sobre a importância de *spillovers* e outras fontes de dependência *cross-section* na estimação da elasticidade de P&D, este estudo considera a estrutura analítica proposta por Eberhardt e Teal (2010). Tal abordagem consiste em representar a função de produção e as relações estruturais implícitas a ela por meio de um modelo de fatores comuns não-observados. Para levar em conta esta estrutura na estimação, foi utilizado o estimador *augmented mean group* (AMG), também proposto por Eberhardt e Teal (2010).

Neste sentido, as principais contribuições deste trabalho para a literatura econômica brasileira são as seguintes. Primeiro, pela produção de estimativas atualizadas da elasticidade de P&D no Brasil. Depois, pela aplicação de uma metodologia empírica mais robusta para estimar a função de produção, gerando estimativas mais confiáveis.

A elasticidade de P&D obtida pelo AMG foi de 0.014. Esta estimativa, porém, não foi estatisticamente significativa. A comparação com outros estimadores² indica que não controlar a dependência *cross-section* nos dados produz viés positivo na estimativa da elasticidade de P&D. Este resultado corrobora o obtido por Eberhardt, Helmers e Strauss (2013). Como discutido pelos autores, isso evidencia a dificuldade de separar o retorno do investimento em P&D do efeito de *spillovers* e outros choques econômicos. Uma consequência disso, por exemplo, é que políticas de inovação devem considerar formas de tornar o conhecimento mais acessível entre os setores, facilitando o *spillover*.

Os resultados do AMG foram utilizados, ainda, para calcular a produtividade total dos fatores (PTF) dos setores analisados como proposto por Eberhardt e Teal (2010). É

² Descritos com mais detalhe no capítulo 3

possível observar que a PTF segue uma evolução temporal próxima à das variáveis, em especial produto e trabalho. Isso indica que há correlação entre a PTF, não-observável, e as variáveis da função de produção. Assim, justifica-se a preocupação de levar em conta a influência da produtividade e de *spillovers* na estimação. Adicionalmente, os resultados apontam para a correlação da PTF com intensidade tecnológica e orientação exportadora.

O restante do trabalho está organizado da seguinte forma. No capítulo 2, é discutida a relação entre P&D e produtividade. Discute-se o modelo da função de produção e sua representação por meio de uma estrutura de fatores comuns não-observados. Além disso, são apresentadas as principais dificuldades empíricas implicadas pela estrutura analítica assumida neste trabalho. Finalmente, os principais resultados na literatura nacional e internacional sobre a elasticidade de P&D são brevemente discutidos.

Já no capítulo 3, é descrita a metodologia empregada neste trabalho. Explica-se o procedimento do estimador AMG e sua relação com a estrutura analítica proposta. São discutidos também os testes de robustez realizados para avaliar a validade das hipóteses assumidas e a qualidade da estimação. O capítulo 4 contém uma descrição das fontes de dados deste trabalho. São explicados o processo de construção da base de dados e das variáveis utilizadas nas estimações, ressaltando a relação entre as variáveis do modelo teórico e as construídas para esta pesquisa.

No capítulo 5 são apresentados os resultados obtidos nesta pesquisa, discutindo como estes se relacionam com as hipóteses feitas ao longo do trabalho sobre a relação entre P&D, *spillovers* e produtividade e sobre a estrutura analítica empregada. Por fim, o capítulo 6 sumariza as discussões teóricas e resultados empíricos do trabalho, ressaltando as principais limitações e possíveis extensões a esta pesquisa.

2 Revisão da Literatura

Como descrito no capítulo 1, este trabalho tem entre seus objetivos estimar, a partir de um modelo de fatores comuns não-observados, a elasticidade de P&D nas indústrias extrativa e de transformação no Brasil. Este capítulo pretende discutir como a literatura econômica justifica e avalia a relação entre P&D e produtividade. Para tanto, é apresentada a estrutura analítica sob a qual se estuda tal relação. Discute-se ainda, seguindo Eberhardt e Teal (2010), como tal estrutura pode ser representada por um modelo de fatores comuns não-observados. Também são descritas as principais dificuldades empíricas na estimação da elasticidade de P&D implicadas pelo modelo teórico. Por fim, os principais resultados sobre este parâmetro são brevemente discutidos.

2.1 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 pesquisa e desenvolvimento (P&D) visando a produção de inovações.

Arrow (1972) argumenta ainda 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 através da imitação, troca de informação entre trabalhadores, limitações do sistema de patentes *etc.* Este efeito, conhecido como *spillover* do conhecimento, indica que, além do gasto próprio em P&D, a produção externa de conhecimento também pode

afetar a produtividade das firmas.

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

Segundo Hall, Mairesse e Mohnen (2010), a forma mais comum 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 insumos 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 *spillovers* de conhecimento. Considere a seguinte função de produção *Cobb-Douglas*:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} R_{it}^{\gamma} [R_{it}^0]^{\phi} e^{u_{it}}, \quad (2.1)$$

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. De acordo com Hall, Mairesse e Mohnen (2010), é comum, ainda, a aplicação de logaritmo sobre (2.1) para se obter um modelo linear, o que facilita a estimação. Após tal transformação, obtém-se

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma r_{it} + \phi r_{it}^0 + \eta_i + \theta_t + u_{it}, \quad (2.2)$$

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 .

Na tentativa de controlar o efeito de *spillovers* na estimação, muitos trabalhos propõem *proxies* desta variável. Como discutido por Hall, Mairesse e Mohnen (2010), as medidas de *spillover* são construídas, em geral, como uma média ponderada dos investimentos em P&D dos setores. Os pesos, segundo os autores, são calculados, por exemplo, com base em relações comerciais entre setores, proximidade tecnológica, no uso de patentes de outros setores *etc.*

Eberhardt, Helmers e Strauss (2013) notam, porém, que esta não é uma boa solução para o problema. Primeiro, porque é difícil saber como o conhecimento se espalha pela economia. Assim, qualquer escolha de pesos para a construção da *proxy* do *spillover* estaria sujeita a críticas. Além disso, alguns dos fatores que determinam a PTF também influenciam a escolha de insumos das firmas, incluindo o investimento em P&D. Assim, a PTF representa outra fonte de endogeneidade na estimação de (2.2), que não seria controlada pela *proxy* de *spillover*¹.

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 tem o mesmo impacto sobre a produção de todos os setores. Se este não for o caso, a *dummy* capturaria apenas parcialmente o componente temporal da PTF, não corrigindo para a correlação deste componente com os insumos da função de produção. A inclusão da tendência, por sua vez, implica assumir que a PTF é estacionária em torno da tendência. Caso a PTF apresente raiz unitária, a tendência não é suficiente para capturar o padrão temporal da variável. Então, o resíduo da função de produção será não-estacionário, causando um problema de regressão espúria.

¹ Exemplos 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)

Como solução a estes problemas, Eberhardt, Helmers e Strauss (2013) sugerem que o *spillover* pode ser visto como um estoque comum de conhecimento que impacta cada setor de forma diferenciada. Mais ainda, os autores argumentam que a evolução da PTF é determinada por uma combinação de choques econômicos comuns, como crises financeiras globais, políticas econômicas ou choques de oferta e demanda. Nesse caso, a produtividade dos setores difere por causa de suas características individuais e de diferenças no impacto dos choques comuns. Os autores sugerem, então, que a PTF e os *spillovers* sejam modelados como uma combinação de fatores comuns latentes.

Com base nessas considerações, a equação (2.2) pode ser reescrita como

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma r_{it} + \eta_i + \mathbf{\Lambda}'_i \mathbf{f}_t + u_{it}, \quad (2.3)$$

em que $\mathbf{\Lambda}'_i \mathbf{f}_t = \sum_{j=1}^M \lambda_{ji} f_{jt}$ e M é o número de fatores comuns. O termo \mathbf{f}_t representa um vetor de fatores comuns associados à PTF, bem como *spillovers* de conhecimento. Já $\mathbf{\Lambda}'_i$ 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 supôr que existe correlação entre a escolha dos insumos e fatores comuns associados à produtividade, além do *spillover* no caso do P&D. Para capturar este efeito, Eberhardt e Teal (2010) propõem modelar as variáveis da função de produção como uma combinação entre os fatores comuns latentes, \mathbf{f}_t e outros componentes não relacionados diretamente à PTF ou ao *spillover*:

$$\mathbf{x}_{it} = \mathbf{\Pi}_i \mathbf{g}_t + \mathbf{\Psi}_i \mathbf{f}_t + v_{it}, \quad (2.4)$$

em que \mathbf{x}_{it} é o vetor de insumos (capital, trabalho e P&D), \mathbf{g}_t é um vetor de fatores comuns não associados à produtividade, como características específicas dos bens

produzidos pelos setores, e v_{it} é um erro idiossincrático.

Outra questão discutida por Eberhardt e Teal (2010) na modelagem da função de produção é a possível presença de raiz unitária nas variáveis do modelo. Intuitivamente, esta propriedade, também chamada de não-estacionariedade, significa que os dados apresentam grande persistência temporal, de modo que um choque em determinado período continua a afetar as variáveis muitos períodos à frente. Os autores notam que existe na literatura forte evidência empírica de que esta é uma característica relevante das variáveis da função de produção, em especial produto e capital.

Neste sentido, Abdih e Joutz (2006) e Jones (1995) trazem evidências de que os gastos em P&D também apresentam alta persistência temporal. Romer (1989) argumenta ainda que a produtividade deve apresentar dependência no tempo. Segundo o autor, produzir conhecimentos mais avançados, por exemplo, só é possível graças ao conhecimento já existente e ao grande esforço para se aprender coisas mais básicas. Sendo assim, seria razoável supor que *spillovers* de conhecimento e choques de produtividade em geral apresentem certo grau de dependência temporal, implicando que alguns dos fatores comuns \mathbf{f}_t podem também não ser estacionários.

Dada a estrutura de fatores comuns proposta para a função de produção, Eberhardt e Teal (2010) propõem assumir que a dependência temporal no modelo pode vir tanto dos fatores comuns latentes da equação (2.3) como dos componentes específicos das variáveis, \mathbf{g}_t . Neste caso, tem-se:

$$\mathbf{f}_t = \mathbf{P}'\mathbf{f}_{t-1} + \epsilon_{it}, \quad (2.5)$$

$$\mathbf{g}_t = \mathbf{K}'\mathbf{g}_{t-1} + \varepsilon_{it}, \quad (2.6)$$

em que \mathbf{K} e \mathbf{P} medem o grau de dependência temporal dos fatores comuns e ϵ_{it} e ε_{it} são erros idiossincráticos.

Outra hipótese importante por trás de (2.3) 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. Desse modo, seria importante permitir que as elasticidades α , β e γ em (2.3) variassem entre setores.

Além da questão conceitual, a imposição de homogeneidade dos parâmetros pode gerar um problema de regressão espúria caso os regressores sejam não-estacionários, como mostrado por Eberhardt e Teal (2011). Mais ainda, os autores mostram que, caso os fatores comuns do erro sejam não-estacionários, este problema pode surgir mesmo que se leve em conta a estrutura fatorial do erro na estimação. Desta forma, Eberhardt e Teal (2011) argumentam que parece mais razoável utilizar uma metodologia econométrica que permita heterogeneidade dos parâmetros na estimação da função de produção.

Para entender este ponto, considere, seguindo Eberhardt e Teal (2011), a regressão com parâmetros homogêneos. Se o processo verdadeiro dos dados for de parâmetros heterogêneos, o resíduo contém $(\beta_i - \hat{\beta})x_{it}$ e $(\lambda_{ji} - \hat{\lambda}_j)f_{jt}$, em que x_{it} é um insumo, f_{jt} é um fator comum, β_i é a elasticidade de x no setor i , λ_{ji} é o impacto do fator f_{jt} no setor i , $\hat{\beta}$ e $\hat{\lambda}$ são os coeficientes comuns estimados. Assim, caso x_{it} ou f_t sejam não-estacionários, o resíduo também será, de modo que a regressão será espúria. Também é possível notar que, como o fator comum f_t ainda vai estar presente no erro, ainda haverá correlação entre os erros dos setores. Nesse sentido, a imposição equivocada de homogeneidade dos parâmetros implica, ainda, que a estimação não controlará de forma completa a dependência *cross-section*.

Com base nas equações (2.4), (2.5), (2.6) e levando em conta a importância de se permitir a heterogeneidade das elasticidades dos insumos, Eberhardt e Teal (2010) propõem o seguinte modelo analítico para a função de produção:

$$y_{it} = \mathbf{B}'_i \mathbf{x}_{it} + \eta_i + \mathbf{\Lambda}'_i \mathbf{f}_t + u_{it}, \quad (2.7)$$

$$\mathbf{x}_{it} = \mathbf{\Pi}_i \mathbf{g}_t + \mathbf{\Psi}_i \mathbf{f}_t + v_{it}, \quad (2.8)$$

$$\mathbf{f}_t = \mathbf{P}' \mathbf{f}_{t-1} + \epsilon_{it}, \quad (2.9)$$

$$\mathbf{g}_t = \mathbf{K}' \mathbf{g}_{t-1} + \varepsilon_{it}. \quad (2.10)$$

Vale ressaltar que, como o erro u_{it} é assumido aleatório, caso alguns dos fatores comuns do erro sejam não-estacionários, o modelo (2.7)-(2.10) implica uma relação de cointegração entre produto, insumos e os fatores comuns. Assim, Eberhardt e Teal (2011) argumentam que a estimação sem levar em conta o efeito dos fatores comuns pode produzir resultados espúrios, uma vez que não captaria esta relação.

2.2 Principais Resultados

2.2.1 Elasticidade de P&D

Cabe, ainda, discutir os valores encontrados para a elasticidade de P&D na literatura, a fim de avaliar melhor os resultados obtidos neste trabalho. Hall, Mairesse e Mohnen (2010) revisam uma série de estudos que estimam essa elasticidade. Os trabalhos analisados utilizam dados de firmas e indústrias para anos entre 1958 e 2006, sendo que a maioria se concentra entre 1970 e 1990. A maior parte dos dados são americanos ou europeus.

Os autores apresentam estimativas com as variáveis em nível – *cross-section* ou painel – e com as variáveis em taxas de crescimento ou desvios das médias individuais. No primeiro caso, os valores variam entre 0.006 e 0.25, com uma média em torno de 0.098. Já no segundo, os valores se situam entre -0.015 e 0.32, com uma média de 0.084. A maior parte dos trabalhos estima a função de produção por MQO. Vale ressaltar que as estimativas obtidas por GMM se encontram abaixo da média nos dois casos.

Hall, Mairesse e Mohnen (2010) notam, ainda, que as estimativas com dados de séries temporais não são estatisticamente significantes em geral. A estimação sem restrição

de retornos constantes de escala também tende a produzir estimativas menores da elasticidade de P&D. Por fim, os autores apontam evidências de que a inclusão de *dummies* individuais, para controlar diferenças fixas entre os setores, também leva a elasticidades menores.

Wieser (2005) realiza uma meta-análise de vários estudos, também sobre firmas e indústrias. Os trabalhos analisados novamente utilizam dados dos Estados Unidos e Europa majoritariamente. Os dados utilizados cobrem o período de 1948 a 1997, se concentrando entre 1977 e 1994. A maior parte dos estudos estima a função de produção por MQO com as variáveis em nível, taxa de crescimento ou desvios das médias individuais.

O autor encontra estimativas da elasticidade de P&D entre -0.016 e 0.38, com uma mediana de 0.1. Também são encontradas evidências de que a inclusão de *dummies* individuais (firmas ou setores) leva a estimativas menores. Finalmente, Wieser (2005) nota que a elasticidade de P&D tende a variar por setor, sendo mais alta, em geral, nos setores mais intensivos em tecnologia.

Os trabalhos discutidos por Hall, Mairesse e Mohnen (2010) e Wieser (2005) não consideram a possível dependência *cross-section* na função de produção. 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. Para avaliar o ganho metodológico da abordagem de fatores comuns, os autores ainda estimam a função de produção com outros estimadores que não consideram a presença de fatores comuns latentes.

Os resultados obtidos por Eberhardt, Helmers e Strauss (2013) indicam que, quando a presença de fatores comuns é controlada, a elasticidade de P&D é menor e não-significante. Isso mostra, segundo os autores, a dificuldade de se separar o retorno privado de P&D do efeito de *spillovers*. Além disso, os resultados indicam que considerar a presença de fatores comuns leva ao controle da não-estacionariedade e dependência *cross-section* nos dados. Finalmente, as estimativas indicam ainda que impôr homogeneidade de parâmetros na estimação tende a distorcer as elasticidades estimadas.

Poucos estudos estimam a elasticidade de P&D no Brasil. Cavalcante, Jacinto e Negri (2015) analisam dados de firmas das indústrias extrativa e de transformação no Brasil entre os anos de 2000 e 2008. As estimações são feitas por MQO para o ano de 2008 individualmente e por efeitos aleatórios para a amostra completa. Todas as estimações incluem *dummies* setoriais. Os autores reportam valores entre 0.06 e 0.1, sendo este valor maior para os setores mais intensivos em tecnologia.

Santana (2017) também analisa dados de firmas das indústrias extrativa e de transformação, mas para o período de 2006 a 2013. Foi utilizado o estimador de efeitos fixos individuais e a função de produção foi estimada com e sem a inclusão de *dummies* de tempo. A autora estima que a elasticidade se encontra entre 0.016 e 0.169, com média em torno de 0.07. Além disso, as estimativas obtidas com a inclusão das *dummies* de tempo foram menores em todos os casos. Este último resultado pode ser visto como evidência da influência de fatores comuns na estimação.

Não foram encontrados trabalhos que apliquem a abordagem de fatores comuns não-observados na função de produção para a estimação da elasticidade de P&D no Brasil. Este trabalho apresenta, assim, uma contribuição metodológica na literatura econômica brasileira, por meio da utilização de um modelo empírico mais robusto para estimar a elasticidade de P&D e pela obtenção de estimativas mais confiáveis para este parâmetro.

2.2.2 Spillovers e PTF

Vale discutir, por fim, os resultados na literatura sobre a relação entre *spillovers*, produtividade e investimento em P&D. Assim, é possível justificar a preocupação com uma correlação entre estas variáveis. Além disso, é possível avaliar a hipótese de que o *spillover* é um fator associado à PTF.

Duguet (2006) analisa dados de firmas da indústria francesa entre 1986 e 1990. O autor encontra evidências de que inovações maiores, como a introdução de novos processos produtivos ou produtos inéditos no mercado, estão associadas a maiores taxas de crescimento da PTF. Além disso, os resultados do estudo mostram que este tipo de inovação depende muito do esforço interno de P&D e, mais ainda, do acesso ao conhecimento

externo, através do uso de patentes ou licenças produzidas em outros setores, por exemplo.

Analisando dados de setores da indústria britânica entre 1970 e 1997, Higon (2007) estima uma função de produção incluindo o investimento interno em pesquisa bem como uma medida de *spillover*. A autora também encontra evidência de que o crescimento da PTF está associado ao investimento interno em P&D e à existência de *spillovers*. Adicionalmente, os resultados indicam que negligenciar a heterogeneidade dos parâmetros e a persistência temporal dos dados produz estimativas distorcidas das elasticidades.

Steingraber e Gonçalves (2010) avaliam o caráter sistêmico da inovação nas indústrias extrativa e de transformação no Brasil. Com dados de firmas dos setores de cada indústria para o ano de 2005, os autores encontram evidência de que diferenças na PTF entre as firmas estão associadas a características dos setores em que elas atuam. Particularmente, os resultados apontam para a importância da cooperação com outras empresas, centros de pesquisa, entre outros, para explicar o desempenho individual das firmas.

Os estudos discutidos nesta seção indicam que é razoável modelar os *spillovers* como fatores associados à PTF, como proposto por Eberhardt e Teal (2010). Além disso, fica também evidente a relação entre o crescimento da PTF e o investimento das firmas em inovação. Assim, se justifica a preocupação de controlar o efeito da PTF e *spillovers* na estimação da elasticidade de P&D.

Por último, é importante ressaltar a evolução da PTF no período analisado neste trabalho no Brasil. Analisando dados macroeconômicos, Filho, Pessoa e Veloso (2010), apontam uma trajetória de crescimento da PTF brasileira entre 2003 e 2007. As estimativas de Ellery (2014), também com dados agregados, corroboram este resultado.

Os resultados de Santana (2017), que utiliza dados de firmas das indústrias extrativa e de transformação, seguem o mesmo padrão das estimativas acima. Além disso, o estudo evidencia uma continuação da trajetória de crescimento até 2011. Finalmente, a autora observa uma queda na PTF entre 2011 e 2013.

3 Metodologia

Existem, segundo Eberhardt e Teal (2011), três formas mais recorrentes na literatura de se controlar a presença de fatores comuns não-observados em modelos de dados em painel. A mais simples seria a inclusão de *dummies* individuais e de tempo na regressão. Coakley, Fuertes e Smith (2006) notam, porém, que tal solução implica assumir que o impacto dos fatores comuns é o mesmo para todos os setores. Como discutido acima, tal hipótese, além de não parecer tão justificável conceitualmente, pode levar a um problema de regressão espúria caso as variáveis analisadas sejam não-estacionárias.

A segunda solução consiste em modelar explicitamente a dependência *cross-section* dos dados. No caso da função de produção, Eberhardt e Teal (2011) argumentam que esta solução requer que se façam hipóteses sobre como setores estão próximos ou distantes entre si para medir seu grau de dependência, distância essa que pode depender de fatores comerciais, institucionais *etc.* Segundo Eberhardt, Helmers e Strauss (2013), tal solução depende de hipóteses *ad hoc* sobre como se dão as relações entre os setores e requer que se conheça exatamente quais são os fatores associados a esta dependência, o que, em geral, não é possível.

A última solução apresentada por Eberhardt e Teal (2011), que será seguida neste trabalho, consiste em utilizar estimadores de dados em painel que assumam explicitamente uma estrutura como a descrita pelas equações (2.7)-(2.10), buscando controlar a presença de fatores comuns sem impôr restrições muito fortes sobre como se dá a dependência gerada por estes.

Estimadores nesta linha, em geral, consistem em aumentar a regressão original com alguma *proxy* dos fatores comuns, como proposto por Eberhardt e Teal (2010) e Pesaran (2006), ou em estimar conjuntamente os parâmetros da regressão e os fatores comuns, como proposto por Bai (2009), entre outros. Estes estimadores são discutidos com mais detalhe neste capítulo, em especial o de Eberhardt e Teal (2010), que será utilizado neste estudo.

3.1 Estimadores de Painel com Fatores Comuns Não-Observáveis

Como discutido no capítulo 2, as relações estruturais implícitas à função de produção serão levadas em consideração neste estudo por meio da estrutura analítica proposta por Eberhardt e Teal (2010):

$$y_{it} = \mathbf{B}'_i \mathbf{x}_{it} + \eta_i + \mathbf{\Lambda}'_i \mathbf{f}_t + u_{it}, \quad (3.1)$$

$$\mathbf{x}_{it} = \mathbf{\Pi}_i \mathbf{g}_t + \mathbf{\Psi}_i \mathbf{f}_t + v_{it}, \quad (3.2)$$

$$\mathbf{f}_t = \mathbf{P}' \mathbf{f}_{t-1} + \epsilon_{it}, \quad (3.3)$$

$$\mathbf{g}_t = \mathbf{K}' \mathbf{g}_{t-1} + \varepsilon_{it}, \quad (3.4)$$

em que i indexa os setores, t representa o ano, y_{it} é o logaritmo do produto, \mathbf{x}_{it} é o vetor de insumos (capital, trabalho e P&D) também em logaritmo, \mathbf{f}_t é o vetor de fatores comuns associados à produtividade (*spillovers*, choques econômicos, efeitos de políticas *etc.*) e $\mathbf{\Pi}_i \mathbf{g}_t$ captura fatores específicos da demanda dos setores pelos insumos, como características dos bens produzidos pelo setor.

O modelo (3.1)-(3.4) foi escolhido por permitir, de forma relativamente simples, considerar a endogeneidade da demanda por insumos através de sua relação com a produtividade e a possibilidade de persistência temporal da produção e das escolhas de insumos. Como discutido por Eberhardt, Helmers e Strauss (2013), tais características são particularmente importantes para a estimação da elasticidade de P&D, dada a relação entre gastos em pesquisa, *spillovers* e choques de produtividade, além da dependência entre a inovação e o conhecimento acumulado ao longo do tempo.

Dada a discussão do início deste capítulo, este estudo fará uso de um estimador que lide explicitamente com a estrutura multifatorial do erro implicada pelo modelo analítico proposto. Para isto, foi escolhido o estimador *augmented mean group* (AMG) de Eberhardt e Teal (2010), que será descrito na próxima seção. Antes, porém, cabe ressaltar as formas mais comuns de se estimar modelos de dados em painel com fatores comuns não-observados.

Uma linha de estimadores deste tipo busca aumentar a regressão original com estimativas dos fatores comuns do erro. Coakley, Fuertes e Smith (2002), por exemplo, propõem estimar primeiro a regressão (2.1) por mínimos quadrados ordinários (MQO) sem controlar os fatores comuns. Realiza-se uma análise de componentes principais do resíduo para estimar os fatores e, finalmente, aumenta-se a regressão original com as estimativas dos fatores.

Para ilustrar o procedimento, considere, um modelo de regressão com uma variável dependente, y_{it} e um regressor, x_{it} :

$$y_{it} = \beta x_{it} + u_{it}, \quad (3.5)$$

$$u_{it} = \mathbf{\Gamma}' \mathbf{f}_t + \epsilon_{it}, \quad (3.6)$$

em que \mathbf{f}_t é um vetor de fatores comuns não-observados e ϵ_{it} é um termo de erro idiossincrático. Estima-se (3.5) por MQO e realiza-se a análise de componentes principais do resíduo, $\hat{u}_{it} = y_{it} - \hat{\beta}x_{it}$. As componentes principais estimadas do resíduo são, então, utilizadas como *proxies* para os fatores comuns, $\hat{\mathbf{f}}_t$. Por fim, estas *proxies* são adicionadas a (3.5), gerando uma nova regressão:

$$y_{it} = \beta x_{it} + \mathbf{\Gamma}' \hat{\mathbf{f}}_t + \epsilon_{it}. \quad (3.7)$$

Pesaran (2006) nota que, em geral, esse estimador não é consistente. Isso porque o resíduo é obtido por MQO, que não é consistente caso os fatores comuns latentes sejam correlacionados com os regressores. Bai (2009) propõe um estimador que consiste basicamente em repetir o procedimento de Coakley, Fuertes e Smith (2002) até convergência das estimativas dos parâmetros. O autor mostra que tal estimador é consistente mesmo sob endogeneidade dos regressores.

Este tipo de abordagem apresenta, segundo Chudik e Pesaran (2013), duas grandes limitações. Primeiro, a análise de componentes principais sobre os resíduos requer a determinação do número de fatores comuns latentes a ser estimados, o que, segundo

os autores, introduz certo grau de incerteza na estimação. Depois, estes métodos assumem, implicitamente, que só existem fatores fortes no erro, ou seja, choques comuns cujo impacto não diminui à medida que o número de indivíduos no painel aumenta. Economicamente, Eberhardt, Helmers e Strauss (2013) notam que fatores comuns fortes representam choques globais, crises financeiras por exemplo, enquanto fatores fracos incluem choques locais como *spillovers* de conhecimento e outras externalidades entre setores próximos. Dessa forma, é importante permitir a presença de fatores fracos na estimação. Por fim, como mostrado por Ahn, Lee e Schmidt (2013), o estimador de Bai (2009) é inconsistente para painéis em que as séries de tempo individuais são pequenas.

Outros estimadores buscam controlar a estrutura fatorial do erro por meio da inclusão de *proxies* para os fatores na regressão. Pesaran (2006), por exemplo, propõe o estimador *common correlated effects* (CCE), que consiste em aumentar a regressão original com médias *cross-section* da variável dependente e dos regressores. A nova regressão pode, então, ser estimada por MQO, caso se assuma parâmetros homogêneos, ou pelo estimador *mean group* (MG) de Pesaran e Smith (1995), caso seja assumida heterogeneidade no modelo¹.

Para entender como os fatores comuns podem ser aproximados por uma combinação linear das médias *cross-section* das variáveis, considere, como em Eberhardt, Helmers e Strauss (2013), um modelo de parâmetros heterogêneos com uma variável dependente, um regressor e, por simplicidade, apenas um fator comum no erro:

$$y_{it} = \beta_i x_{it} + u_{it}, \quad (3.8)$$

$$u_{it} = \gamma_i f_t + \epsilon_{it}. \quad (3.9)$$

Das equações (3.8) e (3.9), segue que

$$\bar{y}_t = \bar{\beta} \bar{x}_t + \bar{\gamma} f_t, \quad (3.10)$$

¹ O estimador MG consiste em estimar regressões individuais para cada unidade do painel e obter uma estimativa da média dos parâmetros por meio da média das estimativas individuais

em que $\bar{y}_t = N^{-1} \sum_{i=1}^N y_{it}$ é a média *cross-section* de y . Isolando f_t em (3.10), tem-se

$$f_t = \pi_1 \bar{y}_t + \pi_2 \bar{x}_t, \quad (3.11)$$

em que $\pi_1 = 1/\bar{\gamma}$ e $\pi_2 = -\bar{\beta}/\bar{\gamma}$. Chudik, Pesaran e Tosetti (2011) mostram que esta relação entre o fator comum e as médias *cross-section* de y e x vale mesmo no caso em que o erro é composto por um número infinito de fatores fracos e finito de fatores fortes². Apenas no caso em que há um número infinito de fatores latentes fortes os autores mostram que tal representação não é válida, uma vez que a variância do erro neste caso não é limitada.

Finalmente, dada a expressão para f_t , (3.8) e (3.9) implicam que

$$y_{it} = \beta_i x_{it} + \delta_{1i} \bar{y}_t + \delta_{2i} \bar{x}_t + \epsilon_{it}, \quad (3.12)$$

em que $\delta_{1i} = \gamma_i \pi_1$ e $\delta_{2i} = \gamma_i \pi_2$. Vale notar que os parâmetros associados às médias *cross-section* em (3.12) são compostos por uma combinação dos parâmetros desconhecidos e suas médias. Ou seja, diferente dos estimadores de componentes principais, em que são obtidas estimativas explícitas para os fatores comuns e seus parâmetros, o CCE não permite identificar os fatores e seus coeficientes. Neste sentido, Eberhardt, Helmers e Strauss (2013) argumentam que não seria razoável interpretar os parâmetros das médias *cross-section*. Segundo os autores, estes termos devem ser vistos meramente como controle para a dependência *cross-section* nos dados e não representativos de algum choque específico por exemplo.

Uma vantagem do CCE é que não requer a determinação do número de fatores comuns no erro. Chudik, Pesaran e Tosetti (2011) demonstram ainda que o estimador é consistente na presença de um número fixo de fatores fortes e infinitos fatores fracos. Os autores também apresentam uma série de simulações de Monte Carlo para avaliar

² Os autores definem um fator f_{it} como fraco se $\lim_{N \rightarrow \infty} \sum_{i=1}^N |\gamma_{il}| = K < \infty$ e como forte se $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N |\gamma_{il}| = K > 0$, em que γ_{il} é a carga do fator f_{it}

a performance do estimador. Os resultados indicam que o viés e erro quadrático médio (RMSE) do CCE diminuem à medida que o tamanho da amostra aumenta.

Kapetanios, Pesaran e Yamagata (2011) mostram que o CCE é consistente mesmo sob não-estacionariedade dos fatores. As propriedades em amostras finitas do estimador também são analisadas via simulações de Monte Carlo. Novamente, o viés e RMSE do CCE são decrescentes no tamanho da amostra, mesmo sob não-estacionariedade dos fatores. Finalmente, como em Chudik, Pesaran e Tosetti (2011), o CCE apresenta boa performance tanto quando se assumem parâmetros homogêneos quanto heterogêneos.

Por outro lado, Eberhardt e Teal (2008) ressaltam que o CCE é relativamente intensivo no uso de dados. Isto é, à medida que o número de regressores cresce, aumenta também a quantidade de médias *cross-section* para se incluir na regressão. Em painéis cujas séries de tempo individuais são curtas, isso pode implicar uma perda significativa dos graus de liberdade, tornando a estimação menos precisa.

O estimador AMG de Eberhardt e Teal (2008) consiste, em um primeiro estágio, na obtenção de uma estimativa da evolução média dos fatores comuns e, em um segundo estágio, na estimação da função de produção aumentada com essa estimativa. Eberhardt e Bond (2009) mostram por meio de simulações que o AMG, em geral, controla a dependência *cross-section* e não-estacionariedade trazida pelos fatores comuns do erro, tendo performance relativamente similar ao CCE.

Eberhardt e Teal (2008) estimam uma função de produção com dados industriais de países desenvolvidos e em desenvolvimento entre os anos de 1970 e 2002. Os resultados indicam que AMG e CCE apresentam performances próximas. Além disso, ambos controlam a dependência *cross-section*, não-estacionariedade e heterogeneidade dos dados. Dessa forma, as estimativas produzidas pelos autores corroboram os resultados das simulações de Eberhardt e Bond (2009).

Finalmente, Eberhardt e Teal (2010) argumentam que a estimativa da evolução dos fatores comuns obtida no primeiro estágio pode ter interpretação econômica a depender do que se supõe estar contido nos fatores comuns. No caso da função de produção, por

exemplo, os autores interpretam o processo dinâmico comum como uma estimativa da evolução média da produtividade total dos fatores (PTF).

Em relação aos estimadores discutidos anteriormente, o AMG permite controlar a presença de fatores comuns no erro sem demandar que o painel analisado contenha séries de tempo individuais muito longas, uma vez que adiciona apenas uma variável à regressão original. Como será discutido no capítulo 4, esta é uma preocupação relevante neste trabalho, uma vez que os dados utilizados contém séries individuais relativamente curtas (11-15 anos). Além disso, a interpretação do processo dinâmico comum estimado pelo AMG como medida da evolução da PTF permite avaliar as hipóteses feitas na construção do modelo analítico sobre o efeito de choques de produtividade. Dadas as características da base de dados analisada e a discussão conceitual sobre a modelagem da PTF apresentada até aqui, foi considerado que o AMG é o estimador mais adequado para este trabalho. O estimador está descrito com mais detalhe na próxima seção.

3.2 Estratégia de Estimação

3.2.1 Estimador AMG

O AMG consiste em um procedimento de dois estágios. No primeiro, estima-se a função de produção por primeiras diferenças incluindo *dummies* de tempo:

$$\Delta y_{it} = \mathbf{B}' \Delta \mathbf{x}_{it} + \sum_{t=2}^T c_t \Delta D_t + \varepsilon_{it}, \quad (3.13)$$

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 de um processo dinâmico comum, que captaria a evolução média dos fatores comuns. No segundo estágio, \hat{c}_t é incluído como um regressor adicional na

função de produção:

$$y_{it} = \mathbf{B}'\mathbf{x}_{it} + \delta_i\hat{c}_t + d_it + \varepsilon_{it}, \quad (3.14)$$

em que t é uma tendência determinística, incluída para controlar possíveis efeitos temporais não captados pelo processo dinâmico comum. Estima-se então (3.14) pelo estimador MG de Pesaran e Smith (1995), que, como discutido acima, assume parâmetros heterogêneos.

Segundo Eberhardt e Teal (2010), como o processo comum é estimado por primeiras diferenças, obtém-se uma estimativa consistente da evolução dos fatores comuns mesmo que estes sejam não-estacionários. Assim, o AMG é capaz de controlar a presença de fatores comuns no erro e sua possível não-estacionariedade por meio da estimação consistente do processo comum.

Outra hipótese assumida por este estimador é que os fatores comuns fazem parte da relação de cointegração entre as variáveis. 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.

3.2.2 Cálculo da PTF

Eberhardt e Teal (2010) apontam outra vantagem do AMG. Os autores mostram como as estimativas produzidas pelo estimador possibilitam um cálculo da PTF mais robusto que os métodos mais tradicionais. Para entender como, vale discutir brevemente as formas de cálculo da PTF mais utilizadas e compará-las com o cálculo a partir das estimativas do AMG³.

³ A discussão sobre os métodos tradicionais de cálculo da PTF segue, em linhas gerais, Beveren (2012).

Considere a seguinte função de produção *Cobb-Douglas* em logaritmo:

$$y_{it} = \alpha k_{it} + \beta l_{it} + \gamma r_{it} + a_{it} + u_{it}, \quad (3.15)$$

em que $a_{it} = \log(A_{it})$ e a notação é a mesma adotada no capítulo 2.

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:

$$a_{it} = y_{it} - \hat{\alpha}k_{it} - \hat{\beta}l_{it} - \hat{\gamma}r_{it}. \quad (3.16)$$

O autor nota que uma das maiores preocupações ao se estimar funções de produção é a simultaneidade entre a escolha de insumos e a produtividade. Quando não controlado, este efeito leva à endogeneidade dos insumos na regressão, produzindo viés nas estimativas. Da equação (3.16) é possível notar que estimativas viesadas dos parâmetros implicam também viés no cálculo da PTF pelo resíduo.

Muitos trabalhos tentam controlar a endogeneidade dos insumos incluindo *dummies* individuais e de tempo na regressão original. Como discutido no capítulo 2, uma das principais limitações desta solução é a hipótese de que o componente temporal comum da PTF afeta os setores de forma igual. Caso tal hipótese não se verifique, a *dummy* de tempo não será suficiente para capturar o componente dinâmico da PTF, de modo que parte do efeito da produtividade ainda não será controlada na estimação. Assim, as elasticidades estimadas ainda terão viés.

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 dos insumos. Por outro lado, ainda se assume que a evolução da PTF é comum e estacionária. O *system-GMM* é descrito com mais detalhe na próxima seção.

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 (3.14), Eberhardt e Teal (2010) propõem calcular a evolução da PTF de cada setor da seguinte forma:

$$a_{i,0} = \hat{\alpha}_i k_{i,0} + \hat{\beta}_i l_{i,0} + \hat{\gamma}_i r_{i,0}, \quad (3.17)$$

$$a_{i,\tau} = a_{i,0} + \hat{\delta}_i \hat{c}_\tau + \hat{d}_i \tau, \quad (3.18)$$

em que $\tau = 1, \dots, T$ é o período em questão, \hat{c}_t 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 do parâmetro da tendência determinística incluída no segundo estágio.

O cálculo proposto difere do tradicional, descrito pela equação (3.16) de duas formas. Primeiro, é usada a tendência comum estimada pelo AMG para calcular a evolução da PTF, como mostra a equação (3.18). Além disso, o método proposto também difere com relação ao cálculo do nível inicial da PTF. Como discutido por Eberhardt e Teal (2010), o cálculo da PTF por meio do resíduo, como na equação (3.16), toma o intercepto da regressão como nível inicial da PTF. O cálculo a partir do AMG leva em conta, adicionalmente, o nível dos insumos no período inicial, como pode ser visto na equação (3.18).

A intuição por trás disso é a seguinte. Mesmo que determinados setores tenham características individuais parecidas, eles ainda podem ter níveis de PTF diferentes se a produtividade de cada fator for diferente entre setores. Assim, a medida de PTF sem

levar em conta a heterogeneidade das elasticidades dos fatores estaria incompleta. Nesse caso, o cálculo do nível inicial da PTF por (3.18) permite capturar diferenças nos níveis da PTF dos setores de forma mais completa que só um efeito fixo individual permitiria.

Com base na discussão acima, este trabalho também utilizará a PTF calculada a partir do AMG para avaliar a qualidade do estimador. Como discutido no capítulo 2, é razoável supor que choques de produtividade estejam associados, por exemplo, ao conhecimento acumulado pelos setores. Adicionalmente, é possível que exista uma correlação positiva entre PTF e orientação exportadora, uma vez que, como discutido por Kannebley (2011), firmas mais produtivas tendem a se engajar mais em atividades de exportação. No caso do Brasil, Bonelli (1992), por exemplo, aponta para a existência de tal correlação para as indústrias extrativa e de transformação no período de 1975 a 1985. Assim, para avaliar a qualidade da PTF calculada, será analisada a correlação desta variável com dados de exportação e intensidade tecnológica dos setores.

3.2.3 Estimações Adicionais

Para avaliar a validade das hipóteses assumidas no AMG sobre o processo gerador de dados, a função de produção será estimada também com outros estimadores mais utilizados, que não assumem uma estrutura multifatorial heterogênea do erro. O primeiro deles é o estimador de MQO aumentado com *dummies* de tempo. Como discutido acima, a inclusão de *dummies* temporais para controlar o efeito de choques comuns implica assumir que o impacto destes choques é homogêneo entre os setores. Assim, o processo gerador de dados assumido nesta estimação é:

$$y_{it} = \mathbf{B}\mathbf{x}_{it} + \theta_t + \varepsilon_{it}, \quad (3.19)$$

em que θ_t é um choque comum e ε_{it} é um erro idiossincrático. Este modelo pode ser visto como um caso particular da estrutura assumida neste trabalho, descrita em (3.1)-(3.4), no qual se impõe homogeneidade dos parâmetros e ausência de dependência temporal das variáveis e fatores comuns.

O segundo estimador adicional é o de efeitos fixos *twoway* (EF) que consiste em adicionar um efeito individual invariante no tempo à equação (3.19):

$$y_{it} = \mathbf{B}\mathbf{x}_{it} + \theta_t + \alpha_i + \varepsilon_{it}. \quad (3.20)$$

Economicamente, o modelo do MQO implica assumir que a produtividade dos setores é a mesma, captada pelo termo θ_t . Já o EF assume que há uma evolução comum da produtividade, dada por θ_t , mas que esta pode diferir entre os setores, efeito capturado pelo efeito fixo, η_i . Como discutido anteriormente, uma vez que estes estimadores assumem ausência de dependência temporal, na presença de raiz unitária das variáveis e da produtividade estes estimadores sofrerão do problema de regressão espúria. Além disso, caso o processo verdadeiro seja de parâmetros heterogêneos, MQO e EF também não controlarão a dependência *cross-section* gerada pela presença de fatores comuns mesmo com a inclusão da *dummy* de tempo, como explicado no capítulo 2.

O estimador de primeiras diferenças (PD) também será utilizado. Este método consiste em tomar as variáveis no modelo (3.20) em primeiras diferenças para eliminar o efeito fixo individual α_i , corrigindo uma possível endogeneidade gerada caso haja correlação deste efeito com os regressores:

$$\Delta y_{it} = \mathbf{B}'\Delta\mathbf{x}_{it} + \Delta\theta_t + \Delta\varepsilon_{it}. \quad (3.21)$$

Além de controlar o efeito fixo individual, o estimador PD é, em geral, consistente mesmo sob não-estacionariedade das variáveis. Isto porque, como discutido em Hendry (1995) e Jones (1995), séries econômicas de valores reais ⁴, tendem a ser integradas de primeira ordem (I(1)), ou seja, são séries com raiz unitárias mas que se tornam estacionárias em primeiras diferenças. Espera-se, assim, que este estimador não sofra do problema de regressão espúria. Por outro lado, como no caso do MQO e EF, caso os parâmetros verdadeiros sejam heterogêneos, o estimador PD não controlará a dependência *cross-section* gerada pelos fatores comuns.

⁴ Séries não afetadas por variações de preço como trabalho ou produto e capital quando deflacionados.

O último estimador adicional utilizado será o *system*-GMM de Blundell e Bond (1998). Para este estimador, assume-se explicitamente uma dependência temporal no erro:

$$y_{it} = \mathbf{B}'\mathbf{x}_{it} + \theta_t + \alpha_i + \varepsilon_{it}, \quad (3.22)$$

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \epsilon_{it}, \quad (3.23)$$

em que ϵ_{it} é um termo de erro idiossincrático. (3.22) e (3.23) implicam que

$$y_{it} = \rho y_{i,t-1} + \mathbf{B}'\mathbf{x}_{it} + \mathbf{\Pi}'\mathbf{x}_{i,t-1} + \theta_t^* + \alpha_i^* + \epsilon_{it}, \quad (3.24)$$

em que $\mathbf{\Pi}' = -\rho\mathbf{B}'$, $\theta_t^* = \theta_t - \rho\theta_{t-1}$, $\alpha_i^* = (1 - \rho)\alpha_i$ ⁵. A restrição imposta por $\mathbf{\Pi}$ pode ser testada, como proposto por Blundell e Bond (2000), para avaliar a validade da representação em (3.22)-(3.23).

Como α_i está correlacionado com y em todos os períodos, $y_{i,t-1}$ é endógeno em (3.24). Para eliminar o efeito fixo, as variáveis são tomadas em primeira diferença para a regressão

$$\Delta y_{it} = \rho\Delta y_{i,t-1} + \mathbf{B}\Delta\mathbf{x}_{it} + \mathbf{\Pi}'\Delta\mathbf{x}_{i,t-1} + \Delta\theta_t^* + \Delta\epsilon_{it}. \quad (3.25)$$

Porém, como $y_{i,t-1}$ depende de θ_{t-1} e $\epsilon_{i,t-1}$ por definição, $\Delta y_{i,t-1}$ em (3.25) também é endógeno. O estimador *system*-GMM consiste em estimar (3.25) utilizando defasagens mais longas de y_{it} (de $y_{i,t-2}$ para trás) como instrumentos para $\Delta y_{i,t-1}$ e diferenças mais longas de y_{it} (de $\Delta y_{i,t-2}$ para trás) como instrumentos para os níveis de y_{it} , uma vez que estas variáveis não estariam correlacionadas com os erros em t e $t-1$. São incluídas também *dummies* de tempo para controlar o efeito de θ_t^* . Assim, o *system*-GMM controla a possível não-estacionariedade do modelo. Porém, o *system*-GMM assume homogeneidade dos parâmetros. Caso esta hipótese não se verifique, o estimador não controla a endogeneidade causada pela dependência *cross-section*. Isso porque, como mostrado por Pesaran e Smith

⁵ Dadas as estimativas de $\mathbf{B}\epsilon\mathbf{\Pi}$ os coeficientes de longo-prazo podem ser calculados. No caso do P&D, por exemplo, o coeficiente é dado por $\beta_r^{lp} = \frac{\beta_r + \pi_r}{1 + \rho}$

(1995), no caso de painéis dinâmicos, a imposição equivocada de homogeneidade dos parâmetros faz com que não exista instrumento que satisfaça as condições de relevância e exogeneidade simultaneamente⁶.

Espera-se que, caso as hipóteses assumidas pelo AMG sejam válidas, este seja o único estimador que controle tanto a dependência *cross-section* quanto a não-estacionariedade nos erros da função de produção. Neste caso, como explicado acima, espera-se que MQO e EF não controlem nenhuma das duas características enquanto PD e *system*-GMM controlem apenas a não-estacionariedade.

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 viés quando os dados apresentam dependência *cross-section*, heterogeneidade e não-estacionariedade. O estimador de primeiras diferenças, quando incluída *dummy* de tempo, não apresenta viés. O AMG é o que apresenta estimativas mais próximas dos parâmetros verdadeiros em geral. Finalmente, os autores apresentam evidência de que o AMG tem uma variância menor que os outros estimadores.

3.2.4 Testes dos Resíduos

Para avaliar a presença de dependência *cross-section* e não-estacionariedade nos modelos descritos acima, serão realizados dois testes sobre os resíduos produzidos por cada estimador. Para a primeira propriedade, será aplicado o teste CD de Pesaran (2004). O teste, cuja hipótese nula é de ausência de dependência *cross-section*, consiste basicamente em analisar a correlação amostral entre as unidades do painel. Para uma dada variável, a estatística de teste é

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right), \quad (3.26)$$

em que $\hat{\rho}_{ij}$ é a correlação amostral da variável analisada entre as unidades i e j .

Quanto mais distante a correlação média estiver de zero, mais fácil de se rejeitar a hipótese

⁶ Pela mesma lógica da relação entre imposição de homogeneidade e regressão espúria explicada no capítulo 2

nula de independência *cross-section*.

Já para avaliar a não-estacionariedade dos resíduos, será aplicado o teste CIPS de Pesaran (2007). A vantagem deste método é que ele é válido mesmo na presença de fatores comuns não-observados. 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:

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + d_i \Delta \bar{y}_t + e_{it} \quad (3.27)$$

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.

Com base nas discussões das seções anteriores, espera-se que, caso as hipóteses assumidas pelo AMG estejam corretas, os resíduos do AMG, PD e *system*-GMM rejeitem a hipótese nula do teste CIPS, enquanto MQO e EF não a rejeitariam. Além disso, espera-se que os resíduos do AMG sejam os únicos que não rejeitem a hipótese nula do teste CD.

4 Base de Dados

Para estimar a função de produção, é necessário construir as variáveis de produto, trabalho, estoque de capital e gasto em P&D. Além disso, também será avaliada a relação entre a estimativa da PTF produzida como resíduo da função de produção e a orientação exportadora dos setores, que será medida pelo volume de exportações. Assim, também são necessários dados de exportação. Este capítulo descreve a construção da base de dados e das variáveis, apontando as principais dificuldades encontradas neste processo. Por fim, são apresentadas estatísticas descritivas para uma análise prévia das hipóteses levantadas ao longo do trabalho.

Serão analisados dados dos setores das indústrias extrativa e de transformação brasileiras, agregados a um nível de 3 dígitos da Classificação Nacional de Atividade Econômica (CNAE), entre os anos de 2003 e 2017. As variáveis nominais da PIA foram deflacionadas utilizando o Índice de Preços por Atacado (IPA) setorial da Fundação Getúlio Vargas (FGV). Os dados de exportação foram deflacionados pelo índice de preços das exportações brasileiras da Fundação Centro de Estudos do Comércio Exterior (Funcex).

4.1 Fontes de Dados

4.1.1 Pesquisa Industrial Anual (PIA)

A Pesquisa Industrial Anual (PIA) do IBGE tem por objetivo apresentar propriedades estruturais e temporais da atividade industrial no Brasil. A PIA é uma pesquisa amostral, sendo sua amostragem composta por empresas com pelo menos uma pessoa ocupada em 31 de dezembro do ano de referência e cuja atividade principal é descrita pelas categorias B e C (indústrias extrativa e de transformação respectivamente) da Classificação Nacional de Atividades Econômicas (CNAE 2.0). A amostra da PIA é dividida em dois estratos:

- estrato certo, composto por empresas com 30 ou mais pessoas ocupadas ou que obtiveram receita bruta da venda de produtos e serviços industriais superior a R\$ 13,6 milhões no ano anterior ao da pesquisa;
- estrato aleatório, composto por empresas com menos de 30 pessoas ocupadas.

Desta pesquisa, serão utilizadas as variáveis de Valor de Transformação Industrial (VTI) como medida de valor adicionado para a construção da variável de produto, pessoal ocupado para a construção da variável de trabalho e 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, serão considerados os dados da PIA para empresas com 5 ou mais pessoas ocupadas. Essa informação é disponibilizada para os setores divididos em grupos da CNAE, categoria que corresponde a um nível de agregação de 3 dígitos do código.

Ao trabalhar com os dados agregados da PIA, foram encontrados dois grandes problemas. O primeiro diz respeito à classificação dos setores analisados, que mudou a partir de 2007, e o segundo é a quantidade de missing nos dados. Estas questões são discutidas com detalhe a seguir.

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 1 . 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 será 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 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:

$$VTI_{121} = 0.83 \times VTI_{160} \quad (4.1)$$

$$VTI_{122} = 0.16 \times VTI_{160} \quad (4.2)$$

$$VTI_{131} = 0.78 \times VTI_{171} + 1 \times VTI_{172}. \quad (4.3)$$

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 observadas¹. Os resultados indicam que não há

¹ Os grupos utilizados no teste local foram definidos em Lucinda e Meyer (2013)

diferença de média entre os valores reais e os calculados com o dicionário. As tabelas 10 e 11 em anexo apresentam estes resultados. Adicionalmente, foram comparadas a distribuição real das variáveis e a calculada pelo dicionário. Em geral, as distribuições são bastante próximas. A figura 6, também em anexo, apresenta esta comparação.

Os resultados indicam que o dicionário realiza uma boa compatibilização entre CNAE 1.0 e 2.0. Assim, os dados antes de 2007 serão convertidos para a CNAE 2.0 e incluídos na análise. O dicionário completo também é apresentado em anexo.

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 1 apresenta o número de *missing* por setor da CNAE 2.0 (dados entre 2007 e 2017) para cada variável que será utilizada da PIA (A, M e B são aquisições, melhorias e baixas de ativo respectivamente)².

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 2 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

² Uma tabela com os setores da CNAE 1.0 (dados entre 1996 e 2007) que apresentam *missing* é apresentada em anexo

Tabela 1 – Ocorrência de Missing por Setor - CNAE 2.0

CNAE	VTI	Pessoal	Salarios	A	M	B
050	0	0	0	0	5	0
060	0	0	0	0	2	0
099	0	0	0	3	8	3
183	0	0	0	0	0	1
191	0	0	0	1	7	2
204	0	0	0	0	1	0
255	0	0	0	0	2	0
267	0	0	0	0	1	0
268	0	0	0	0	8	2
303	7	8	8	6	6	6
304	3	3	3	3	3	3
305	11	11	11	10	11	10
322	0	0	0	0	3	1
332	0	0	0	0	1	0

Fonte: Elaboração própria

2002. Sendo assim, este trabalho utilizará 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. A divisão dos setores a partir da classificação de

Tabela 2 – 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

intensidade tecnológica da OCDE está disponível em Cavalcante (2014) ³.

Primeiramente, foi realizada uma comparação gráfica do comportamento das 3 variáveis entre os setores com informação completa e incompleta. Não parece haver diferença 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 restrita aos setores com informação completa parecem ter uma composição parecida dos grupos de intensidade tecnológica. Esta análise gráfica pode ser vista pelas figuras 7 e 8 em anexo.

Para tornar a análise mais precisa, foi estimado um modelo logit usando uma

³ A análise realizada segue, em partes, a metodologia proposta por Alves e Silva (2008)

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 indicam ausência de correlação entre características econômicas e tecnológicas dos setores e o fato de sua informação estar completa. A tabela 14, em anexo, apresenta os resultados da estimação.

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. Foram considerados neste trabalho, então, apenas tais setores, uma vez que estes devem apresentar dados mais confiáveis.

4.1.2 Relação Anual de Informações Sociais (RAIS)

A Relação Anual de Informações Sociais (RAIS) é uma base de dados da Secretaria Especial de Previdência e Trabalho do Ministério da Economia, que fornece informações sobre o mercado de trabalho formal brasileiro. A RAIS traz características dos trabalhadores, como gênero, escolaridade e ocupação, bem como dos estabelecimentos, como número de pessoas ocupadas e natureza jurídica. Desta base, serão utilizados os dados de ocupação dos trabalhadores, 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.

4.1.3 Fundação Centro de Estudos do Comércio Exterior (Funcex)

A Funcex é uma instituição sem fins lucrativos que visa promover estudos técnicos e acadêmicos bem como coletar e divulgar dados sobre comércio exterior no Brasil. Dos dados disponibilizados pela Funcex, será utilizado o valor em dólar das exportações *free on board* (FOB).

4.2 Variáveis

4.2.1 Investimento em P&D

No Brasil, dados sobre inovação, incluindo gastos em P&D, são coletados 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. Seria necessária, então, outra medida para esta variável.

Por outro lado, Hall (1993) aponta evidências de que ao menos 50% dos gastos em P&D correspondem aos salários de profissionais ligados à pesquisa, como cientistas e engenheiros. Araújo, Cavalcante e Alves (2009) argumentam que, além dos salários de pesquisadores serem componente importante dos gastos em P&D, a presença de pesquisadores está associada ao esforço inovativo das firmas. Nesse caso, as firmas contratam mais pesquisadores quando estão envolvidas em mais projetos de P&D.

Nesse sentido, Araújo, Cavalcante e Alves (2009) sugerem que a quantidade de pessoas ocupadas em pesquisa seria uma boa *proxy* para o investimento em P&D. Utilizando dados da RAIS, os autores propõem a construção da variável de pessoal em ocupações técnico-científicas (Potec). O procedimento 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). A tabela 3 apresenta as categorias que compõem a variável de Potec, identificadas pela CBO.

Seguindo 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 correlação estimada entre gasto total em inovação e o Potec foi de 0.84 quando é levada em conta apenas a indústria de transformação e 0.68 para a amostra completa ⁴. A alta correlação encontrada, que corrobora os resultados de Araújo, Cavalcante e Alves (2009), indica que o Potec é uma boa *proxy* para o investimento em

⁴ Tal diferenciação foi considerada porque os dados da indústria extrativa só são divulgados de forma agregada na Pintec, sem distinção de setor

Tabela 3 – Ocupações do Potec

Grupo Ocupacional	CBO
Pesquisadores	203
	202
Engenheiros	214
	222
Diretores e Gerentes de P&D	1237
	1426
	201
	211
Profissionais "Científicos"	212
	213
	221

Fonte: Elaboração própria

P&D. Adicionalmente, os dados de Potec foram comparados com a divisão dos setores com base na intensidade tecnológica, proposta pela OCDE e adaptada para a CNAE por Cavalcante (2014). O resultado desta comparação, apresentado em anexo, indica que a variável de Potec também tem aderência às categorias de intensidade tecnológica.

A principal vantagem do Potec em relação aos dados de inovação da Pintec é que, uma vez que a variável é baseada na RAIS, sua disponibilidade é anual. Assim, este trabalho fará uso da variável de Potec como *proxy* para o gasto em P&D por setor porque, deste modo, não será necessário reduzir a amostra analisada.

Finalmente, 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 pelo autor, o número de pessoas incluídas no Potec por setor foi multiplicado por uma jornada média anual de trabalho de 2112 horas/homem:

$$Potec_{it} = \frac{R_{it} \times 2112}{N_{it}^{RAIS}}, \quad (4.4)$$

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.

4.2.2 Trabalho

Para a série de trabalho foi utilizado o valor médio da variável de pessoal ocupado (PO) por setor da PIA, multiplicado pela jornada média anual de trabalho de 2112 horas/homem, como discutido acima. Para evitar dupla contagem dos profissionais ocupados em pesquisa, o Potec foi subtraído deste valor. Assim, a variável utilizada neste estudo para o trabalho representa pessoal que não desempenha ocupações técnico-científicas (NPotec). Sua construção pode ser descrita como:

$$NPotec_{it} = \frac{PO_{it} \times 2112}{N_{it}^{PIA}} - Potec_{it}, \quad (4.5)$$

em que N_{it}^{PIA} é o número de empresas do setor i amostradas na PIA no período t .

4.2.3 Capital

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%. Para o cálculo do estoque inicial de capital, segue-se Garcia (2003): com base no coeficiente do trabalho na distribuição funcional da renda, $\alpha_i^l = W_{i0}/Y_{i0}$, em que W_{i0} é o total de salários do setor i no período inicial – no caso deste trabalho, o ano de 1996 –, 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

$$K_{i0} = \frac{(Y_{i0})^{\frac{1}{\alpha_i^k}}}{(W_{i0})^{\frac{\alpha_i^l}{\alpha_i^k}}}. \quad (4.6)$$

4.2.4 Produto

Para a série de produto, foi utilizado o valor médio por setor da 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.

4.3 Estatísticas Descritivas

A base final contém informações de 96 setores entre 2003 e 2017, totalizando 1396 observações. A fim de avaliar a relevância das características dos parâmetros e variáveis da função de produção discutidas ao longo deste trabalho, esta seção apresenta uma análise descritiva dos dados utilizados.

A tabela 4 apresenta estatísticas descritivas das variáveis utilizadas. Uma análise do coeficiente de variação (CV), definido como a razão entre desvio-padrão (DP) e a média, indica que a variável de Potec é a que apresenta mais variabilidade.

Tabela 4 – Estatísticas Descritivas

	Média	Mediana	DP	CV	Min	Max
VTI	15.62	15.289	1.776	11.4%	11.585	21.431
NPotec	11.924	11.64	1.141	9.6%	9.279	16.264
Capital	16.759	16.581	1.881	11.2%	12.411	22.519
Potec	7.347	7.292	1.724	23.5%	2.294	12.212

Fonte: Elaboração própria. Valores em logaritmo

A figura 1 mostra a comparação entre setores – agregados em divisões da CNAE, que corresponde aos dois primeiros dígitos do código, para facilitar a visualização – da razão do produto, capital e Potec por hora trabalhada (Y/L, K/L e R/L, respectivamente)⁵.

O número de horas trabalhadas em pesquisa e o produto por hora trabalhada variam bastante entre setores. Mesmo o estoque de capital por trabalhador, que aparenta ser relativamente mais homogêneo, ainda apresenta certa variabilidade. Além disso, é possível notar que o ordenamento dos setores também muda entre os gráficos.

⁵ Valores em logaritmo

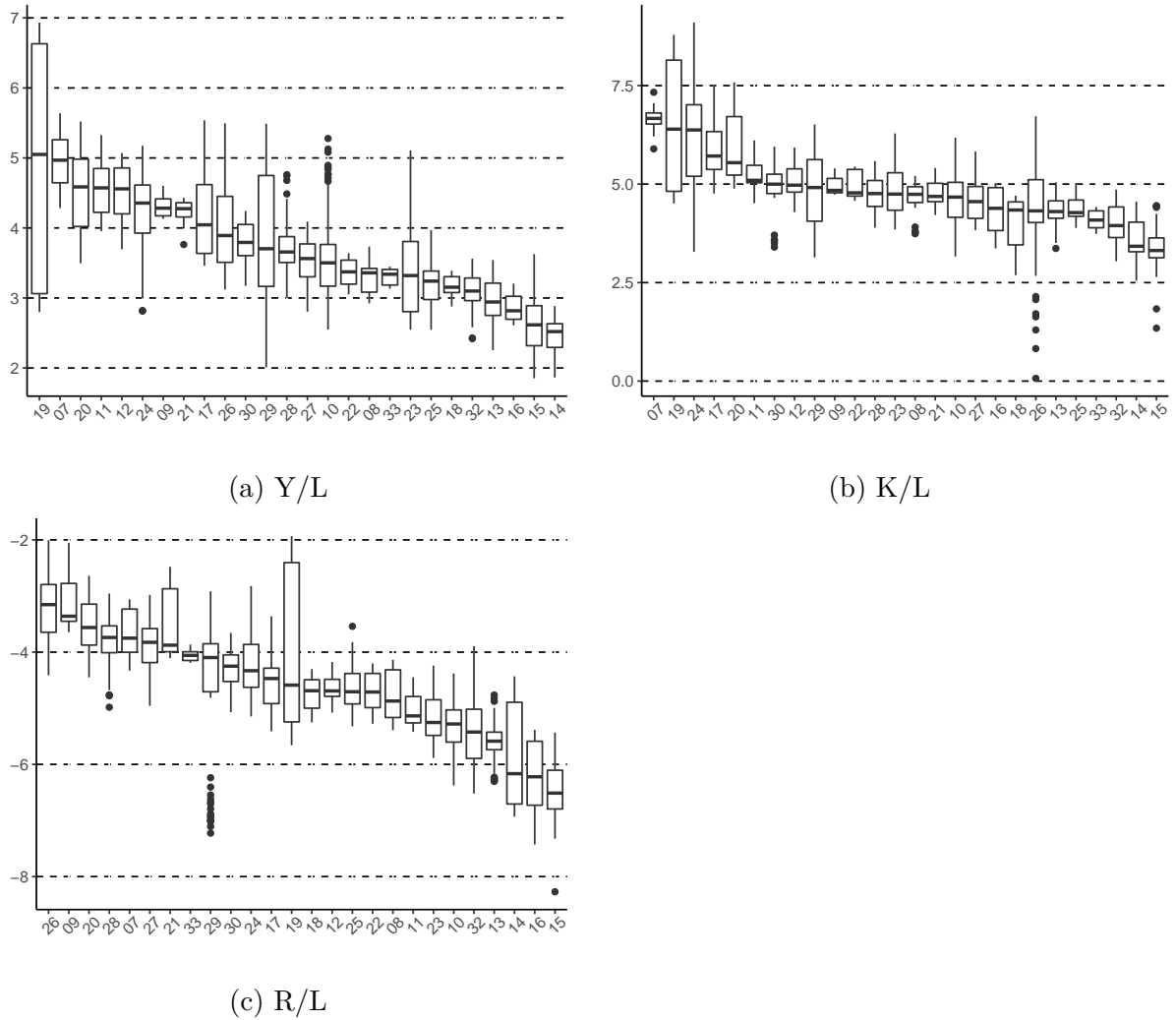


Figura 1 – Produto e Insumos por Hora de Trabalho

O fato dos setores com maior produtividade do trabalho não serem necessariamente os mesmos com maior capital ou P&D por trabalhador pode ser visto como indício da heterogeneidade das elasticidades de cada insumo ou da PTF. Para entender este ponto, considere por simplicidade a função de produção com capital, trabalho e P&D e retornos constantes de escala. Neste caso, tem-se:

$$y_{it} = a + \alpha k_{it} + \gamma r_{it}, \tag{4.7}$$

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. Neste caso, setores com maiores valores de k_{it} e r_{it} deveriam ter valores maiores também de y_{it} .

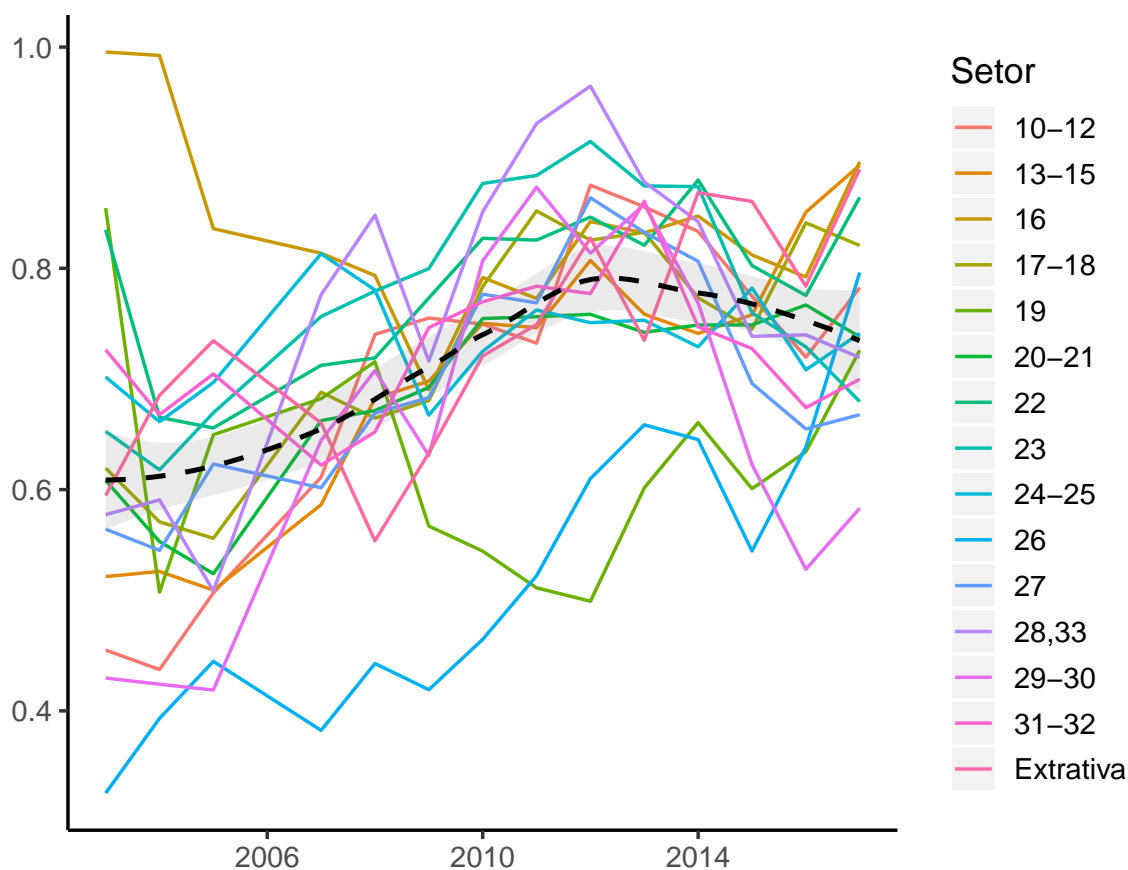


Figura 2 – Evolução Média do Produto por Setor

É possível notar pela figura 1 que este padrão não se observa entre alguns setores. Por exemplo, o setor 07 apresenta níveis maiores de k e r que o 19, enquanto o último é o que apresenta maior y^6 . Pela equação (4.7), para que isso ocorra é necessário que os setores apresentem diferenças na PTF ou nas elasticidades dos insumos.

A figura 2 apresenta a evolução do produto por setor. A fim de facilitar a visualização, os setores foram agrupados por proximidade com base em sua divisão da CNAE⁷ e as médias anuais de cada grupo são reportadas. Além disso, as séries estão expostas como proporção do valor máximo da série, o que evita distorções na escala do gráfico⁸. A descrição dos grupos reportados é apresentada em anexo.

Analisando a figura 2, é possível observar uma tendência temporal comum às séries de crescimento até 2010, relativa estagnação até 2013 e queda a partir de 2014. A linha

⁶ A mesma comparação vale entre os setores 14 e 15 ou 23 e 10 por exemplo.

⁷ Grupos construídos com base em Eberhardt, Helmers e Strauss (2013)

⁸ Para cada setor, tem-se $y_t^* = y_t / \max(y_t)$; são reportadas as médias de y_t^* entre os grupos de setores

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%. A análise das demais variáveis, disponível em anexo, também aponta uma tendência temporal comum entre os setores.

A presença de uma tendência temporal comum aos setores pode ser vista como evidência da influência de fatores comuns não-observados nas séries individuais dos setores. Este padrão aponta para a importância de se levar em conta tal propriedade dos dados na estimação da função de produção. Além disso, o fato das séries apresentarem períodos de crescimento ou queda pode indicar que, pelo menos para alguns setores, é razoável assumir que as variáveis sejam não-estacionárias.

4.3.1 Testes Preliminares

A fim de avaliar a pertinência das hipóteses acerca da função de produção apresentadas no capítulo 2, e complementar a análise descritiva dos dados apresentada nesta seção, foram aplicados os testes CIPS e CD, descritos no capítulo 3, sobre as variáveis construídas. As hipóteses nulas dos testes são, respectivamente, de não-estacionariedade e independência *cross-section* dos dados.

A tabela 5 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 5 – Teste 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

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. Vale notar, ainda, que a hipótese nula do teste CIPS é de que todas as séries do painel são não-estacionárias, de modo que mesmo sua rejeição não garante a estacionariedade das variáveis.

Os testes preliminares indicam que não-estacionariedade e dependência *cross-section* são características presentes nos dados. Estes resultados confirmam a discussão do capítulo 2 e a análise deste capítulo. Assim, justifica-se a estimação da função de produção com um modelo que leve em conta tais características, como o proposto no capítulo 3.

5 Resultados

5.1 Resultados Principais

5.1.1 Estimação Sem P&D

Como descrito no capítulo 3, além do estimador AMG, a função de produção será estimada com os estimadores de MQO, Efeitos Fixos, primeiras diferenças, *system-GMM* e MG. A fim de avaliar como a qualidade da estimação muda entre os estimadores, serão primeiramente reportados os resultados das estimações da função de produção sem incluir a variável de Potec. Assim, quando ela for incluída na estimação, será possível avaliar com mais precisão como os estimadores se comparam em termos da elasticidade de P&D estimada e como os resultados são afetados pela inclusão da variável. Neste caso, foi utilizada a variável de pessoal ocupado total da PIA, sem descontar o Potec. A tabela 6 apresenta os resultados desta primeira estimação^{1,2}.

Tabela 6 – 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

Fonte: Elaboração própria

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 coeficientes estão dentro do intervalo obtido por Lucinda e Meyer (2013) com dados da PIA ao nível de 3 dígitos da CNAE entre 1996 e 2007, que vai de 0.6 a 1.33, com

¹ Para o *system-GMM*, são apresentados os coeficientes de longo-prazo, cujo cálculo está descrito no capítulo 3

² 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

média de 0.97. Particularmente, a estimativa do *system*-GMM deste trabalho se encontra próxima do valor máximo obtido por Lucinda e Meyer (2013).

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 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. Com exceção do MQO, as estimativas são mais baixas que as obtidas por Lucinda e Meyer (2013), que variam entre 0.207 e 0.571, com média de 0.37.

Como esperado, não é possível rejeitar a hipótese de raiz unitária para os resíduos de MQO e efeitos fixos. Os estimadores de primeiras diferenças e *system*-GMM, que modelam a dinâmica dos dados explicitamente, produzem resíduos estacionários. A hipótese de raiz unitária também é rejeitada para os resíduos de MG e AMG.

A comparação entre as estimativas de MQO e EF com o PD e *system*-GMM aponta para o possível viés gerado na estimação ao negligenciar a não-estacionariedade das variáveis. Além disso, a comparação entre MQO e MG evidencia como a imposição equivocada de homogeneidade dos parâmetros pode levar a resultados espúrios. Finalmente, uma comparação do RMSE indica que o AMG é o estimador que melhor se ajusta aos dados.

A hipótese de retornos constantes de escala é rejeitada em todas as especificações. Lucinda e Meyer (2013) encontram este resultado para 6 dos 11 grupos de setores que avaliam. Garcia (2003) também encontra evidência de retornos crescentes de escala em todas as divisões da CNAE³. Vale ressaltar que é possível que este resultado seja consequência da omissão da variável de P&D. Isso porque, se capital e trabalho forem positivamente correlacionados com esta variável, seus coeficientes podem estar capturando parte desta correlação e, por isso, estariam inflados.

Os resultados preliminares não permitem inferir muito acerca do impacto de não levar em conta a dependência *cross-section* das variáveis, uma vez que nenhum dos estimadores controlou esta propriedade. Na próxima subseção, será discutido como essa limitação pode estar relacionada à omissão da variável de P&D.

5.1.2 Estimação com P&D

A discussão acima aponta para as consequências de se negligenciar a dependência *cross-section*, não-estacionariedade e heterogeneidade na estimação da função de produção. Tendo em vista o comportamento dos estimadores utilizados neste trabalho na presença de tais propriedades, procede-se com a estimação da elasticidade de P&D, que

³ Nível de agregação de 2 dígitos

é o principal objetivo deste trabalho. A tabela 7 apresenta os resultados das estimações da função de produção incluindo o Potec.

Tabela 7 – 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

Fonte: Elaboração própria

Após a inclusão do Potec, 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. Os coeficientes ainda estão dentro do intervalo obtido por Lucinda e Meyer (2013), sendo a estimativa do *system*-GMM inclusive mais próxima da média de 0.97 obtida pelos autores.

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. No caso do MG, a elasticidade do capital deixa de ser estatisticamente significativa. Novamente, apenas o MQO produz uma estimativa dentro do intervalo obtido por Lucinda e Meyer (2013) para o capital.

O coeficiente do Potec varia de 0.014 (AMG) a 0.267 (MQO). Os valores obtidos por MQO e *system*-GMM são os que mais se distanciam dos outros, sendo que apenas estes produzem estimativas estatisticamente significantes para o Potec. O restante das estimativas se encontra abaixo dos valores reportados por Cavalcante, Jacinto e Negri (2015), entre 0.06 e 0.1, e Santana (2017), com média de 0.07.

As estimativas das elasticidades de trabalho e capital caíram em quase todos os casos, sendo que apenas a elasticidade do capital no EF se manteve igual ao caso da regressão sem o Potec. Isso indica uma possível correlação dos gastos em P&D com as variáveis de capital e trabalho. Dessa forma, os parâmetros estimados na regressão sem Potec captam parte desta correlação, além do próprio efeito das variáveis sobre o produto.

Com exceção do AMG, os resultados sobre as propriedades dos resíduos são os mesmos. O resíduo do AMG, além de estacionário como antes, não apresenta dependência *cross-section*. Uma possível explicação seria a importância da variável de P&D para explicar a dependência entre os setores. Novamente, o AMG é o estimador que apresenta menor RMSE, indicando melhor ajuste aos dados. A análise dos resíduos corrobora o argumento feito neste trabalho de que o AMG é o estimador mais adequado dentre os analisados para estimar a função de produção.

Os resultados indicam que uma elasticidade alta e significativa do P&D pode ser consequência de ignorar a dependência *cross-section* e não-estacionariedade das variáveis. O mesmo resultado foi encontrado por Eberhardt, Helmers e Strauss (2013), que estimam a elasticidade do P&D com o estimador CCE de Pesaran (2006), que também controla a presença de fatores comuns no erro.

Eberhardt, Helmers e Strauss (2013) argumentam que esse resultado evidencia a dificuldade de separar os retornos individuais de P&D do efeito de *spillovers*. Os autores notam também que a melhor performance do modelo de fatores comuns latentes, que não faz hipóteses sobre o que está presente nestes fatores, indica que *spillovers* não são a única fonte de endogeneidade nos dados. Nesse caso, a estimação da elasticidade de P&D em uma função de produção não seria a forma mais adequada de se estudar sobre retornos privados de pesquisa.

5.2 Cálculo da PTF

Como discutido no capítulo 3, Eberhardt e Teal (2010), ao propor o AMG, propõem também uma nova forma de calcular a PTF a partir da tendência comum obtida pelo estimador. Este procedimento, descrito também no capítulo 3, foi realizado para obter a evolução da PTF de cada setor na amostra. Como a PTF depende das elasticidades e da tendência temporal comum estimadas pelo AMG, a qualidade da medida proposta depende da boa performance do AMG. Assim, é possível avaliar a qualidade das estimativas do estimador a partir de uma análise da PTF produzida.

A figura 3 apresenta a evolução da PTF por setor de forma análoga à figura 2. As séries, em geral, apresentam a mesma tendência que a reportada para a PTF no Brasil nos trabalhos discutidos no capítulo 2: uma trajetória de crescimento entre 2003 e 2013 e uma trajetória de queda entre 2013 e 2017, esboçando uma leve recuperação ao final deste período.

Vale notar ainda que essa tendência é comum a praticamente todas as séries. Como discutido para o caso do produto no capítulo 4, esta característica aponta para a presença de dependência *cross-section* e raiz unitária na PTF. Assim, é razoável supor que a PTF seja composta por uma combinação de fatores comuns não-estacionários.

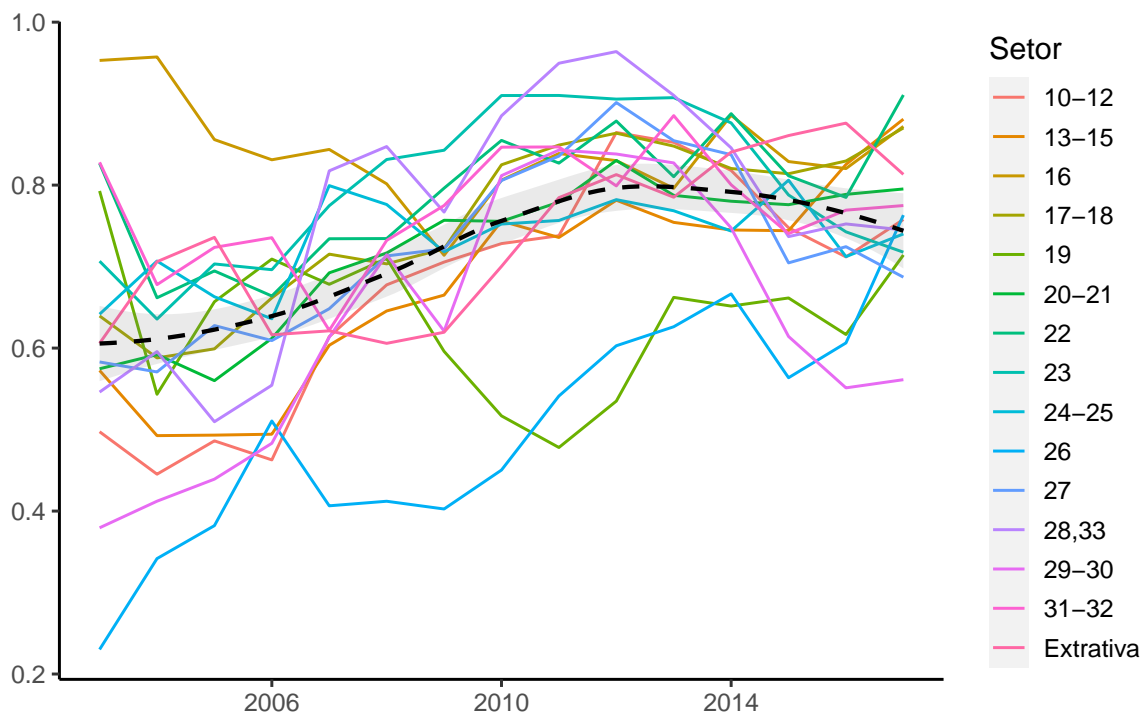
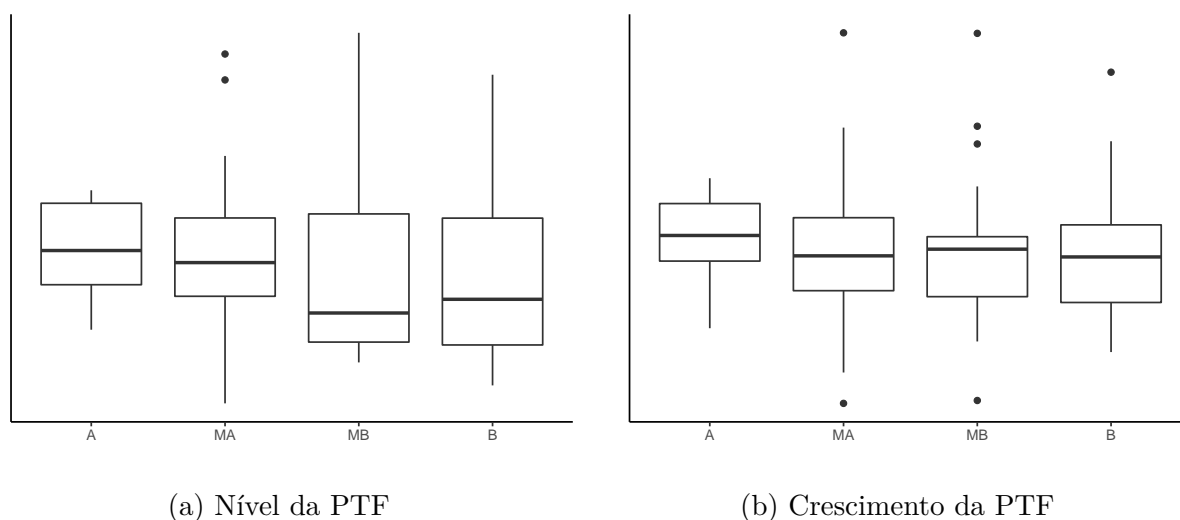


Figura 3 – Evolução média da PTF por Setor

A figura 4 apresenta a distribuição da PTF em termos de nível e taxa de crescimento por setor. Os setores estão agrupados pela classificação de intensidade tecnológica da OCDE⁴. O painel 4a indica que os setores de alta e média-alta intensidade tecnológica apresentam uma produtividade maior. O painel 4b aponta que o crescimento da PTF é maior para os setores de alta intensidade tecnológica e que os outros grupos não apresentam grande diferença entre si.



(a) Nível da PTF

(b) Crescimento da PTF

Figura 4 – Relação entre PTF e intensidade tecnológica

⁴ Médias dos setores entre 2003 e 2017

A figura 5 apresenta a relação do *share* de exportações⁵ com o nível e crescimento da PTF. A análise da figura aponta para a correlação entre um alto *share* de exportações e altos níveis e taxas de crescimento da PTF. Os coeficientes de correlação calculados foram de 0.32 para o nível da PTF e 0.19 para a taxa de crescimento. Como discutido por Kannebley (2011), uma explicação para isso é a autosseleção das firmas exportadoras, ou seja, o fato de que firmas mais produtivas são mais propensas a entrar e permanecer no mercado exportador, enquanto firmas menos produtivas tendem a ter pouca participação no mesmo. O autor apresenta evidência favorável a esta interpretação no caso brasileiro.

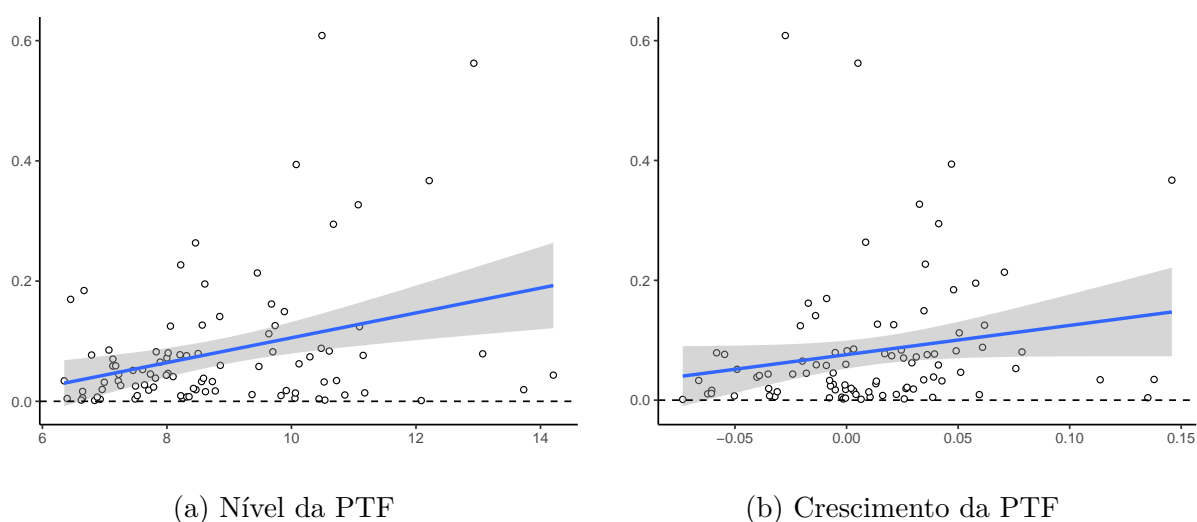


Figura 5 – Relação entre PTF e Exportações

A estimativa da PTF obtida pelo AMG apresenta correlação com a intensidade tecnológica e volume de exportações dos setores. Esse resultado corrobora a evidência na literatura sobre a correlação entre gastos em P&D, produtividade e *spillovers* — como apontado por Cohen e Levinthal (1989) e Kumbhakar et al. (2012) — e sobre a relação entre produtividade e orientação exportadora — como discutido por Kannebley (2011). Isso pode ser visto como evidência da qualidade desta medida da PTF e, conseqüentemente, do AMG.

5.3 Estimação com dados a partir de 2007

Como teste adicional de robustez, foi avaliada também a estimação com pessoal ocupado total e com distinção entre Potec e NPotec para os dados a partir de 2007. A tabela 8 apresenta os resultados do primeiro grupo de estimações.

As estimativas para o trabalho apresentaram média de 1.06, valor próximo do obtido com a amostra completa, de 1.043. Novamente, MQO e *system*-GMM foram os que mais se distanciaram, estando 1.5 desvios-padrão abaixo e 1.2 desvios-padrão acima

⁵ Razão entre volume de exportações e valor bruto de produção

Tabela 8 – Estimação sem Potec com Dados a Partir de 2007

	MQO	EF	PD	Sys-GMM	MG	AMG
Capital	0.41 (5.53)***	0.09 (1.38)	0.06 (1.55)	0.07 (0.82)	0.15 (2.02)**	0.13 (1.60)
Pessoal	0.86 (6.79)***	0.95 (10.06)***	1.11 (15.90)***	1.23 (7.75)***	1.1 (18.40)***	1.11 (17.77)***
Obs	1056	1056	960	960	1056	960
CRS	17.97***	0.18	5.04**	0.74	7.17***	5.68**
CD	-1.57	-1.79*	-0.33	-1.49	-1.1	-1.43
CIPS	-0.47	-0.22	-3.5***	-5.39*	0.35	-3.39***
RMSE	0.45	0.16	0.14	0.14	0.08	0.06

Fonte: Elaboração própria

da média respectivamente. Os coeficientes ainda se encontram dentro do intervalo obtido por Lucinda e Meyer (2013).

Os coeficientes do capital variaram entre 0.06 e 0.41, intervalo próximo ao obtido com a amostra completa. Agora, porém, a média caiu para 0.15. A estimativa de MQO é a que mais se distancia, estando quase 2 desvios-padrão acima da média. Os valores obtidos por efeitos fixos e *system*-GMM caíram quase pela metade em comparação com as estimativas para a amostra completa. A estimativa do capital passou a ser estatisticamente significativa apenas para MQO e MG.

Quanto à dependência *cross-section*, apenas o resíduo produzido por efeitos fixos não controla esta propriedade. As conclusões sobre a presença de raiz unitária nos resíduos são as mesmas que no caso da amostra completa, com exceção do MG, que apresenta raiz unitária. Finalmente, a análise do RMSE indica novamente que o AMG é o estimador que melhor se ajusta aos dados.

Por fim, a tabela 9 apresenta os resultados das estimações incluindo Potec. A média das estimativas da elasticidade do trabalho cai para 0.95 após a inclusão do Potec. Novamente, os valores obtidos pelos estimadores foram próximos, com exceção do MQO e *system*-GMM, cujas estimativas estão 1.36 e 1.05 desvios-padrão abaixo da média respectivamente. Os coeficientes continuam dentro do intervalo obtido por Lucinda e Meyer (2013) e, assim como no caso da amostra completa, a média é mais próxima da média de 0.97 obtida pelos autores.

As estimativas da elasticidade do capital caíram para MQO e AMG e aumentaram para *system*-GMM e MG. Os novos valores variaram entre 0.06 e 0.31, com média de 0.159. A estimativa de MQO continua sendo a que mais se distancia das outras, estando cerca de 1.43 desvios-padrão acima da média. Com relação às estimações sem Potec, o capital passa a ser estatisticamente significativo para o *system*-GMM.

Tabela 9 – Estimação com Dados a Partir de 2007

	MQO	EF	PD	Sys-GMM	MG	AMG
Capital	0.311 (5.30)***	0.09 (1.4)	0.06 (1.51)	0.16 (1.71)*	0.174 (2.06)**	0.09 (1.08)
Npotec	0.694 (6.67)***	0.952 (10.06)***	1.09 (15.89)***	1.009 (5.3)***	1.11 (21.20)***	1.14 (19.67)***
Potec	0.254 (6.12)***	-0.04 (-0.76)	0.07 (2.41)**	0.17 (2.28)**	0.028 (0.63)	0.05 (1.00)
Obs	1056	1056	960	960	1056	960
CRS	23.14***	0	7.43***	1.31	8.26***	6.01**
CD	-1.91*	-1.81*	-0.45	-1.79**	-0.73	-0.89
CIPS	3.38	4.11	-8.3***	-3.01***	-6.17***	-2.02***
RMSE	0.4	0.16	0.14	0.15	0.07	0.05

Fonte: Elaboração própria

O coeficiente do Potec varia de -0.04 (efeitos fixos) a 0.254 (MQO). Apenas os estimadores de MQO, *system*-GMM e primeiras diferenças produzem estimativas estatisticamente significantes para o Potec, sendo que apenas os dois primeiros produziam tal resultado no caso da amostra completa.

Após a inclusão do Potec, os resíduos de MQO, efeitos fixos e *system*-GMM apresentam dependência *cross-section*. Como na regressão com a amostra completa, apenas MQO e efeitos fixos não controlam a não-estacionariedade dos dados. A análise do RMSE mais uma vez indica que o AMG é o estimador com melhor ajuste aos dados.

A estimação com os dados a partir de 2007, em geral, corrobora os resultados obtidos na estimação principal. Novamente o AMG controla dependência *cross-section*, apesar de que não é mais o único. Dos 3 estimadores que produzem uma estimativa estatisticamente significativa para o Potec, apenas o de primeiras diferenças produz resíduos estacionários e sem dependência *cross-section*. A comparação entre MQO e MG novamente evidencia que a imposição errônea de homogeneidade dos parâmetros leva a estimativas distorcidas.

Os resultados desta seção devem ser avaliados com cautela. Como as séries de cada setor já eram pequenas na amostra original, a restrição do período de análise traz algumas dificuldades. Primeiro, os estimadores MG e AMG estimam séries individuais para os setores. Com a redução da amostra, as séries individuais consistem de 11 observações apenas. O *system*-GMM, por sua vez, utiliza as defasagens das variáveis como instrumentos para corrigir a endogeneidade dos dados. A redução da amostra implica um menor número de instrumentos possíveis para o estimador. Dessa forma, esses estimadores são menos precisos com a amostra reduzida.

Outra dificuldade diz respeito aos testes CD e CIPS. O primeiro parte da correlação entre as séries de resíduos dos setores enquanto o segundo estima regressões ADF aumentadas por médias *cross-section* para as séries dos resíduos de cada setor. Assim, a redução da amostra também diminui a precisão no cálculo das estatísticas de teste.

6 Conclusão

Este trabalho teve por objetivo estimar a elasticidade de P&D em um painel de setores da indústria extrativa e de transformação no Brasil. Para isso, foi estimada uma função de produção incluindo uma variável de mão-de-obra ocupada em P&D. Discutiu-se como a função de produção e as relações estruturais implícitas a ela podem ser representadas por um modelo de fatores comuns não-observados com alta persistência temporal e parâmetros heterogêneos, ressaltando as consequências disso para a estimação da elasticidade de P&D. A análise descritiva dos dados apresentou evidência favorável a essa representação da função de produção.

Nesse sentido, a principal contribuição deste trabalho é a utilização de um estimador que leve em conta a estrutura dos dados descrita acima. O estimador escolhido foi o AMG de Eberhardt e Teal (2010). A análise dos resíduos produzidos pelo AMG indica que o estimador controla a presença de fatores comuns e não-estacionariedade dos dados. Com isso, foi possível obter uma estimativa mais robusta da elasticidade de P&D na indústria brasileira.

A elasticidade de P&D estimada pelo AMG, de 0.014, não foi estatisticamente significativa. Isso indica a dificuldade de se separar o efeito individual do P&D do efeito de *spillovers* de conhecimento e choques comuns de produtividade. O mesmo resultado pode ser encontrado em Eberhardt, Helmers e Strauss (2013). Além disso, a comparação entre as estimativas com e sem a variável de P&D indica um possível viés causado pela omissão da variável de P&D. Esse efeito aponta para a correlação entre pesquisa, capital e trabalho.

A análise acima mostra como o impacto do investimento em pesquisa na produtividade está associado a *spillovers* de conhecimento e outros choques de produtividade. Além disso, pode-se inferir que o gasto em P&D também pode afetar a produtividade por meio de sua relação com trabalho e capital. Assim, é difícil estimar o retorno de P&D, mesmo que privado, sem levar em conta o efeito de *spillovers* e a relação entre inovação e

os fatores de produção. Eberhardt, Helmers e Strauss (2013) e Hall, Mairesse e Mohnen (2010) apresentam evidência favorável a esta conclusão.

Com relação à PTF calculada a partir dos resultados do AMG, é possível notar que, em geral, setores de alta intensidade tecnológica apresentam maior crescimento da produtividade. Esse resultado aponta para a correlação entre PTF e gastos em P&D. Isso indica que *spillovers* podem não ser a única fonte de endogeneidade da variável de P&D, caso outras fatores estejam presentes na PTF. Assim, incluir apenas uma *proxy* para *spillovers* pode não ser o suficiente para produzir estimativas consistentes.

As séries individuais da PTF apresentaram uma evolução próxima entre os setores, com uma trajetória de crescimento até 2013 e um movimento de queda desde então. Este resultado indica que a PTF de cada setor está correlacionada e que, em geral, as séries individuais não são estacionárias. Assim, justifica-se o uso de métodos que levem em conta a dependência *cross-section* e temporal dos dados, como é o caso do modelo de fatores comuns correlacionados.

Finalmente, outro resultado obtido sobre a PTF é que esta variável apresenta correlação positiva com a orientação exportadora dos setores, medida pelo volume de exportações. Este resultado também foi apontado por Sun, Hone e Doucouliago (1999) e Grossman e Helpman (1991). Os autores argumentam que a abertura comercial pode levar a ganhos de produtividade por gerar economias de escala, aumentar o acesso a fontes externas de conhecimento e por gerar uma pressão competitiva dentro dos setores. Kannebley (2011) nota que a correlação entre produtividade e exportação também pode se dar porque as firmas mais produtivas são as mais propensas a se engajar e obter ganhos com atividades de exportação.

A discussão dos resultados indica que a análise do impacto de P&D na produção por meio de sua inclusão como um insumo aditivamente separável na função de produção não é adequada. Como notam Eberhardt, Helmers e Strauss (2013), é necessário um maior entendimento de como se dá a dependência entre os setores e de como ocorre o *spillover* de conhecimento. Além disso, é importante estudar como o gasto em P&D pode afetar a produção indiretamente, por meio de sua influência nos outros insumos.

A principal limitação deste trabalho diz respeito aos dados utilizados. Primeiramente, as séries de tempo individuais dos setores eram relativamente pequenas. Como o AMG depende de estimativas individuais, o tamanho das séries dos setores afeta a qualidade da estimativa final. Ademais, com séries de tempo maiores, seria possível capturar melhor a dependência temporal dos dados. Apesar disso, o AMG ainda foi o estimador que melhor lidou com os problemas discutidos neste trabalho.

Referências

- ABDIH, Y.; JOUTZ, F. Relating the knowledge production function to total factor productivity: an endogenous growth puzzle. **IMF Staff Papers**, Springer, v. 53, n. 2, p. 242–271, 2006.
- AHN, S. C.; LEE, Y. H.; SCHMIDT, P. Panel data models with multiple time-varying individual effects. **Journal of econometrics**, Elsevier, v. 174, n. 1, p. 1–14, 2013.
- 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.
- _____. Gmm estimation with persistent panel data: an application to production functions. **Econometric reviews**, Taylor & Francis, v. 19, n. 3, p. 321–340, 2000.
- BONELLI, R. Growth and productivity in brazilian industries: impacts of trade orientation. **Journal of development economics**, Elsevier, v. 39, n. 1, p. 85–109, 1992.
- 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.
- CHUDI, A.; PESARAN, M. H. Large panel data models with cross-sectional dependence: a survey. **CAFE Research Paper**, n. 13.15, 2013.
- CHUDI, 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. A principal components approach to cross-section dependence in panels. Citeseer, 2002.

- _____. Unobserved heterogeneity in panel time series models. **Computational Statistics & Data Analysis**, Elsevier, v. 50, n. 9, p. 2361–2380, 2006.
- 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.
- 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.
- _____. Productivity analysis in global manufacturing production. 2010.
- _____. 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.
- ELLERY, R. Desafios para o cálculo da produtividade total dos fatores. **Produtividade no Brasil: desempenho e determinantes**, IPEA: ABDI, v. 1, p. 53–86, 2014.
- FAGERBERG, J. Technology and international differences in growth rates. **Journal of economic Literature**, JSTOR, v. 32, n. 3, p. 1147–1175, 1994.
- FILHO, F. d. H. B.; PESSÔA, S. d. A.; VELOSO, F. A. Evolução da produtividade total dos fatores na economia brasileira com ênfase no capital humano-1992-2007. **Revista Brasileira de Economia**, SciELO Brasil, v. 64, n. 2, p. 91–113, 2010.
- 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.
- _____. **The search for R&D spillovers**. [S.l.], 1991.
- GROSSMAN, G. M.; HELPMAN, E. Trade, knowledge spillovers, and growth. **European economic review**, Elsevier, v. 35, n. 2-3, p. 517–526, 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.
- HALL, B. H. R&d tax policy during the 1980s: success or failure? **Tax policy and the economy**, National Bureau of Economic Research and The MIT Press, v. 7, p. 1–35, 1993.

- 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.
- HENDRY, D. F. **Dynamic econometrics**. [S.l.]: Oxford University Press on Demand, 1995.
- 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.
- JONES, C. I. Time series tests of endogenous growth models. **The Quarterly Journal of Economics**, MIT Press, v. 110, n. 2, p. 495–525, 1995.
- KANNEBLEY, S. Firms heterogêneas e exportações: uma resenha à luz das evidências brasileiras. **Revista de Economia Contemporânea**, SciELO Brasil, v. 15, n. 1, p. 143–170, 2011.
- KAPETANIOS, G.; PESARAN, M. H.; YAMAGATA, T. Panels with non-stationary multifactor error structures. **Journal of Econometrics**, Elsevier, v. 160, n. 2, p. 326–348, 2011.
- KUMBHAKAR, S. C. et al. Corporate r&d and firm efficiency: evidence from europe's top r&d investors. **Journal of Productivity Analysis**, Springer, v. 37, n. 2, p. 125–140, 2012.
- 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.
- PESARAN, M. H. General diagnostic tests for cross section dependence in panels. CESifo working paper series, 2004.
- _____. Estimation and inference in large heterogeneous panels with a multifactor error structure. **Econometrica**, Wiley Online Library, v. 74, n. 4, p. 967–1012, 2006.
- _____. 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.
- ROMER, P. **Endogenous Technological Change**. [S.l.], 1989.
- SANTANA, B. G. **Retornos do P&D e incentivos fiscais: uma análise para o caso brasileiro**. Dissertação (Mestrado) — Universidade de São Paulo, 2017.
- 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.

SUN, H.; HONE, P.; DOUCOULIAGO, H. Economic openness and technical efficiency: a case study of chinese manufacturing industries. **Economics of Transition**, Wiley Online Library, v. 7, n. 3, p. 615–636, 1999.

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.

A Construção da Base de Dados

A.1 PIA

Este capítulo apresenta algumas das análises realizadas para a construção da base de dados deste trabalho, descritas em mais detalhe no capítulo 4.

A.1.1 Mudança da CNAE

Tabela 10 – Teste de Diferença de Média Global

VTI	Pessoal	Aquisições	Melhorias	Baixas
0.52	0.44	0.76	0.82	0.56

Fonte: Elaboração própria

Tabela 11 – 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

¹ P-Valores maiores que 0.99 arredondados para 1

Fonte: Elaboração própria

A tabela 12 apresenta o dicionário construído para realizar a compatibilização entre CNAE 1.0 e CNAE 2.0.

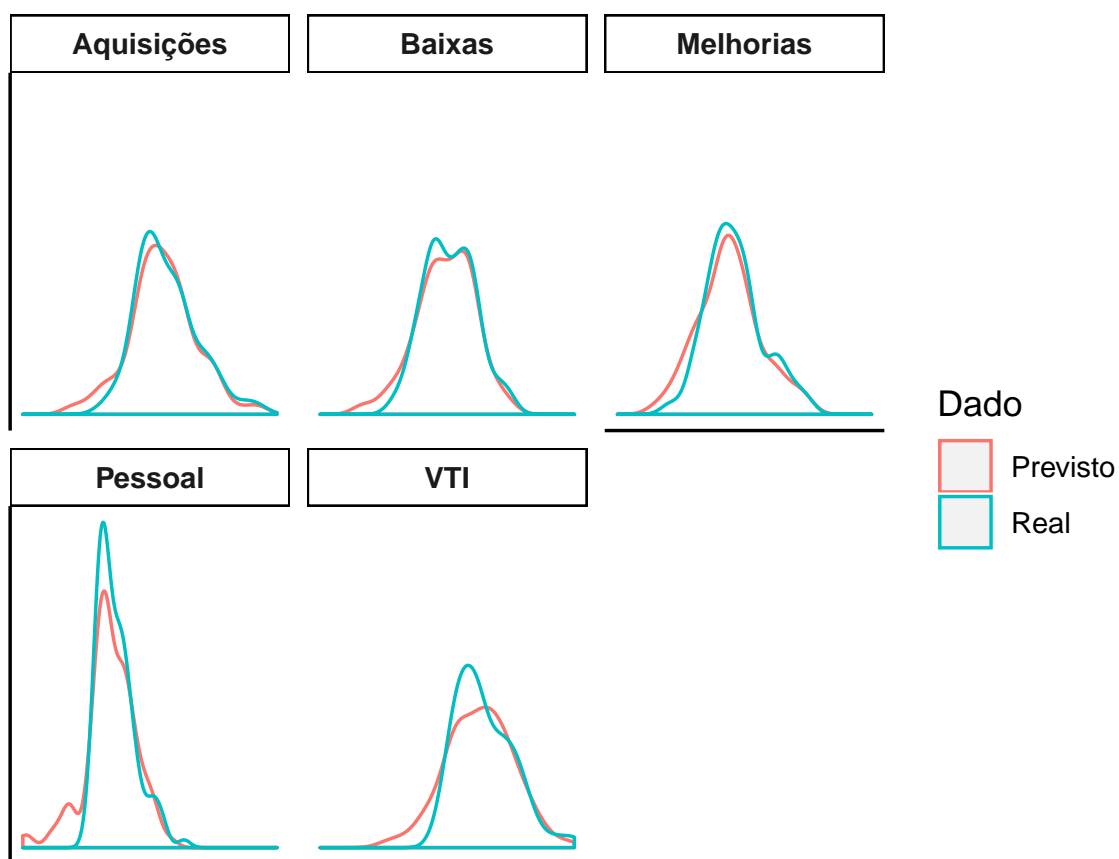


Figura 6 – Distribuições reais e previstas pelo dicionário

Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0

CNAE 1.0	CNAE 2.0	Porcentagem
100	050	0.98
100	089	0.02
111	060	1
112	091	1
131	071	1
132	072	1.00
141	081	0.95
142	089	0.93
151	101	0.90
151	102	0.02

Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0

CNAE 1.0	CNAE 2.0	Porcentagem
151	109	0.0002
152	103	0.99
153	104	0.99
154	105	1
155	106	0.98
156	107	0.99
157	108	1
158	109	0.97
159	111	0.51
159	112	0.47
160	121	0.83
160	122	0.17
171	131	0.79
172	131	1
173	132	0.99
174	132	0.32
175	134	1
176	132	0.002
176	134	0.004
176	135	0.68
176	329	0.01
177	133	0.84
177	142	0.13
181	141	0.99
182	141	0.42
182	325	0.03
182	329	0.51

Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0

CNAE 1.0	CNAE 2.0	Porcentagem
191	151	1
192	152	0.97
193	153	0.73
193	154	0.08
201	161	1
202	162	1
211	171	1
212	172	0.92
213	173	1
214	174	0.97
221	174	0.09
222	181	0.82
222	182	0.16
223	183	1
231	191	1
232	192	0.99
234	193	1
241	201	0.96
242	193	0.02
242	202	0.57
243	203	0.96
244	204	1
245	211	0.01
245	212	0.88
245	325	0.01
246	205	1
247	206	0.92

Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0

CNAE 1.0	CNAE 2.0	Porcentagem
248	207	1
249	209	0.85
249	268	0.01
251	221	0.97
251	329	0.004
252	222	0.79
252	329	0.003
261	231	0.99
262	232	1
263	233	1
264	234	0.99
269	239	0.78
271	241	0.98
272	242	1
273	243	1
274	244	0.99
275	245	0.29
281	251	0.95
282	252	0.94
283	253	0.95
284	251	0.004
284	254	0.93
288	331	1
289	259	0.95
289	282	0.001
289	329	0.003
291	281	0.84

Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0

CNAE 1.0	CNAE 2.0	Porcentagem
292	282	0.72
292	286	0.002
293	283	0.78
294	284	0.83
295	285	0.89
296	254	0.09
296	286	0.76
297	255	0.71
297	332	0.29
298	275	0.93
298	282	0.001
299	331	0.96
301	282	0.67
302	262	0.92
311	271	0.77
312	273	0.99
313	273	1
314	272	1
315	274	0.90
315	329	0.10
316	265	0.01
316	294	0.93
318	331	1
319	279	0.76
321	261	0.59
322	261	0.19
322	263	0.78

Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0

CNAE 1.0	CNAE 2.0	Porcentagem
322	279	0.004
323	264	0.94
329	331	1
331	266	0.17
331	310	0.002
331	325	0.73
332	265	0.93
332	325	0.003
333	265	0.78
334	267	0.30
334	325	0.67
335	265	1
339	331	0.14
341	291	0.10
342	292	0.66
343	293	0.75
343	331	0.01
344	294	0.91
345	295	1
351	301	0.50
351	331	0.50
352	303	0.89
352	331	0.11
353	304	0.29
353	331	0.71
359	309	0.98
361	310	1.00

Tabela 12 – Dicionário de Correspondência entre CNAE 1.0 e CNAE 2.0

CNAE 1.0	CNAE 2.0	Porcentagem
361	332	0.004
369	209	0.02
369	282	0.0003
369	309	0.02
369	321	0.11
369	322	0.01
369	323	0.05
369	324	0.16
369	329	0.55

Fonte: Elaboração própria

A.1.2 Missing

A tabela abaixo apresenta os setores com *missing* na CNAE 1.0.

Tabela 13 – Setores com Missing - CNAE 1.0

CNAE	VTI	Pessoal	Salarios	A	M	B
100	0.00	0.00	0.00	0.00	4	0.00
111	8.00	8.00	8.00	8.00	9	8.00
112	0.00	0.00	0.00	0.00	1	0.00
231	3.00	3.00	3.00	3.00	11	6.00
233	12.00	12.00	12.00	12.00	12	12.00
244	0.00	0.00	0.00	0.00	1	0.00
288	7.00	7.00	7.00	7.00	9	7.00
299	7.00	7.00	7.00	7.00	7	7.00
301	0.00	0.00	0.00	0.00	2	0.00
318	7.00	7.00	7.00	7.00	7	7.00
329	7.00	7.00	7.00	7.00	8	7.00
339	7.00	7.00	7.00	7.00	8	7.00
371	0.00	0.00	0.00	0.00	3	0.00

Fonte: Elaboração própria

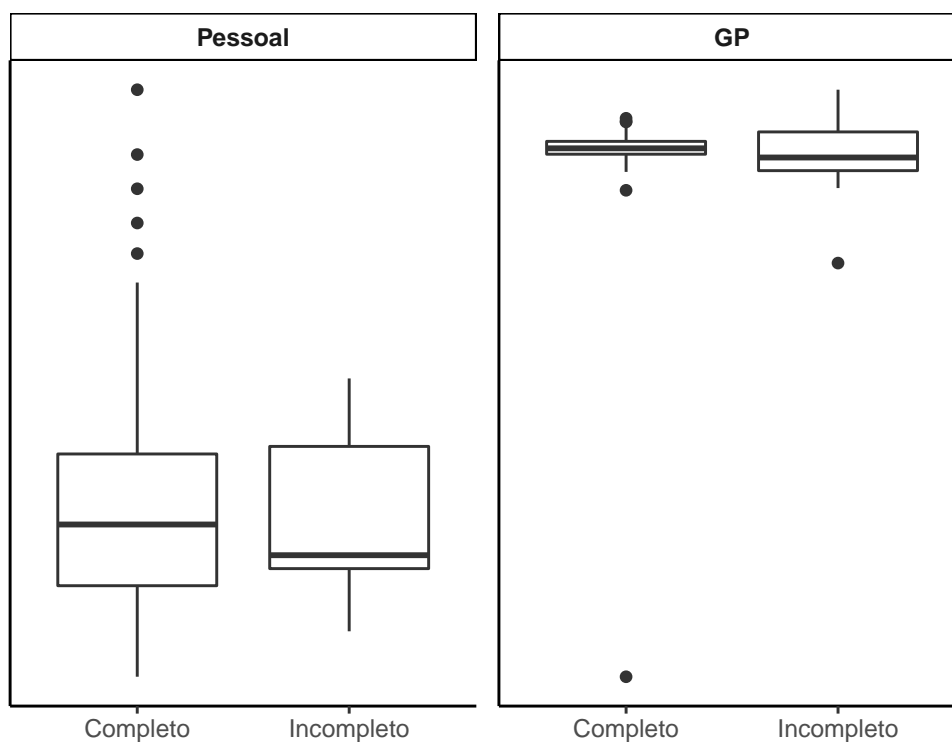
As figuras 7 e 8 apresentam a comparação gráfica entre setores com informação

completa e incompleta. A tabela 14 apresenta os resultados da estimação do logit para avaliar a relação entre características econômicas e informação completa ¹.

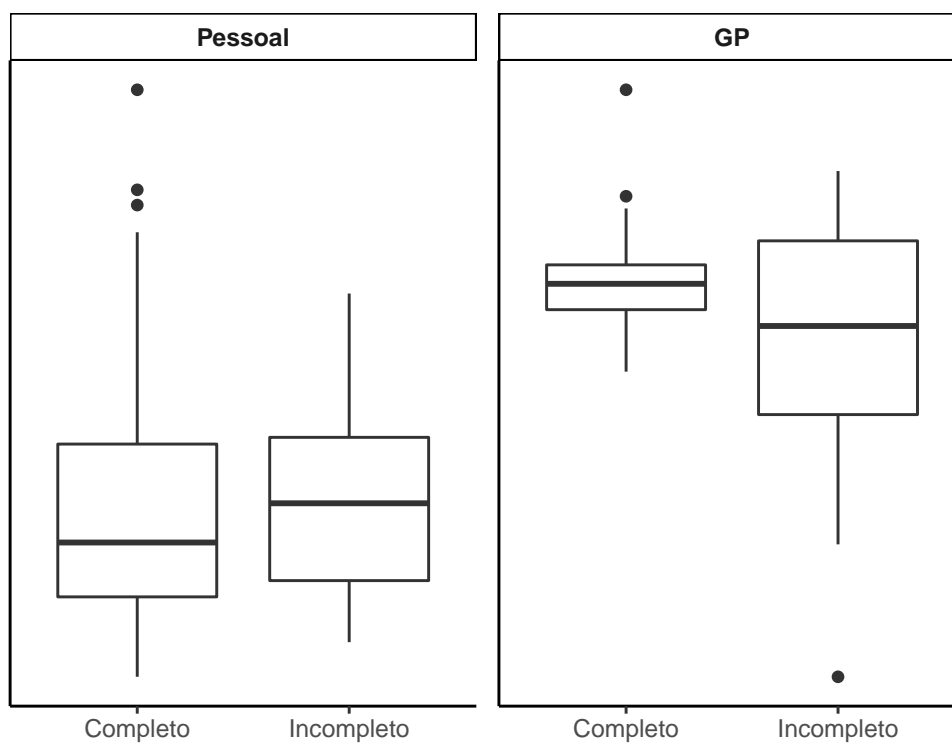
Tabela 14 – Resultados - Logit

	<i>Variável Dependente:</i>	
	Missing	
	(1)	(2)
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
<i>Nota:</i>	*p<0.1; **p<0.05; ***p<0.01	

¹ Erro-Padrão entre parênteses

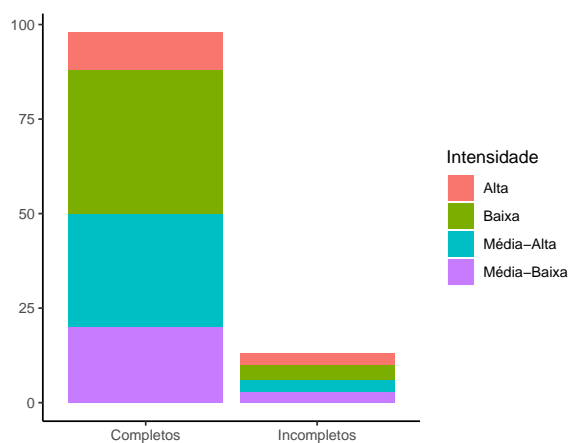


(a) 1996

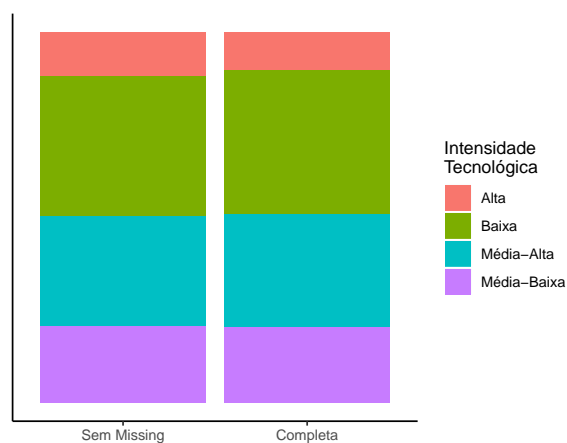


(b) 2007

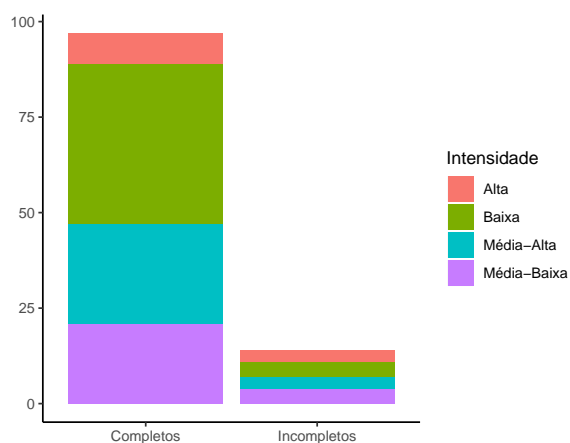
Figura 7 – Quantidade de Missing por Intensidade Tecnológica



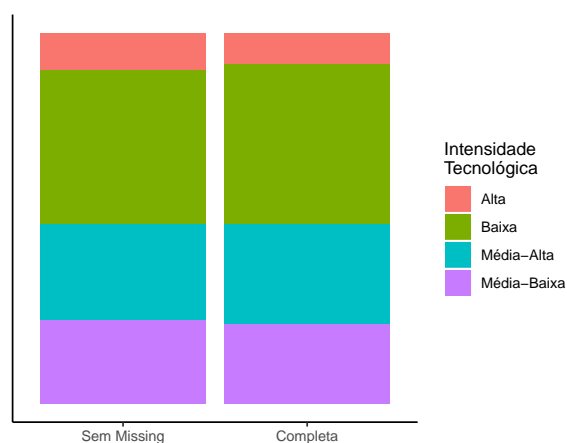
(a) Setores com Missing - 1996



(b) Amostra completa x restrita - 1996



(c) Setores com Missing - 2007



(d) Amostra completa x restrita - 2007

Figura 8 – Quantidade de Missing por Intensidade Tecnológica

B Potec x Dados de Inovação

Seguindo o procedimento de Araújo, Cavalcante e Alves (2009), serão comparados os números do Potec com os dados de gasto em inovação da Pintec ¹. As tabelas 15a e 15b apresenta a correlação do Potec com cada variável da Pintec associada a gasto em P&D.

Tabela 15 – Comparação entre Potec e dados de gastos em inovação da Pintec

(a) Apenas Indústria de Transformação

Gastos em Inovação	Correlação
Total	0.84
Atividades internas de pesquisa e desenvolvimento	0.59
Aquisição externa de pesquisa e desenvolvimento	0.61
Aquisição de outros conhecimentos externos	0.87
Aquisição de software	0.45
Aquisição de máquinas e equipamentos	0.79
Treinamento	0.66
Introdução das inovações tecnológicas no mercado	0.85
Projeto industrial e outras preparações técnicas	0.83

(b) Indústria Extrativa e de Transformação

Gastos em Inovação	Correlação
Total	0.68
Atividades internas de pesquisa e desenvolvimento	0.49
Aquisição externa de pesquisa e desenvolvimento	0.49
Aquisição de outros conhecimentos externos	0.75
Aquisição de software	0.36
Aquisição de máquinas e equipamentos	0.64
Treinamento	0.56
Introdução das inovações tecnológicas no mercado	0.79
Projeto industrial e outras preparações técnicas	0.76

Fonte: Elaboração própria

Como é possível ver, os gastos com P&D na indústria extrativa parecem ter menos correlação com o Potec do que na de transformação. Parte dessa diferença pode se dar porque a Pintec divulga os dados da segunda divididos por setores enquanto os da primeira são divulgados apenas agregados.

¹ Escolheu-se trabalhar com o Potec ao invés da variável de gasto em P&D interno da Pintec porque os dados da Pintec são muito mais agregados, além de serem disponibilizados apenas a cada 3 anos, dificultando uma análise mais precisa

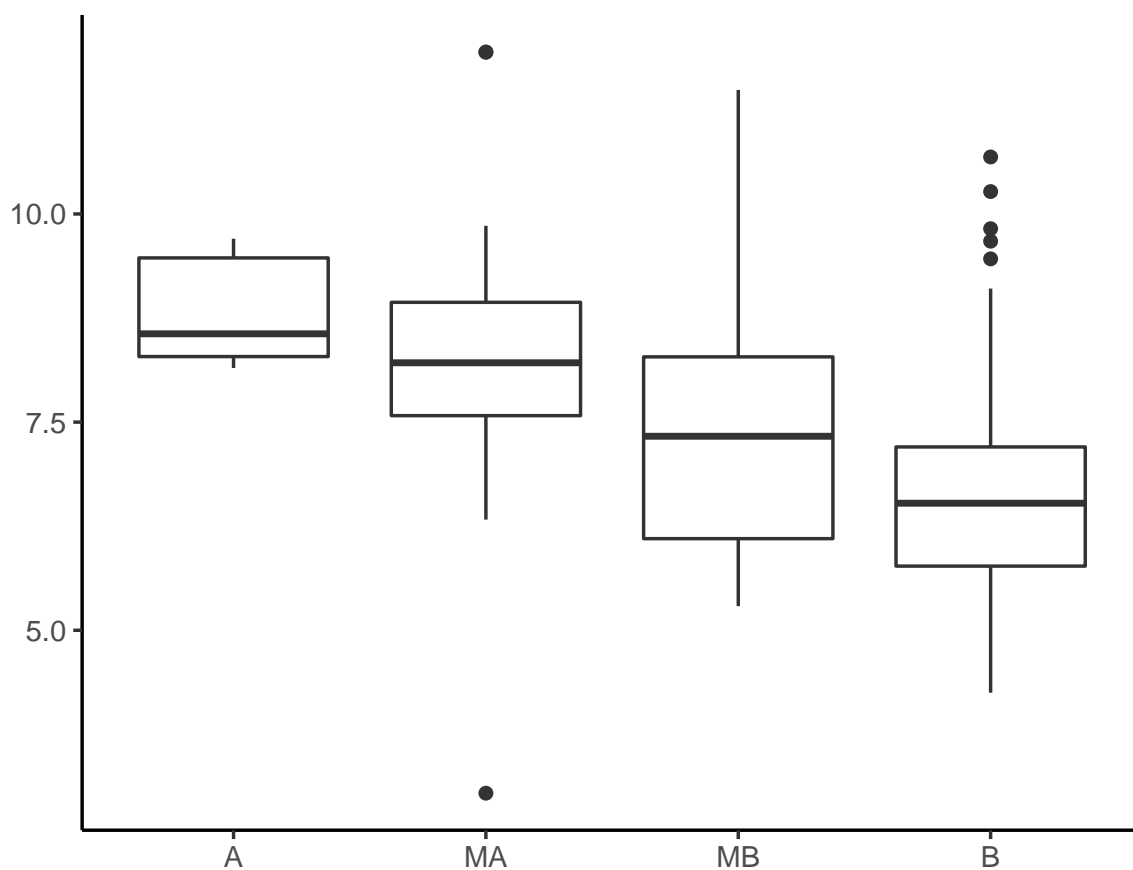


Figura 9 – Potec por Intensidade Tecnológica da OCDE

A figura 9 apresenta a distribuição do Potec pelas categorias de intensidade definidas pela OCDE². Os dados de Potec parecem corresponder bem à classificação da OCDE.

² Alta (A), Média-Alta (MA), Média-Baixa (MB) e Baixa (B)

C Descritivas Adicionais

A figura 10 apresenta a evolução média das séries de cada variável, exceto produto, por setor.

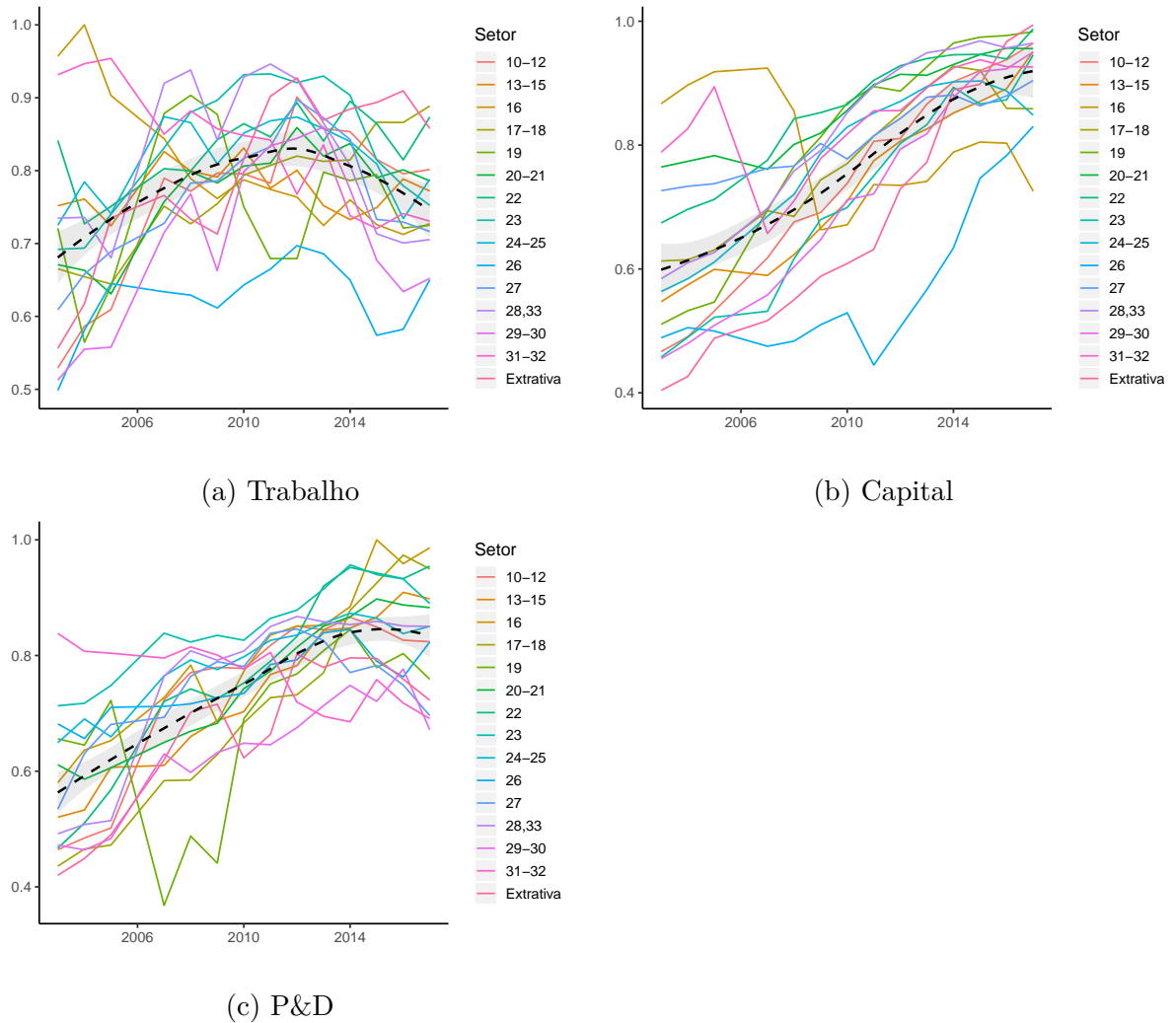


Figura 10 – Evolução Média das Variáveis por Setor