

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
MESTRADO EM ECONOMIA APLICADA

MARCEL FERREIRA DE OLIVEIRA

**DINÂMICA DA PRODUTIVIDADE: UMA ABORDAGEM POR MEIO DE
MODELO DE FATORES DINÂMICOS**

JUIZ DE FORA
2015

MARCEL FERREIRA DE OLIVEIRA

**DINÂMICA DA PRODUTIVIDADE: UMA ABORDAGEM POR MEIO DE
MODELO DE FATORES DINÂMICOS**

Dissertação apresentada ao programa de Pós-Graduação em Economia Aplicada da Faculdade de Economia da Universidade Federal de Juiz de Fora como parte das exigências do Mestrado em Economia Aplicada.

Orientador: Prof. Dr. Sidney Martins Caetano

Co-Orientador: Prof. Dr. Guilherme Valle Moura

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Ferreira de Oliveira, Marcel.

DINÂMICA DA PRODUTIVIDADE: : UMA ABORDAGEM POR MEIO DE MODELO DE FATORES DINÂMICOS / Marcel Ferreira de Oliveira. -- 2015.

115 p.

Orientador: Sidney Martins Caetano

Coorientador: Guilherme Valle Moura

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Economia. Programa de Pós-Graduação em Economia, 2015.

1. Produtividade Total dos Fatores. 2. Modelo de Fatores Dinâmicos. 3. GMM Sistêmico. 4. Crescimento Econômico. I. Martins Caetano, Sidney, orient. II. Valle Moura, Guilherme, coorient. III. Título.

MARCEL FERREIRA DE OLIVEIRA

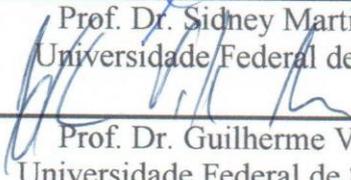
**DINÂMICA DA PRODUTIVIDADE: UMA ABORDAGEM POR MEIO DE
MODELO DE FATORES DINÂMICOS**

Dissertação submetida à Faculdade de Economia da Universidade de Juiz de Fora, como requisito parcial para a obtenção do título de Mestre em Economia Aplicada, com área de concentração em Macroeconomia Aplicada.

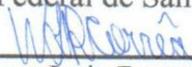
APROVADA EM 11/06/2015



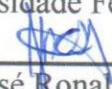
Prof. Dr. Sidney Martins Caetano
Universidade Federal de Juiz de Fora



Prof. Dr. Guilherme Valle Moura
Universidade Federal de Santa Catarina



Prof. Dr. Wilson Luiz Rotatori Corrêa
Universidade Federal de Juiz de Fora



Prof. Dr. José Ronaldo de Castro Souza Júnior
Instituto de Pesquisa Econômica Aplicada - IPEA

Agradecimentos

Primeiramente agradeço ao meu orientador e co-orientador, os Professores Sidney Martins Caetano e Guilherme Valle Moura, pela paciência e solicitude durante a confecção desta dissertação. Também agradeço aos membros convidados para a banca, que aceitaram o convite e se dispuseram a ler e avaliar esta dissertação.

Agradeço aos professores do mestrado em Economia Aplicada da Universidade Federal de Juiz de Fora pela disposição em auxiliar e ensinar os alunos do curso. Também agradeço aos funcionários do serviço de pós-graduação e aos funcionários responsáveis pela limpeza do prédio de economia da UFJF, pela solicitude e prestatividade dos mesmos.

Agradeço aos meus colegas e amigos tanto da minha turma quanto das turmas anteriores e posteriores do corpo discente do Programa de Pós-Graduação em Economia Aplicada, que me fizeram companhia durante os anos em que estive em Juiz de Fora. A companhia de vocês certamente tornou a estadia nesta cidade bem mais agradável.

Agradeço à FAPEMIG pelo apoio financeiro durante a maior parte do curso.

Agradeço, finalmente, aos meus pais por todo o suporte na minha decisão de vir para Juiz de Fora fazer mestrado aqui.

“what is clear to historians and laypersons alike is that capitalism is unique in the extraordinary growth record it has been able to achieve in its recurring industrial revolutions that have produced an outpouring of material wealth unlike anything previously seen in human history.”

(William J. Baumol)

Resumo: O objetivo desta dissertação é estudar a dinâmica da produtividade total dos fatores (PTF) através de um modelo de fatores dinâmicos e verificar a importância do *catch up* e outros determinantes da PTF por meio do estimador do GMM Sistemático, utilizando dados da Penn World Table 9.0 entre 1970 e 2014. Devido à existência de transbordamentos tecnológicos e comércio internacional, é esperado que existam variações comuns entre as produtividades de diferentes países. Sendo assim, estimamos um modelo de fatores dinâmicos para as PTFs de um conjunto de países tecnologicamente avançados para capturar esses efeitos. Dessa estimativa fomos capazes de extrair seu fator comum e usá-la como *proxy* para PTF mundial. Isso nos permite incluir a tendência comum na regressão da PTF e estimar seus efeitos sobre a PTF. Nosso fator comum estimado reflete bem aos principais choques de produtividade que ocorreram no período e a sensibilidade de cada país com relação ao fator comum parece estar negativamente relacionada com o nível de desenvolvimento de cada país – essa correlação parece estar de acordo com o argumento de que há um efeito de *catch up* na produtividade: países mais distantes da fronteira têm crescimento maior por absorverem mais da tecnologia mundial. Além disso, nossas estimativas para o painel dinâmico estimado utilizando este fator estimado reforça esse resultado.

Palavras-chave: Modelo de Fatores Dinâmicos, Produtividade Total dos Fatores, GMM Sistemático, Crescimento Econômico.

Abstract: The purpose of this dissertation is to study the global dynamics of total factor productivity (TFP) through a dynamic factors model and verify the importance of the catch up effect and other determinants of TFP through the System GMM estimator, using data from the Penn World Table 9.0 from 1970 to 2014. Because there're technological spillovers and international trade, it would be expected to see the existence of common variations in the productivity of different countries. Therefore, we apply a dynamic factor model to these productivities and its growth rates in order to capture these effects. These estimates allow us to extract the common factor and use it as a proxy for the world TFP. It allows us to include the common trend in the dynamic equation in the TFP regression in order to estimate its effects on the TFP. Our estimated common factor captures well the main productivity shocks that occurred in the period, and the sensibility of each country to the common factor seems to be negatively related to its country level of technological development – this correlation is in accordance with the argument that there's a catch up effect in productivity: countries more distant from the frontier have higher growth rates because there is more room to absorb from the global technology. Moreover, our estimates for the dynamic panel using this estimated factor reinforces this result.

Keywords: Dynamic Factors Model, Total Productivity of Factors, System GMM, Economic Growth.

LISTA DE ILUSTRAÇÕES

Figura 1 - Estimativa da distribuição de países de acordo com o PIB por trabalhador (em log e ajustado pela PPP) em 1960, 1980 e 2000.....	34
Figura 2 - Séries de <i>proxies</i> da PTF mundial entre 1970 e 2014 (em log e centradas em torno de suas médias).....	49
Figura 3 - PTFs de países selecionados entre 1970 e 2014 (em log).....	50
Figura 4 - Evolução da PTF de países selecionados, entre 1970 e 2014.	51
Figura 5 - Taxas de crescimento das PTFs (países selecionados) entre 1970 e 2014.	52
Figura 6 - Fator comum e Fatores específicos.....	54
Figura 7 - Variáveis selecionadas e loadings (DFM)	58
Figura 8 - Correlação entre Índices de Instituições e Produto per capita (em log).	62
Figura 9 – Fator comum e Fatores específicos (G7, PWT 8.0).....	111
Figura 10 - Correlação entre variáveis selecionadas e loadings (G7, PWT 8.0)	112
Figura 11 – Fator comum e Fatores específicos (G7, PWT 9.0)	114
Figura 12 - Correlação entre variáveis selecionadas e loadings (G7, PWT 9.0)	115

LISTA DE TABELAS

Tabela 1 - Decomposição do desenvolvimento (em %)	15
Tabela 2 - Variáveis utilizadas na construção do segundo indicador de instituições.	47
Tabela 3 - Correlações entre <i>proxies</i> da PTF mundial	49
Tabela 4 - Correlações entre as PTFs dos países selecionados.	51
Tabela 5 – Estimativas do modelo de fatores dinâmicos.....	53
Tabela 6 - Indicadores de tecnologia e produtividade vs <i>loadings</i>	56
Tabela 7 - Correlações entre <i>loadings</i> e indicadores de tecnologia e produtividade.....	57
Tabela 8 - Decomposição da variância	58
Tabela 9 - Estatísticas para os <i>gaps</i>	59
Tabela 10 - Estimativas para o <i>gap</i> entre PTF mundial e PTFs. Médias temporais dos <i>gaps</i> médios.....	60
Tabela 11 - GMM Sistemico em um estágio para diferentes índices de instituições	64
Tabela 12 - GMM Sistemico em dois estágios para diferentes índices de instituições	65
Tabela 13 - GMM Sistemico em um estágio para diferentes <i>proxies</i> para PTF mundial.	66
Tabela 14 - Limites 'razoáveis' para o coeficiente autorregressivo do GMM Sistemico (limites dados pelas estimativas do MQO e Efeitos Fixos (FE).....	67
Tabela 15 - GMM Sistemico em um estágio para diferentes <i>proxies</i> para PTF mundial.	67
Tabela 16 - Resultados para o modelo de fatores dinâmicos (G7, PWT 8.0).....	110
Tabela 17 - Decomposição da Variância (G7, PWT 8.0)	112
Tabela 18 - Resultados para o modelo de fatores dinâmicos (G7, PWT 9.0).....	113
Tabela 19 - Decomposição da Variância (G7, PWT 9.0)	113

Sumário

1. INTRODUÇÃO	8
2. REFERENCIAL TEÓRICO	12
2.1. Modelo Neoclássico de Crescimento	12
2.2. Modelos de Crescimento Endógeno	16
2.3. Difusão Tecnológica e Modelo para PTF	17
2.4. Tecnologia e Capital Humano	24
2.5. Eficiência Alocativa e PTF	34
3. METODOLOGIA	36
3.1. O Modelo	36
3.2. Procedimento de Estimação	40
3.2.1. Modelo de Fatores Dinâmicos	40
3.2.2. Decomposição da Variância	41
3.2.3. Método dos Momentos Generalizado Sistêmico	42
3.3. Base de Dados	43
4. RESULTADOS	49
4.1. Análise Descritiva	49
4.2. Resultados: Modelo de Fatores Dinâmicos	52
4.3. Recuperando a constante do Fator Comum	59
4.4. Resultados: GMM Sistêmico	60
4.4.1. Instituições e Produtividade	61
4.4.2. Proxies para a PTF mundial	65
5. CONSIDERAÇÕES FINAIS	69
6. REFERÊNCIAS	72
7. APÊNDICE A	78
A1. Análise Fatorial	78
A2. Modelos de Estado Espaço e Filtro de Kalman	86
A3. Modelo de Fatores Dinâmicos	95
A4. Decomposição da Variância	102
A5. Método dos Momentos Generalizados	103
8. APÊNDICE B	110
B1. Modelo de Fatores Dinâmicos com os países do G7	110

1. INTRODUÇÃO

Durante a maior parte da história da humanidade o padrão de vida de todos os povos era semelhante, caracterizado por pobreza generalizada e sem grandes perspectivas de mudança: por séculos o crescimento da renda per capita no mundo todo foi baixo ou mesmo inexistente. No entanto, desde a revolução industrial o cenário econômico mundial mudou radicalmente e o que se observou desde então foi um aumento persistente dos níveis de renda de diversos países. Todavia, alguns cresceram mais do que outros, gerando grandes desigualdades de renda per capita a nível internacional (ver ACEMOGLU, 2009; VILLELA, 2013). Duas questões que se colocam são, então, (i) como se deu e se dá esse crescimento da renda per capita e (ii) por que alguns países cresceram menos do que outros? Este trabalho busca contribuir para a resposta desta última questão.

Uma forma de ver essa questão é partindo dos fatores que compõe o produto per capita. Se pudermos definir o produto de um país enquanto uma função de produção agregada, esta depende de sua dotação de insumos e também da capacidade ou eficiência com a qual se transformam os insumos disponíveis em produto: isto é, por um elemento de eficiência. Esta é a chamada Produtividade Total dos Fatores (PTF). Um fato estilizado na literatura recente de crescimento econômico é que este último elemento, a PTF, explica a maior parte das diferenças de renda per capita observada entre os países. Essa evidência está presente nos trabalhos de Hall e Jones (1999), Jones e Romer (2009), Klenow e Rodríguez-Clare (1992), Prescott (1998), Veloso, Ferreira e Pessoa (2013), Hsieh e Klenow (2010) e outros. Considerada a relevância, não é de se surpreender que boa parte das pesquisas recentes sobre crescimento e desenvolvimento econômico têm se voltado para entender melhor este componente (DANQUAH, MORAL-BENITO E BAZOUMANA, 2011). Nas palavras de Easterly e Levine (2001): “in searching for the secrets of long-run economic growth, a high priority should be placed on rigorously defining the term “TFP” [Total Fator Productivity], empirically dissecting TFP, and on identifying the policies and institutions most conducive to TFP growth”.

É possível decompor o crescimento da PTF em três componentes: o progresso técnico, mudança de eficiência técnica e eficiência alocativa (ver PIRES E GARCIA, 2011). Se for também possível assumir que os fatores são alocados de maneira eficiente, então a PTF pode ser chamada de tecnologia indiscriminadamente, e é por isso que boa parte da literatura de crescimento de longo prazo se refere à PTF simplesmente como

tecnologia. O desenvolvimento do componente tecnologia – do qual se acredita que contabiliza a maior parte da PTF, em tese deveria ser comum a todos os países no longo prazo. Inclusive, é assim que o modelo neoclássico de crescimento trata a tecnologia: afinal, o conhecimento usado na produção de novas tecnologias é um bem comum, ele está disponível sem grandes dificuldades para praticamente todo mundo. O mesmo, porém, não é necessariamente válido para a tecnologia, pois o processo de difusão tecnológica pode ser mais lento que a difusão do conhecimento. Para ilustrar isso, suponha que a difusão tecnológica siga uma curva-S: quando uma tecnologia é criada ela começa se espalhando de maneira lenta até chegar a um ponto em que ela se torna popular e é adotada massivamente. Quando a maior parte das pessoas já estiver usando essa tecnologia, há então um decaimento brusco em sua adoção. Aquele conhecimento já foi difundido e até a nova tecnologia já está disponível, mas ela não é imediatamente adotada por todo mundo. Assim, o adequado entendimento das diferenças de produtividade no mundo deve considerar também os elementos que influenciam a adoção de novas tecnologias.

Os modelos de difusão tecnológica comumente supõem a existência de uma PTF mundial, que pode ser descrita como a fronteira tecnológica ou mesmo tomando uma forma mais geral como um índice agregado de tecnologia para o mundo (ver ACEMOGLU, 2009; AIYAR E FEYRER, 2002; DAMSGAARD E KRUSELL, 2010). Supõe-se, então, que a evolução da tecnologia de um país é função de seus próprios investimentos em pesquisa e desenvolvimento (P&D) e também da tecnologia produzida pelos demais países.

Como explica Acemoglu (2009), existem grandes diferenças de produtividade e emprego de tecnologia tanto internacionalmente quanto nacionalmente: há diferenças de PTF mesmo entre plantas dentro de setores próximos. Evidentemente que as explicações para as diferenças de tecnologia a nível setorial/nacional e internacional são diferentes. Em nível internacional o autor foca em três causas para as diferenças de produtividade e tecnologia: (i) difusão tecnológica lenta; (ii) existência de tecnologias apropriadas para contextos específicos; e (iii) instituições.

A primeira hipótese trabalha com a noção de que o processo de difusão tecnológico é lento. Assim, alguns países podem ter condições de adotar essas tecnologias produzidas mais ou menos rapidamente. Sobre esse aspecto, Prescott (1998) e Parente e Prescott (2000) levantam o argumento de que muitas inovações são poupadoras de trabalho, e isto gera uma espécie de conflito distributivo: ao introduzir a

tecnologia poupadora de trabalho, põe-se em risco o emprego dos trabalhadores já estabelecidos naquele setor. Estes, por sua vez, têm incentivos para barrar a adoção da nova tecnologia, logo o poder de barganha dos trabalhadores e instituições que atuam sobre o mercado de trabalho tem o potencial para barrar a adoção de determinadas tecnologias por anos, mesmo que a tecnologia já exista e o país tenha plenas condições de utilizá-la.

Outra linha de argumentação vê no capital humano um fator essencial para a absorção de novas tecnologias: Nelson e Phelps (1966) levantam a hipótese de que empreendedores mais bem educados têm maior facilidade para adotar novos produtos e técnicas, pois estes teriam maior capacidade de avaliar essas inovações e determinar com mais facilidade se são promissoras ou não. Assim, a adoção de tecnologias próximas à fronteira seria mais rápida para países com altos níveis de capital humano.

Nesta dissertação buscamos contribuir com essa dissecação empírica da PTF, mais especificamente, buscamos entender melhor como esta se relaciona com a chamada PTF mundial ou fronteira tecnológica e se o efeito de *catch up* tecnológico entre países realmente se verifica na prática. Nossas estimativas vão ao sentido de confirmar a presença de *catch up* tecnológico, bem como convergência condicional entre as PTFs.

Inicialmente buscamos construir uma *proxy* para a PTF mundial dividindo a PTF em dois fatores: um ‘global’, comum à todos os países, e outro específico de cada país. A ideia aqui é que a PTF de cada país não está num vácuo: na presença de transbordamentos tecnológicos e comércio internacional, é de se esperar que as PTFs se relacionem entre si. Desse modo, existem choques de produtividade comuns a todos os países, tal como pela produção de uma nova tecnologia. É de se esperar, então, que diferentes países se comportem de maneira diferente frente aos choques comuns.

Sendo assim, neste trabalho buscamos fazer duas coisas: uma primeira é construir uma *proxy* para PTF mundial por meio de um modelo de fatores dinâmicos. Enquanto trabalhos como o de Aiyar e Feyrer (2002) assumem uma tendência comum entre os países, mas o elimina do modelo transformando os dados, nós buscamos estimar essa tendência. Para tal fim, estimamos um modelo de fatores dinâmicos para os 10 países com maior eficiência técnica segundo o *ranking* de Pires e Garcia (2011). Os países utilizados no modelo são: Estados Unidos, Japão, Suíça, Grã Bretanha, Israel, Suécia, Canadá, Hong Kong, Dinamarca e Noruega. O nosso recorte temporal é entre 1970 e 2014. Nossos resultados com esse fator comum sugerem uma relação negativa

entre a sensibilidade de cada país em relação ao fator comum e o nível de desenvolvimento tecnológico do mesmo. Isso vai em direção da hipótese de que há um efeito de *catch up* da produtividade: países mais distantes da fronteira tecnologia têm bastante espaço para absorver tecnologia dos outros países, enquanto os países mais próximos não desfrutam desse bônus.

Nós levamos essas primeiras evidências adiante inserindo o fator estimado em um painel dinâmico composto de 144 países entre 1970 e 2014. A estimação foi feita a partir do Método dos Momentos Generalizados Sistemático (GMM Sistemático). Os resultados dão suporte para a relação sugerida anteriormente: os países da OCDE absorvem sistematicamente menos da PTF mundial do que os países fora desse grupo. Seguindo a formulação teórica de Aiyar e Feyrer (2002), também encontramos evidências de convergência condicional na PTF, porém a uma taxa mais modesta: algo entre 0.56 e 1.13% (em contraposição a taxa de 3% ao ano, encontrado pelos autores). Em nosso modelo, consideramos diferentes potenciais determinantes da produtividade (inflação, gasto do governo, abertura comercial, capital humano e instituições), utilizando três variáveis de instituições e duas outras *proxies* para a PTF mundial (média das PTFs dos países da OCDE e PTF dos Estados Unidos). Uma coisa que se nota é que, enquanto a relação da PTF com os demais determinantes não parecem ser muito robustas em nosso modelo (sua significância depende muito da especificação do modelo), a relação da PTF com nossas *proxies* para PTF mundial parecem relativamente mais robustas.

Além desta introdução, esta dissertação é composta por mais quatro seções. A segunda seção apresenta uma revisão da literatura de crescimento econômico, apresentando um breve histórico da emergência da PTF como fator explicativo para tanto o crescimento de longo prazo como para explicar as diferenças de renda entre os países, e uma revisão dos principais modelos e evidências relativas ao processo de difusão tecnológica entre países. A terceira seção contém nossa metodologia, com o modelo a ser usado, as técnicas de estimação empregadas e a base de dados utilizada. A quarta seção discute os resultados. E, por fim, a quinta e última seção apresenta nossas conclusões finais.

2. REFERENCIAL TEÓRICO

2.1. Modelo Neoclássico de Crescimento

A moderna teoria do crescimento econômico toma como ponto de partida as contribuições de Solow (1956, 1957) e Swan (1956). O modelo de Solow ou Solow-Swan foi uma das primeiras tentativas de formalizar a dinâmica do crescimento do produto e é importante por servir de base para diversos modelos posteriores. Este modelo é construído em torno de uma função de produção agregada com retornos constantes de escala:

$$Y(t) = F[A(t), K(t), L(t)] \quad (2.1)$$

Tal que t é tempo, Y é o produto agregado, K é o estoque de capital físico, L é o trabalho e A é um resíduo¹, também chamado por Produtividade Total dos Fatores (PTF). Como pressuposto, o trabalho cresce a uma taxa constante n , a PTF cresce a uma taxa constante g e a dinâmica do estoque de capital é descrita pela seguinte equação de movimento:

$$\dot{K}(t) = I(t) - \delta K(t) = sY(t) - \delta K(t) \quad (2.2)$$

Em linhas gerais, a formulação do modelo dá ênfase na forma como a acumulação de capital impacta no crescimento de longo prazo do produto por trabalhador, mas a acumulação de capital per si não gera crescimento, pois o seu rendimento é decrescente. Uma das conclusões do modelo é que o crescimento de longo prazo do produto por trabalhador de um país é explicado exclusivamente pela taxa de progresso tecnológico, g . Além disso, o nível de bem-estar é avaliado em termos de consumo no longo prazo e a variável chave para definir a trajetória que leva ao maior nível de consumo no estado estacionário é a taxa de poupança. A fim de refinar essa análise de bem estar, o modelo de Ramsey-Cass-Koopmans torna a poupança endógena e permite que se avalie o nível ótimo de bem-estar através de uma função de utilidade de um agente representativo. Outro refinamento no modelo neoclássico é a inclusão do capital humano.

O modelo neoclássico, então, coloca ênfase na acumulação de fatores para explicar a dinâmica da evolução da produtividade do trabalho e suas diferenças entre países. A fim de verificar quão bem o modelo neoclássico explica as variações de renda per capita entre países, Mankiw, Romer e Weil (1992) aplicam métodos de regressão em dados de *cross-section* para testar a capacidade preditiva do modelo de Solow

¹ O parâmetro A representa tudo o que não é explicado pelos insumos. É comum chamá-lo de tecnologia, mas na prática ele é bem mais abrangente que isso.

clássico e de sua versão aumentada pelo capital humano para três conjuntos de países entre 1960 e 1985². A especificação usada pelos autores segue os pressupostos básicos do modelo de Solow, com taxas de crescimento da tecnologia e depreciação constantes.

Os autores definem a tecnologia como:

$$A(t) = A(0)e^{gt} \quad (2.3)$$

$$A(0) = e^{a+\varepsilon_i} \Rightarrow \ln A(0) = a + \varepsilon_i \quad (2.4)$$

Onde g é o progresso técnico, $A(0)$ é o estoque de tecnologia inicial, composto por um parâmetro comum a todos os países, a , e um termo específico, ε_i . Note que $A(0)$ não capta apenas a tecnologia, mas outros elementos, tal como a dotação de recursos, instituições e outras coisas que podem diferir entre países. Como hipótese de identificação, supõe-se que os choques no termo específico independem da poupança, s , e da taxa de crescimento populacional, n . Isso permitiria estimar o modelo via mínimo quadrados ordinários em uma *cross-section*. Na estimação do modelo aumentado com capital humano, os autores constroem uma variável de capital humano que é definida como a fração de pessoas elegíveis ao mercado de trabalho com ensino médio completo.

Os resultados dão suporte para as predições do modelo de Solow de que, no longo prazo, o nível de produto por trabalhador está positivamente correlacionado com a taxa de poupança e negativamente correlacionado com o crescimento populacional, bem como teriam demonstrado uma forte capacidade preditiva do modelo, de onde as estimativas para o modelo de Solow aumentado com capital humano seriam capazes de explicar 78% da variância do produto por trabalhador entre os países. Sob essa perspectiva a acumulação de fatores de produção seria peça essencial para explicar a diferença entre os países.

Apesar dos resultados empíricos atraentes e concordância com as premissas do modelo de Solow aumentado pelo capital humano, Islam (1995) questiona a hipótese de identificação de Mankiw, Romer e Weil (1992), mostrando que a eficiência econômica é correlacionada com a taxa de poupança, violando uma das hipóteses básicas do modelo de mínimos quadrados e, portanto, enviesando as estimativas. Isso acaba dando um peso excessivo ao capital, pois desconsidera o efeito indireto da PTF sobre o mesmo. A conclusão do autor é que se deve controlar adequadamente a PTF de cada país, bem como esta é uma variável de grande relevância dentro de cada país. Logo, as

² Os três conjuntos são: (a) todos os países da base disponível em Summers e Herston (1988), excluídos aqueles cuja produção petrolífera é a indústria dominante; (b) países dentro da amostra anterior que possuíam mais do que um milhão de habitantes; e (c) os países da OCDE que tinham mais de um milhão de habitantes.

diferenças de eficiência tornam-se fundamentais na determinação dos diferenciais de renda per capita entre países.

Outra leva de críticas contra Mankiw, Romer e Weil (1992) vem sobre a forma como o autor constrói a série de capital humano. Klenow e Rodríguez-Clare (1997) consideram outras medidas alternativas de capital humano, que julgam serem mais robustas por levarem em conta o setor primário e terciário da educação, e comparam com o modelo original de Mankiw, Romer e Weil (1992) através de uma série de exercícios de decomposição do desenvolvimento a fim de verificar a contribuição da PTF e dos demais fatores para explicar a variação no produto por trabalhador entre os países.

Nesses exercícios os autores chegam à conclusão de que dependendo da medida de capital humano utilizado e calibração, a capacidade explicativa dos fatores diminui bastante. Se no estudo original Mankiw, Romer e Weil (1992) encontraram que ambos os tipos de capital (físico e humano) eram capazes de explicar 78% da variação no PIB por trabalhador, isso praticamente se inverte para outras medidas de capital humano e diferentes calibrações. Por exemplo, quando se supõe que L é o número de trabalhadores (ao invés da população economicamente ativa) e incluído todos os níveis de ensino (ao invés de apenas o ensino secundário) na medida de capital humano, o poder explicativo dos insumos para explicar as variações no PIB per capita cai para 40%, dando, assim, um papel proeminente na PTF para explicar a diferença entre os países, que passa então a explicar 60% da variação de produto por trabalhador.

Exercícios semelhantes, mas com outras bases de dados também colocam um papel importante na PTF para explicar as grandes diferenças de produtividade entre os países, sendo esta tão ou mais importante para explicar a variação de renda per capita entre os países. Veloso et. al. (2013), por exemplo, realizam exercícios de decomposição do desenvolvimento – decomposição que permite calcular a contribuição relativa da PTF e dos fatores de produção sobre a variação no produto por trabalhador. Eles apresentam duas formas alternativas de representação: a forma tradicional, que usa a razão capital-trabalho como medida de capital físico, e também uma forma alternativa que considera a medida de capital como a razão capital-produto (utilizada também nos exercícios de Klenow e Rodríguez-Clare (1992)). Os resultados calculados por eles, utilizando dados da Penn World Table 7.0³ estão expostos na Tabela 1⁴.

³ Os resultados abrangem 189 países entre 1950 e 2009, os dados são calculados segundo o conceito da paridade de compra.

Tabela 1 - Decomposição do desenvolvimento (em %)

Contribuição Tradicional para a Variação de Y/L		
	A	Fatores
1990	47%	53%
2000	48%	52%
2009	48%	52%
Contribuição Alternativa para a Variação de Y/L		
1990	77%	23%
2000	79%	21%
2009	79%	21%

Fonte: Veloso et. al. (2013).

Aiyar e Feyrer (2002) e Hall e Jones (1999) também fazem este exercício. Ambos encontram, assim como na tabela anterior, que a maior parte da diferença de PIB por trabalhador entre os países é devido a diferenças de produtividade. Inclusive, Jones e Romer (2009) incluem dentre os fatores estilizados na recente literatura de crescimento as grandes diferenças de renda e PTF entre países, argumentando que a PTF é quantitativamente ao menos tão importante quanto os fatores de produção. As diferenças de renda e PTF são bem correlacionadas, assim, países pobres não seriam pobres apenas porque tem pouco capital e educação, mas também por usarem os recursos que tem de maneira menos eficiente.

Prescott (1998) também avalia a contribuição dos diferentes tipos de capital e sua capacidade de explicar as grandes variações de produto per capita entre os países, este também chega à mesma conclusão. Conforme o autor, a teoria neoclássica de crescimento (modelo de Solow e extensões de Ramsey-Cass-Koopmans) não constitui uma teoria de desenvolvimento, por ser incapaz de explicar as grandes diferenças de produto por trabalhador entre os países. Para mostrar isso, são feitos vários exercícios de calibração onde o autor tenta mostrar que seria necessário assumir valores irrealistas para certos parâmetros para que o modelo neoclássico realmente explicasse satisfatoriamente as diferenças observadas. Se for adicionado o capital intangível no modelo de Solow, o investimento nesse tipo de capital teria que ser tão grande quanto o próprio PIB para que as diferenças de poupança dessem conta das diferenças de produto por trabalhador. Mesmo o modelo com capital humano é insuficiente enquanto teoria de desenvolvimento, na medida em que a fração alocada para investimento de capital

⁴ Na decomposição tradicional, o produto por trabalhador é escrito como:

$$y = Ak^\alpha h^{1-\alpha}$$

Onde y é o produto por trabalhador, k é o capital físico por trabalhador e h é o capital humano por trabalhador. Na decomposição alternativa é utilizada a razão capital-produto, assim,

$$y = A^{\frac{1}{1-\alpha}} (k/y)^{\frac{\alpha}{1-\alpha}} h$$

Na decomposição do desenvolvimento, os fatores são calculados como

$$Cov[\ln(y), \ln(Z)] / Var[\ln(y)]$$

Onde Z é a variável explicativa multiplicada por seu coeficiente. No exercício $\alpha = 0.4$.

humano teria que ser maior do que o tempo alocado no mercado para que as diferenças fossem tão significativas quanto o esperado. O modelo neoclássico, conclui o autor, só pode dar conta dessas diferenças se a PTF variar bastante entre os países.

O reconhecimento da importância da PTF para explicar as grandes desigualdades de renda serviu de motivação para que se buscassem modelos mais adequados para explicar os mecanismos internos da PTF. Prescott (1998), por exemplo, conclui que uma teoria para a PTF é necessária. Nessa linha, há pelo menos dois corpos de pesquisa delineados no campo teórico: (i) a teoria do crescimento endógeno, que explica o crescimento de longo prazo a partir de modelos de inovação; e (ii) o problema da adoção tecnológica e diferenças entre PTFs. Essas duas questões serão tratadas a seguir.

2.2. Modelos de Crescimento Endógeno

Para modelar adequadamente e inserir uma lógica interna para a produtividade, dois pressupostos do modelo neoclássico são postos em questão: (i) o pressuposto de rendimentos decrescentes no capital e (ii) a hipótese de concorrência perfeita. Se for possível supor retornos constantes ou crescentes para o capital, então seria possível ter crescimento sustentado explicado pelo próprio modelo. Além disso, uma vez que o conhecimento é um bem não rival, para que haja incentivos para investir na produção de inovações é necessário relaxar a hipótese de competição perfeita (FERREIRA E ELLERY JR., 1996).

Romer (1986) formula um modelo de crescimento endógeno em que se supõe a existência de externalidades positivas geradas individualmente pelas firmas ao investirem em capital físico. A tecnologia é detonada por $\Psi(E)$, onde $\Psi'(E) > 0$ e E representa o estoque de ideias na economia, que é produzido como externalidade da acumulação de capital. A firma, ao produzir novos bens também despense esforços em pesquisa e desenvolvimento (P&D), logo E segue os movimentos do capital. Sob essa hipótese, E é uma função monótona e positiva de K . Isso permite separar o capital em capital privado e capital social, sendo este último um bem público. Note que, como se trata de um bem público, há a possibilidade de transbordamento para todos agentes da economia, bem como para os agentes de outros países (FERREIRA E ELLERY JR., 1996).

Todavia, o modelo desconsidera os fundamentos que explicariam a racionalidade de uma empresa investir em P&D. O processo de P&D é custoso e o

conhecimento é um bem público, então não há incentivos para que haja uma produção eficiente de ideias num ambiente competitivo. Essa característica do mercado de ideias justifica em grande parte a existência de patentes e outras formas de proteção à propriedade intelectual. Esses direitos, porém, surgem exatamente para dar monopólio ao criador das invenções sobre seu produto. Portanto, adequar os modelos de crescimento endógeno aos incentivos do processo de inovação exige um distanciamento do pressuposto de competição perfeita.

Na busca por maior fundamentação para os modelos de crescimento endógeno, surge no início dos anos 90 uma série de novos modelos de crescimento endógeno com competição imperfeita. Romer (1990) considera uma economia com três setores: um setor de pesquisas, produção de bens intermediários e outro de bens finais. Nos modelos de Romer (1990) e Grossman e Helpman (1991) os insumos novos se somam aos antigos e o crescimento se dá por uma maior divisão do trabalho. Para dar mais realismo aos modelos, Aghion e Howitt (1990) sugerem que existe um processo de ‘destruição criativa’, onde os insumos antigos são eventualmente substituídos por novos, devido à superioridade dos novos insumos (FERREIRA E ELLERY JR., 1996).

2.3. Difusão Tecnológica e Modelo para PTF

Os modelos de crescimento endógeno, entretanto, só conseguem explicar o deslocamento da fronteira tecnológica, mas não conseguem dizer muito sobre as grandes diferenças de tecnologia entre os países. Se o conhecimento é um bem público e a informação possui grande mobilidade no mundo, então seria de se esperar que todos os países se beneficiassem direta ou indiretamente da produção de uma inovação, independente do país de origem. Dado esse efeito de transbordamento tecnológico, a questão que se põe é o que leva um país a adotar novas tecnologias e se a existência de barreiras à adoção é suficiente para explicar a diferença observada entre as PTFs (GONÇALVES, 2013).

Para responder essa questão, Prescott (1998) sugere que as diferenças de produtividade total dos fatores não são devidas ao estoque de conhecimento usável, mas sim às práticas de trabalho e a resistência em adotá-las ou não. O autor ilustra isso com dois casos específicos: na indústria têxtil e na mineração de carvão. O primeiro se refere às fábricas de algodão na Índia e Japão entre 1920 e 1938. A produtividade do trabalho no Japão aumentou 120%, enquanto na Índia esse aumento foi de apenas 40% no período citado. Segundo Wolcott (1994 *apud* PRESCOTT, 1998), essa diferença de

produtividade não foi por conta de melhores equipamentos, mas sim devido à capacidade de resistência dos empregados em adotar certas práticas de trabalho. No caso, os trabalhadores japoneses tinham menos incentivos e habilidade para resistir às tentativas de aumentar o número de máquinas operadas por trabalhador. O que explicaria essa diferença é que, no Japão, os trabalhadores da indústria têxtil eram em geral garotas analfabetas de 15 a 18 anos, que largavam esses empregos logo após se casar, e quaisquer tentativas de greve resultavam na deportação dessas trabalhadoras de volta para suas respectivas províncias. Em resumo, a capacidade de mobilização e incentivos para tal era limitado entre essas trabalhadoras. Ao contrário, na Índia a maior parte dos trabalhadores no setor têxtil eram homens adultos que ficavam a vida inteira naquele emprego, tendo, portanto, maiores incentivos para resistir a pressões de se impor tecnologias poupadoras de trabalho.

A segunda evidência anedótica se refere às minas de carvão nos Estados Unidos entre 1949 e 1994. Segundo o autor, esse setor tinha um sindicato bastante poderoso nos primeiros anos do período, e essa capacidade de mobilização teria sido sentida na produtividade. O autor cita, por exemplo, a introdução da tuneladora (*boring machine*), que só foi introduzido no setor entre 1949 e 1969. A tecnologia já existia antes, mas os contratos com o sindicato proibiam explicitamente seu uso. Ela só teria sido liberada devido à diminuição do preço do petróleo, aumentando a competitividade e pondo as empresas do setor de carvão em risco se não aumentassem sua produtividade. Durante o período o autor mostra que a produtividade do trabalho esteve correlacionada com as variações nos preços de bens substitutos e entrada de novos competidores. Por exemplo, entre 1978 e 1994 a produtividade cresceu como resultado da queda no preço do petróleo e aumento da competição com minas a céu aberto, que não eram sindicalizadas.

Prescott (1998) também nota que há uma grande variação de produtividade entre setores mesmo em países com níveis de educação e capital por trabalhador similares, logo a incapacidade dos trabalhadores em adotar determinadas técnicas, ainda que ocorra em alguns casos, não explica toda a história. Além do conflito distributivo no mercado de trabalho, o autor também ressalta que leis específicas de certos países podem impedir a adoção de certas tecnologias redutoras de trabalho.

Sobre esse argumento, Parente e Prescott (2000 *apud* GONÇALVES, 2013) tentam modelar essa ideia supondo um custo de transação exógeno, de maneira que, dentro de certos intervalos e parâmetros, se torna possível explicitar satisfatoriamente a grande diferença entre as PTFs. Esta explicação, porém, não parece ser possível num

ambiente competitivo. Se fosse possível a entrada de concorrentes com a nova tecnologia, a empresa já estabelecida não teria como barrar a adoção da mesma tecnologia sem colocar em risco a própria sobrevivência, que em última instância serviria de pressão suficiente para conseguir a cooperação dos sindicatos e trabalhadores. Com isso, torna-se necessário trabalhar dentro de um contexto de competição imperfeita e com significativas barreiras à entrada. Nesse sentido, a tese de Prescott (1998) e Parente e Prescott (2000) levam a duas prescrições: (i) a necessidade de incentivos à entrada no mercado de bens (competição); e (ii) uma maior flexibilização no mercado de trabalho. A segunda prescrição consiste em facilitar o processo de contratações e demissões, a fim de desincentivar esse tipo de conluio.

Gonçalves (2013) ressalta também a interação entre o poder político e econômico nessa discussão. Note que, a princípio, os trabalhadores poderiam usar seu poder de barganha para simplesmente aceitar a nova tecnologia e trabalhar menos horas ao invés de barrar aquela tecnologia. Todavia, isso pode não ser a decisão mais racional num contexto intertemporal, pois a menor importância do trabalho no processo produtivo tem efeitos negativos sobre o poder de barganha dos trabalhadores. Acemoglu e Robinson (2000 *apud* GONÇALVES, 2013) sugerem que essa lógica tem um pano de fundo mais geral, de maneira que grupos politicamente organizados têm incentivos para impedir reformas pró-crescimento na medida em que as mesmas possam futuramente minar seu próprio poder político, devido ao empoderamento econômico e (posteriormente) político de outros atores.

Para entender melhor e de maneira mais geral as maneiras pelas quais a tecnologia difere entre países, Acemoglu (2009) apresenta um conjunto de modelos para explicar o processo de difusão tecnológica. O ponto de partida é o modelo com crescimento exógeno à la Solow com tecnologia Horrod-Neutra e função de produção por trabalho agregada

$$y_j(t) = A_j(t)f(k_j(t)) \quad (2.5)$$

Onde y é o produto por trabalhador, A é a tecnologia, $f(.)$ é a função de produção e k é o capital por trabalhador efetivo. O subscrito j denota o j -ésimo país (de um conjunto de J países) e t é o tempo.

A equação de movimento do capital é a mesma do modelo de Solow com uma taxa de depreciação, δ , igual para todos os países. A diferença nesse modelo em relação

ao modelo do modelo de Solow é que, além de serem consideradas várias economias, o modelo supõe uma equação de movimento para a tecnologia de cada país.

Seja $g_j = \dot{A}_j(t)/A_j(t)$ a taxa de crescimento da tecnologia de j em t e suponha a existência de uma *fronteira tecnológica* denotada por $A(t)$, que cresce exogenamente a uma taxa constante $g = \dot{A}(t)/A(t) > 0$ e $A(0) > 0$. Essa fronteira denota o conhecimento máximo que qualquer país pode ter em dado período de tempo, logo $A(t) \geq A_j(t)$ para todo j e t . A partir disso, a equação de movimento da tecnologia de cada país é dada por

$$\dot{A}_j(t) = \sigma_j [A(t) - A_j(t)] + \lambda_j A_j(t) \quad (2.6)$$

Onde $\sigma_j \in (0, \infty)$ e $\lambda_j \in [0, g), \forall j = 1, \dots, J$. Sob essa especificação, a evolução da tecnologia de cada país dependerá do quanto ela consegue absorver da fronteira tecnológica e do que ela produz localmente. Cada país é capaz de absorver tecnologia da fronteira a uma taxa de absorção exógena σ_j e a tecnologia produzida localmente é construída em cima do estoque de conhecimento presente, onde λ_j captura a rapidez com que isso ocorre. Uma coisa a se notar é que países mais distantes da fronteira tecnológica crescerão relativamente a taxas maiores, pois o primeiro termo do lado direito da expressão será tão maior quanto menor for o estoque de tecnologia do país em relação à fronteira.

Definindo $a_j = A_j(t)/A(t)$ como a medida inversa da distância de j da fronteira, podemos reescrever (2.6) como:

$$\dot{a}_j(t) = \sigma_j - (\sigma_j + g - \lambda_j) a_j(t) \quad (2.7)$$

Cuja condição inicial é $a_j(0) > 0$. Para o modelo como um todo, cada país tem uma lei de movimento para o capital e outra para a_j , formando $2J$ equações no sistema para caracterizar a distribuição de tecnologia e renda no mundo e sua dinâmica de transição.

O equilíbrio de estado estacionário do mundo é descrito pela alocação $\{k_j(t), a_j(t)\}_{j=1}^J$ tal que a lei de movimento do capital e lei de movimento da tecnologia são satisfeitos para cada país começando a partir das condições iniciais $\{k_j(0), a_j(0)\}_{j=1}^J$. Já o equilíbrio de estado estacionário do mundo é aquele em que $\dot{k}_j(t) = \dot{a}_j(t) = 0$ para cada país. Esse estado estacionário, porém, exhibe crescimento constante, sendo, portanto, também uma trajetória de equilíbrio de crescimento balanceado.

Existe um único equilíbrio de estado estacionário mundial e ele é globalmente estável. E mesmo com diferenças nas taxas de poupanças e absorção de tecnologia,

todos os países convergem para a mesma taxa g . Segundo o autor, isso acontece devido ao efeito de *catch up* proporcionado pelo fato de que a taxa de absorção da fronteira é maior quanto mais distante da fronteira aquele país estiver, e esse efeito é forte o suficiente para fazer com que todos os países convirjam para a mesma taxa de crescimento. Todavia, ainda persistirão diferenças no nível de produto por trabalhador entre os países devido às diferenças nas taxas de poupança e absorção de tecnologia. Esse modelo base, no entanto, desconsidera quaisquer outras interações entre países. O nível de renda per capita no estado estacionário depende apenas do comportamento da fronteira tecnológico e dos parâmetros específicos de cada país. Esses resultados, pelo menos em termos qualitativos, não dependem da hipótese de taxa de poupança exógena, os resultados se mantêm mesmo quando se incorpora a decisão de consumo das famílias.

O passo seguinte consiste em incluir crescimento endógeno nesse modelo de difusão. Para isso, o autor considera primeiro o caso em que a taxa de crescimento da PTF mundial é exógena e em seguida a torna endógena também.

Acemoglu (2009) parte de um modelo simples de inovação, onde todo o esforço de Pesquisa e Desenvolvimento (P&D) é voltado para a expansão da variedade de insumos ou máquinas usadas na produção. Os ganhos de produtividade advêm, portanto, da maior divisão do trabalho – esse modelo simples é chamado de “modelo de laboratórios-equipamentos”, pois são necessários apenas investimentos em equipamentos e laboratórios, em contraposição ao nível de habilidade dos trabalhadores e cientistas, para gerar crescimento. Nesse modelo a função de produção agregada para a j -ésima economia no tempo t é dada por:

$$Y_j(t) = \frac{1}{1-\beta} \left(\int_0^{N_j(t)} x_j(v, t)^{1-\beta} dv \right) L_j^\beta \quad (2.8)$$

Onde L_j é o trabalho, que nessa especificação é constante no tempo, $N_j(t)$ é o número de máquinas disponíveis no país j no tempo t e $x_j(v, t)$ é o montante de máquinas do tipo v sendo usadas em t . Apenas por questão de simplicidade, pressupõe-se que cada máquina se deprecia totalmente após o uso. Além disso, cada variedade de tecnologia é de propriedade de um monopolista, que vende as suas máquinas ao preço $p_j^x(v, t)$ e o custo de produção é $\psi = 1 - \beta$ unidades do bem final. As diferenças de produtividade surgem do fato de que cada país só pode usar as tecnologias ofertadas pelo monopolista de seu próprio país.

A restrição orçamentária de cada país é dada por:

$$C_j(t) + X_j(t) + \zeta_j Z_j(t) \leq Y_j(t) \quad (2.9)$$

Onde $X_j(t)$ é o investimento nos insumos, $Z_j(t)$ é o gasto com adoção de tecnologia e $C_j(t)$ é o consumo agregado. O parâmetro ζ_j permite a existência de diferenças no custo de adoção da tecnologia de outros países, resultado de barreiras institucionais, políticas de fomento a pesquisa, etc.

Nesse caso a lei de movimento para a expansão do número de máquinas disponíveis é dada por:

$$\dot{N}_j(t) = \eta_j \left(N(t)/N_j(t) \right)^\phi Z_j(t) \quad (2.10)$$

Onde $\eta_j > 0$ para todo j e $\phi > 0$. O termo $N(t)$ é entendido como a fronteira tecnológica da expansão de variedades, que cresce a uma taxa constante exógena g . Tanto o estoque inicial de tecnologia, $N_j(0)$, quanto g são positivos.

Uma implicação desse modelo é que o fluxo de lucros do monopolista em t na economia j é dado por $\pi_j(t) = \beta L_j$. Assim, para uma taxa de juros constante, r_j^* , positiva e supondo existência de um equilíbrio de estado estacionário, o valor presente descontado líquido de uma nova máquina será $V_j^* = \beta L_j / r_j^*$. Além disso, se no estado estacionário todos os países crescerem a mesma taxa, então $N_j(t)$ também irá crescer a taxa g e o *gap* relativo $N_j(t)/N(t)$ se manterá constante num nível μ_j^* . Para que haja livre entrada, então teremos que ter:

$$\mu_j^* = (\eta_j \beta L_j / \zeta_j r_j^*)^{1/\phi} \quad (2.11)$$

Desse modo, países com melhores possibilidades de inovação (η_t alto) ou menores gastos com P&D (ζ_j baixo) terão um μ_j alto e conseqüentemente serão tecnologicamente mais avançados e prósperos. Além disso, vale notar que o tamanho da força de trabalho também impacta positivamente sobre a renda do país, pois isso se reflete numa maior demanda por novas máquinas e torna o processo de P&D mais lucrativo.

Mesmo com essas mudanças, os padrões observados de equilíbrio são similares ao do modelo *benchmark*: há apenas um estado estacionário e ele é globalmente estável em ponto de sela. O que muda de um modelo para o outro é que agora se torna possível identificar os fatores que afetam a taxa de adoção tecnológica e sua relação com o comportamento das firmas.

Para explicar o processo de mudança tecnologia a nível global, o autor se restringe para as forças mantendo os países crescendo a taxas similares. Para realizar a

mudança, o pressuposto de que $N(t)$ cresce a uma taxa constante g deve ser trocado por uma equação que combine de alguma forma todos os incrementos tecnológicos de cada país numa tecnologia global. Para ilustrar isso, Acemoglu (2009) considera o caso em que essa agregação é dada pela média aritmética da tecnologia dos diferentes países:

$$N(t) = (1/J) \sum_{j=1}^J N_j(t) \quad (2.12)$$

Vale notar que sob essa especificação o que temos não é mais a fronteira tecnológica, mas sim a tecnologia média do mundo. Isso, conforme Acemoglu (2009), não implica em grandes problemas e é uma generalização natural do que foi apresentado até agora. Todavia, é importante notar que a escolha da média não é necessariamente a melhor forma de agregar as tecnologias, pois implica em dizer que todos os países contribuem para esse componente de maneira igual. Ainda assim, o autor defende que essa ou outra forma de agregação não tem impactos qualitativos relevantes sobre os principais resultados. Sobre esses resultados, eles são similares aos anteriores. A diferença aqui é que o equilíbrio de estado estacionário onde todas as tecnologias e a tecnologia mundial crescem a taxa g irá exigir a restrição de que $(1/J) \sum_{j=1}^J \mu_j^* = 1$, ou, alternativamente, que $(1/J) \sum_{j=1}^J (n_j \beta L_j / \zeta_j (\rho + \theta g))^{1/\phi} = 1$. Mesmo com essas alterações ainda é possível estabelecer a estabilidade local para o equilíbrio de estado estacionário.

Esses modelos até agora apenas consideram as diferenças como resultado de um processo lento de difusão tecnológica, mas também é possível pensar essas diferenças como persistentes mesmo quando há um livre fluxo de ideias. Outras possibilidades incluem, por exemplo, a possível distinção entre tecnologias apropriadas e inapropriadas, partindo da ideia de que algumas tecnologias são pensadas para as necessidades específicas do país que a desenvolveu, mas que não se aplicariam com a mesma eficácia em outro país. Um exemplo desse tipo de modelo é o modelo de Basu e Weil (1998), em que uma nova tecnologia é desenvolvida por países com produção intensiva em capital, de modo que essa tecnologia resultará nos mesmos ganhos de eficiência aos países com baixa razão capital-trabalho. A introdução da presença de tecnologias adequadas e inadequadas tem potencial para fazer com que isso resulte numa diferença de produto por trabalhador muito maior.

Outra forma de considerar a questão da tecnologia ser apropriada ou não é partindo da ideia de que algumas tecnologias exigem certas habilidades presentes em trabalhadores de maior qualificação. Para ilustrar esse caso, o autor apresenta o modelo

de Acemoglu e Zilibotti (2001), que considera a existência de um descompasso entre tecnologias desenvolvidas em economias avançadas e a habilidade da força de trabalho em países menos desenvolvidos. O modelo distingue os países entre norte e sul: os países do norte investem em P&D e têm uma maior proporção de trabalhadores qualificados, enquanto o sul apenas copia as tecnologias do norte. No equilíbrio, as duas tecnologias irão afetar os dois países de maneira diferente, beneficiando os países do norte relativamente mais do que os do sul. Isto porque a tecnologia é desenvolvida no norte e desenhada para suas necessidades, elas acabam sendo inapropriadas para o sul, que resulta em diferenças de renda entre os dois grupos de países.

Outra potencial causa para diferenças na tecnologia e sua adoção são instituições. Para isso, o autor toma como base uma versão simplificada do modelo de Acemoglu, Antras e Helpman (2007) e apresenta um modelo explicando a forma como as instituições de contrato podem afetar a relação entre produtores e ofertantes, alterando a lucratividade da adoção tecnológica. A principal conclusão do modelo é que fricções contratuais causam escassez de investimento em qualidade, desencorajando a adoção de novas tecnologias e consequentemente reduzindo a produtividade. O trabalho de Acemoglu, Antras e Helpman (2007) também mostra que diferenças modestas de instituições de contratos podem resultar em diferenças de tecnologia e renda bastante grandes (ACEMOGLU, 2009).

2.4. Tecnologia e Capital Humano

Os modelos apresentados por Acemoglu (2009) focam em três causas para as diferenças de produtividade: (i) difusão tecnológica lenta; (ii) tecnologias apropriadas para certos contextos e não para outros; e (iii) instituições de contrato. Sobre o primeiro ponto levantado, o autor não modelou, mas apontou que duas potenciais causas para que alguns países tenham maiores ou menores dificuldades para absorver a tecnologia produzida pelo mundo: (a) existência de barreiras para a adoção tecnológica, na linha do argumento apresentado por Parente e Prescott (2000); ou (b) devido à possível existência de um *link* entre o capital humano e a produtividade, como sugerido por Nelson e Phelps (1966). Eles levantam o argumento de que empreendedores mais bem educados têm maior facilidade para adotar novas tecnologias. A explicação é que o empreendedor bem educado tem maior probabilidade de tomar conhecimento e está mais apto a avaliar as inovações tecnológicas próximas da fronteira. Em consequência disso, eles teriam mais interesse em adotar essas novas tecnologias e com mais rapidez.

Em contrapartida, um empresário menos educado será mais conservador, preferindo esperar até que essa tecnologia esteja mais difundida e consolidada para ter certeza de sua eficácia.

Sob essa hipótese relativa ao capital humano, a seguir serão apresentados dois modelos que Nelson e Phelps (1966) usam para sustentar essa argumentação. Ainda sob essa hipótese, Aiyar e Feyrer (2002) constroem um modelo inspirado em Nelson e Phelps (1966) e avaliam essa hipótese empiricamente para um conjunto de países heterogêneos. Por fim, também é apresentado o modelo de Damsgaard e Krusell (2010), que busca explicar a distribuição de PTF entre os países, levando em consideração a hipótese de Nelson-Phelps.

Se essa hipótese é razoável, então o capital humano tem um efeito indireto sobre a própria taxa de crescimento da produtividade. Para mostrar isso, Nelson e Phelps (1966) postulam dois modelos de difusão de tecnologia a partir da seguinte função de produção agregada:

$$Q(t) = F[K(t), A(t)L(t)] \quad (2.13)$$

$$T(t) = T(0)e^{\lambda t}, \quad \lambda > 0 \quad (2.14)$$

Onde t é o tempo, Q é o produto, K é o capital, L é o trabalho, A é o índice de tecnologia praticada e T é o nível teórico de tecnologia⁵. Pela segunda equação T cresce exogenamente a uma taxa constante λ .

No *primeiro modelo* a defasagem temporal entre a criação de uma nova técnica e sua adoção é função crescente de algum índice educacional do indivíduo em posição de inovar, denotado h . Seja w a defasagem, então:

$$A(t) = T(t - w(h)), \quad w'(h) < 0 \quad (2.15)$$

Isto é, o nível de tecnologia em prática é igual ao nível teórico de w anos atrás, e o tamanho da defasagem é função decrescente do capital humano.

Substituindo (2.31) em (2.30):

$$A(t) = T(0)e^{\lambda[t-w(h)]} \quad (2.16)$$

Dessa formulação temos que A e T crescem à mesma taxa λ e o nível ou caminho da tecnologia é crescente no nível de capital humano, pois quanto maior o nível, menos distante a PTF estará de seu nível teórico. A implicação dessa versão é que o retorno da educação será tão maior quanto mais rápido for o crescimento de T .

⁵ Nelson e Phelps (1966) definem o nível teórico de tecnologia como o nível de tecnologia que prevaleceria se a difusão tecnológica fosse completamente instantânea (i.e. não existem barreiras para adoção tecnológica).

Essa especificação, apesar de simples, ignora a lucratividade da nova técnica não introduzida, então para dar um pouco mais de realismo os autores propõe um *segundo modelo*, especificando a seguinte equação de movimento:

$$\dot{A}(t) = \phi(h)[T(t) - A(t)], \quad \phi(0) = 0, \quad \phi'(h) > 0 \quad (2.17)$$

Nessa hipótese, a própria taxa de crescimento da tecnologia prática é função crescente do capital humano e do *gap* entre a tecnologia teórica e prática como proporção de $A(t)$. Pelas equações desse modelo, no longo prazo a tendência é que a taxa da tecnologia em prática, \dot{A}/A , seja igual a λ . Isto porque quando $\dot{A}/A > \lambda$ o *gap* entre T e A irá diminuir, diminuindo \dot{A}/A até λ . Note também que o *gap* de equilíbrio será $\lambda/\phi(h)$, e, portanto, este é uma função decrescente do nível educacional. Como resultado, um aumento no nível de capital humano tem como resultado a elevação da trajetória da tecnologia em prática no longo prazo.

Assim como no modelo anterior, neste a produtividade marginal da educação é função crescente na taxa de crescimento da tecnologia teórica, λ e positiva somente se $\lambda > 0$. Outro resultado do modelo é que o *payoff* de um nível maior de educação será maior quanto mais tecnologicamente progressiva for essa economia.

Esses modelos são admitidamente simples, mas conseguem ilustrar bem a forma pela qual o nível de capital humano de um país pode influenciar o nível de produtividade daquele país.

O link proposto entre o estoque de capital humano e crescimento da PTF é avaliado empiricamente por Aiyar e Feyrer (2002). Para isso, eles calculam a PTF de 86 países entre 1960 e 1990 e buscam avaliar a presença de convergência condicional na PTF relativa ao capital humano. Para isso, foram aplicados estimadores GMM num contexto de painel dinâmico com efeitos fixos.

Para isso eles formulam um modelo que parte do pressuposto de que alguns países avançados investem em P&D, gerando *spillovers* tecnológicos para os demais países, e o sucesso em absorver tais transbordamentos depende do estoque de capital humano daquele país. Dessa forma, os autores assumem que o estoque de capital humano impacta a absorção de *spillovers* de uma maneira dinâmica. Isto é, um incremento no estoque de capital humano em t impacta a PTF em t e em todos os demais períodos subsequentes. Para isso, é formulado um modelo com três suposições em relação aos *spillovers*: (a) para a maior parte dos países a expansão da fronteira tecnológica é tomada como exógena; (b) os países tiram vantagem desses *spillovers* de

maneira diferente um do outro e sua capacidade de fazer isso depende do estoque de capital humano; e (c) há convergência condicional ao nível de capital humano na PTF.

A PTF é construída tomando como base a especificação de Hall e Jones (1999), onde a tecnologia é Horrod-Neutro e $\alpha = 1/3$. O modelo trabalhado pelos autores parte da suposição de que a taxa de crescimento da PTF em um dado momento é positivamente relacionado com o *gap* entre o nível observado de PTF e seu nível *potencial* naquele dado período, assim:

$$\frac{\dot{A}}{A}(t) = \lambda(\log A^*(t) - \log A(t)) \quad (2.18)$$

Onde $A(t)$ é o nível de PTF e $A^*(t)$ é o nível potencial de PTF, e os autores sugerem que esse nível potencial é determinado da seguinte forma:

$$A^*(t) = Fh^\phi T(t) \quad (2.19)$$

Onde F é um índice de efeitos fixos, h é o capital humano por trabalhador e $T(t)$ é um índice de tecnologia que cresce exogenamente ao longo do tempo e tem um valor positivo. Dessa especificação se depreende também que, conquanto o nível de capital humano seja monotonicamente crescente, todas as PTFs observadas estarão num nível abaixo de seu potencial. Ressalta-se, também, que λ é entendida com o coeficiente de convergência condicional.

Dito isso, são feitas algumas transformações em (2.18) e (2.19) para chegar a especificação da regressão⁶:

$$a_{i,t} = f_i + \rho a_{i,t-1} + \beta x_{i,t-1} + \eta_t + u_{i,t} \quad (2.20)$$

Onde $a_{i,t} = \log A(t_2)$, $a_{i,t-1} = \log A(t_1)$, $x_{i,t-1} = \log h(t_1)$, $f_i = (1 - e^{-\lambda\tau}) \log F$, $\beta = \phi(1 - e^{-\lambda\tau})$, $\rho = e^{-\lambda\tau}$, $\eta_t = e^{-\lambda t_2} \int_{t_1}^{t_2} e^{\lambda t} \log T(t) dt$ e $\tau = t_2 - t_1$. Note que η_t é interpretado como uma tendência comum a todos os países, captando assim o deslocamento da fronteira tecnológica ou comportamento global da PTF. Essa especificação é importante porque, como os autores comentam, esse modelo pode não ser adequado para os principais países inovadores do mundo, já que implicitamente se parte do pressuposto que cada país é incapaz de influenciar esse componente global, e, por isso, tomam este como dado.

⁶ Os autores rearranjam as equações (2.34) e (2.35), multiplicam ambos os lados por $e^{\lambda t}$ e depois integram de ambos os lados entre t_1 e t_2 :

$$\int_{t_1}^{t_2} e^{\lambda t} ([\log A(t)] + \lambda \log A(t)) dt = \log F \int_{t_1}^{t_2} \lambda e^{\lambda t} dt + \phi \log h \int_{t_1}^{t_2} \lambda e^{\lambda t} dt + \int_{t_1}^{t_2} \lambda e^{\lambda t} \log T(t) dt$$

Resolvendo essas integrais e multiplicando o resultado por $e^{-\lambda t_2}$, chega-se ao resultado final:

$$\log A(t_2) = e^{-\lambda\tau} \log A(t_1) + \phi(1 - e^{-\lambda\tau}) \log h + (1 - e^{-\lambda\tau}) \log F + e^{-\lambda t_2} \int_{t_1}^{t_2} e^{\lambda t} \log T(t) dt$$

Que é a equação (2.20).

Para a aplicação do GMM em painel dinâmico, duas transformações são aplicadas sob (2.20). Na primeira transformação cada variável é subtraída de sua média no período, $\tilde{x}_{i,t} = x_{i,t} - \sum_{i=0}^N x_{i,t}/N$, a fim de remover a tendência global, η_t . Depois disso é tomada a primeira diferença de cada variável, a fim de remover o efeito fixo, f_i .

Os resultados obtidos confirmam que o estoque de capital humano tem um impacto positivo e significativo sobre a trajetória de crescimento de longo prazo da PTF. Países com maiores níveis de capital humano convergem para maiores níveis de produtividade. Segundo a melhor estimativa, os países convergem para essas trajetórias a uma taxa de 3% ao ano. Além disso, o nível de capital humano parece ter uma influência forte sobre a trajetória da PTF, de tal forma que a elasticidade da produtividade potencial relativo ao capital humano é de 4%.

Os autores também realizaram um exercício de decomposição da variância para a PTF potencial. Como é possível recuperar os efeitos fixos, é possível construir o nível potencial da PTF transformada de modo que a média entre países seja removida. Assim, a PTF potencial é escrita como:

$$\overline{\log A_{i,t}^*} = \overline{\log F_i} + \phi \overline{\log h_{i,t-1}} \quad (2.21)$$

Onde a barra sinaliza que essas variáveis foram subtraídas da média entre países. Em seguida é possível verificar quanto da variação na PTF potencial entre países é explicada pelos efeitos fixos e pelo capital humano. Aparentemente há estabilidade nas proporções explicadas por cada componente, e a proporção explicada pelo capital humano varia entre 87% e 96%. Esses resultados, tanto das estimações quanto da decomposição, porém, valem apenas para os países da amostra considerada que não pertencem à OCDE. Posteriormente os autores realizam estimações para a amostra toda e verificam que os parâmetros estruturais (λ e ϕ) não são significantes, indicando que o modelo descrito nesse trabalho é inapropriado para países que realizam esforços significativos em pesquisa e desenvolvimento e têm maior importância na determinação da fronteira tecnológica.

Ainda sob a hipótese de Nelson-Phelps, Damsgaard e Krusell (2010) também desenvolvem um modelo para a PTF, mas dessa vez com o objetivo de explicar a distribuição de produtividade no mundo. Idealmente, segundo eles, uma teoria de distribuição da PTF deve ter pelo menos duas características: (i) que o crescimento de longo prazo da PTF seja endógeno; e (ii) enquanto cada país pode influenciar seu nível de PTF relativa, eles não conseguem modificar a taxa de crescimento de longo prazo da

própria PTF. Isso parte do pressuposto de que cada país é pequeno e a taxa de crescimento de longo prazo de cada PTF é igual a da “PTF mundial”. Ou seja, a taxa de crescimento da PTF no mundo até pode ser endógena, mas ela é vista como exógena para cada país individualmente.

Supõe-se que cada país tem uma dotação de trabalhadores com baixa ou alta qualificação (*low-skilled e high-skilled*), denotados, respectivamente, por $L_{i,t}$ e $e_{i,t}$, onde os trabalhadores *low-skilled* são usados na produção de bens e serviços e são imóveis entre países, enquanto os trabalhadores *high-skilled* contribuem para a acumulação de PTF e há livre-mobilidade desse tipo de trabalhador entre os países. O total de trabalhadores altamente qualificados é fixo no mundo e denotado por e_W . Todos os países são simétricos e a dotação do número de trabalhadores pouco qualificados foi normalizado para 1, enquanto a dotação inicial do número de trabalhadores altamente qualificados é igual a média mundial, e_W .

Nessa especificação, a renda agregada do país i é definida:

$$Y_{i,t} = T_{i,t} - w_t e_{i,t} + w_t e_W \quad (2.22)$$

Onde w_t é o salário mundial do trabalhador altamente qualificado e $T_{i,t}$ é o nível de PTF. Nesse modelo os custos com investimento consistem nos salários pagos aos trabalhadores qualificados estrangeiros. Todos os países acumulam a PTF da mesma forma, ainda que o nível de PTF inicial seja diferente entre os países. O investimento na PTF é definido por um planejador central.

A dinâmica da PTF é inspirada na especificação de Nelson e Phelps (1966), de maneira a depender do *gap* relativo entre a PTF do país i e do mundo e também do investimento na PTF, assim:

$$\frac{T_{i,t+1}}{T_{i,t}} = \left(\frac{\bar{T}_t}{T_{i,t}} \right)^\gamma H(e_{i,t}) \quad (2.23)$$

Onde $\gamma \in [0,1]$, \bar{T}_t é a média mundial ou fronteira tecnológica. Como todos têm acesso a essa fronteira, naturalmente sua capacidade de absorver dependerá de sua distância relativa a essa variável, o parâmetro γ dá a força desse efeito de *catch up* no modelo. Os investimentos na PTF são dados por $H(e_{i,t})$, que é uma função estritamente crescente e estritamente côncava em $e_{i,t}$. Além disso, $H(0) = 1$; $H'(0) = \infty$; e $H \in [0, b]$.

Para atender esses pressupostos, os autores supõem a seguinte forma funcional para $H(\cdot)$:

$$H(e_{i,t}) = (b - 1)(1 + e_{i,t}^{-1})^{-\kappa} + 1 \quad (2.24)$$

A PTF mundial média é definida como:

$$\bar{T}_t = \left(\int T_{i,t}^\psi di \right)^{\frac{1}{\psi}} \quad (2.25)$$

Onde ψ é um parâmetro que dita o quanto a PTF mundial depende do país líder ou da fronteira. Por sua vez, a PTF mundial cresce por:

$$\bar{T}_{t+1} = T_t(1 + g_t) \quad (2.26)$$

Onde g_t é a taxa de crescimento e é exogenamente determinada.

Desse conjunto de equações é possível notar que os investimentos na PTF têm efeitos dinâmicos específicos para o país e outros internacionais. O parâmetro γ é o que determina a parcela dos efeitos dinâmicos que são específicos do país.

Do lado do consumidor, pressupõe-se que cada país tem uma dinastia que maximiza a função de utilidade

$$U(C_i) = \sum_{t=0}^{\infty} \beta^t \log(C_{i,t}) \quad (2.27)$$

Onde β é o fator de desconto e C é o consumo.

O problema do planejador do país é uma solução para o país como um todo, sem se preocupar com eventuais diferenças dentro do país (pressupõe-se que a tecnologia circula livremente dentro do país e todos os efeitos dinâmicos nesse caso são internalizados dentro do país). Assim, o planejador irá escolher a sequência de consumo e investimento (na PTF) que maximize a utilidade em (2.43):

$$\max_{\{T_{i,t}, C_{i,t}, e_{i,t}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \log(C_{i,t}) \quad (2.28)$$

s. a.

$$\sum_{t=0}^{\infty} p_t C_{i,t} = \sum_{t=0}^{\infty} p_t (T_{i,t} - w_t e_{i,t} + w_t e_w)$$

$$T_{i,t+1} = T_{i,t}^{1-\gamma} \bar{T}_t^\gamma H(e_{i,t})$$

Onde p_t é o preço no tempo 0 do bem no tempo 1 e $p_0 = 1$. Vale chamar atenção para o fato de que o planejador toma \bar{T}_t como dado, uma vez que se pressupõe que cada país é pequeno demais para influenciar significativamente a PTF mundial.

Como o modelo supõe que o mercado de capitais é perfeito, então as taxas de juros serão exógenas para cada país individualmente. Assim, a decisão de maximização do planejador pode ser dividida no problema da maximização de renda e problema de alocação intertemporal do consumo. Da alocação de consumo, temos que, no equilíbrio mundial, o crescimento do consumo do mundo e de cada país entre dois períodos será igual a taxa de juros vezes o fator de desconto:

$$\frac{C_{i,t+1}}{C_{i,t}} = \beta \frac{p_t}{p_{t+1}} \quad (2.29)$$

O modelo apresenta crescimento balanceado, onde $\bar{T}_t, T_{i,t}, C_{i,t}$ e w_t crescem a uma mesma taxa g . Com uma taxa comum, os autores então definem a PTF ajustada pelo salário dos trabalhadores altamente qualificados: $\hat{w} = w_0/\bar{T}_0 = w_t/\bar{T}_t$ para todo t . Similarmente, $C_{i,t+1}/C_{i,t} = 1 + g$. Isso permite reescrever o problema de maximização da renda num contexto de equilíbrio com crescimento balanceado. Assim, o planejador irá escolher a sequência de PTF relativa futura de maneira a maximizar o produto descontado os custos com investimento, e o problema de maximização da renda passa a poder ser escrito como:

$$\max_{\{z_{i,t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left(z_{i,t} - \hat{w} H^{-1} \left((z_{i,t+1}/z_{i,t}^{1-\gamma})(1+g) \right) \right) \quad (2.30)$$

Onde \hat{w} e g são dados e $z_{i,t} = T_{i,t}/\bar{T}_t$.

A solução do problema (2.46) resulta na seguinte equação de euler:

$$\frac{\hat{w}}{H'(e_{i,t})z_{i,t}^{1-\gamma}} = \frac{\beta}{1+g} + \frac{\hat{w}\beta(1-\gamma)z_{t+2}}{H'(e_{i,t+1})z_{i,t+1}^{2-\gamma}} \quad (2.31)$$

Dessa equação é possível perceber três efeitos de um aumento na PTF relativa na renda:

(i) há um aumento no custo de investimento no tempo t ; (ii) há um aumento na renda em $t+1$; e (iii) há um decréscimo nos custos de investimento em $t+1$, que depende de γ .

Uma vez definida as condições de equilíbrio para a trajetória de crescimento balanceado e assumindo uma distribuição $\Gamma(z)$ estacionária, é possível discutir os pormenores de como essa distribuição se dará. Os autores consideram duas possibilidades: um equilíbrio de crescimento balanceado simétrico (SBGE) e outro assimétrico (ABGE).

No primeiro, todos os países agem de maneira idêntica e $\Gamma(z)$ é degenerada em $z_i = 1$ para todo i , sendo, assim, trivialmente unimodal. Além disso, todos os países escolhem um nível de emprego altamente qualificado igual ao da média mundial. Neste caso a taxa de crescimento é determinada por $H(e_w)$. E a alocação será ótima para cada país de maneira que o salário, \hat{w} , também é função de e_w e é negativamente relacionado com γ . Uma coisa que chama atenção é a importância de γ para o equilíbrio: a função objetivo poderá ser côncava ou não dependendo da magnitude de γ . Assim, para diferentes valores de γ teremos diversas possibilidades.

No caso simétrico, há apenas um único ponto estacionário quando γ é alto, sendo este ponto o equilíbrio simétrico $z = 1$. Conforme γ vai diminuindo a função se torna cada vez mais instável e começa a se dobrar até chegar num ponto em que surgem

dois novos pontos estacionários: um acima e outro abaixo de $z = 1$. Para valores suficientemente pequenos de γ o equilíbrio simétrico deixa de existir, sobrando apenas a possibilidade de equilíbrios assimétricos.

Num equilíbrio assimétrico todas as variáveis crescem a taxas constantes e há mais do que um nível de PTF relativa para ser escolhido pelos planejadores. Nesse caso, os autores consideram a possibilidade de um equilíbrio assimétrico com dois grupos: isso se sustenta pela evidência de que a produtividade do trabalho teria uma distribuição bimodal no mundo (conforme gráfico em ACEMOGLU, 2009). Nesse equilíbrio, a taxa de crescimento do mundo é determinada pela divisão entre trabalhadores altamente qualificados nos grupos com PTF alta e baixa e do tamanho de cada grupo. Uma diferença com o equilíbrio simétrico é que nesse caso γ afeta a taxa de crescimento.

A determinação do valor de e para cada grupo é obtida combinando as equações de euler de cada grupo no estado estacionário: $H(e_{i,1})^{\frac{1-\gamma}{\gamma}} H'(e_{i,1}) = H(e_{i,2})^{\frac{1-\gamma}{\gamma}} H'(e_{i,2})$. Sob essa condição, para que o ABGE com dois grupos exista é necessário também que o termo $H(e_{i,j})^{\frac{1-\gamma}{\gamma}} H'(e_{i,j})$ não seja monotônico, e a não-monotonicidade dependerá do valor de γ e de $H(\cdot)$. A forma funcional para $H(\cdot)$ satisfaz as condições do ABGE com dois grupos.

A razão entre as PTFs dos dois grupos dependerá de suas respectivas quantidades de trabalhadores altamente qualificados, bem como do efeito de *catch up*, γ . No equilíbrio assimétrico o crescimento do consumo é igual para todos, mas os países do grupo com menor PTF também terão uma renda menor e um nível de consumo menor. Vale notar que, conforme simulações dos autores, para um determinado valor de γ suficientemente baixo é possível existir um SBGE instável e também um continuum de ABGE coexistindo. Abaixo desse valor apenas o ABGE passa a existir, e acima disso apenas o SBGE existe.

Essas diferenças ocorrem porque o modelo proposto tem duas forças distintas: há um efeito de *catch up*, cuja importância é medida por γ e há também a taxa de investimento em mais PTF, que gera divergência. Se o efeito de *catch up* for suficientemente importante, haverá um único equilíbrio para todos. Se esse efeito for baixo, haverá divergência, pois os países com alta PTF também investirão mais e o efeito do investimento se sobreporá ao efeito da convergência. Os autores também notam que a importância de cada país na constituição da PTF mundial pode impactar a

dinâmica fora do estado estacionário no ABGE, mas isso não é analisado a fundo no trabalho deles.

Uma coisa a se notar nos exercícios com as simulações dos autores é que há apenas estabilidade desses equilíbrios para valores muito altos de γ ou muito baixos (quando ABGE existe). Valores intermediários tendem a ter um SBGE instável, ainda que todos os valores em que ABGE existe são estáveis. Adicionar choques específicos para cada país na PTF não altera os principais resultados.

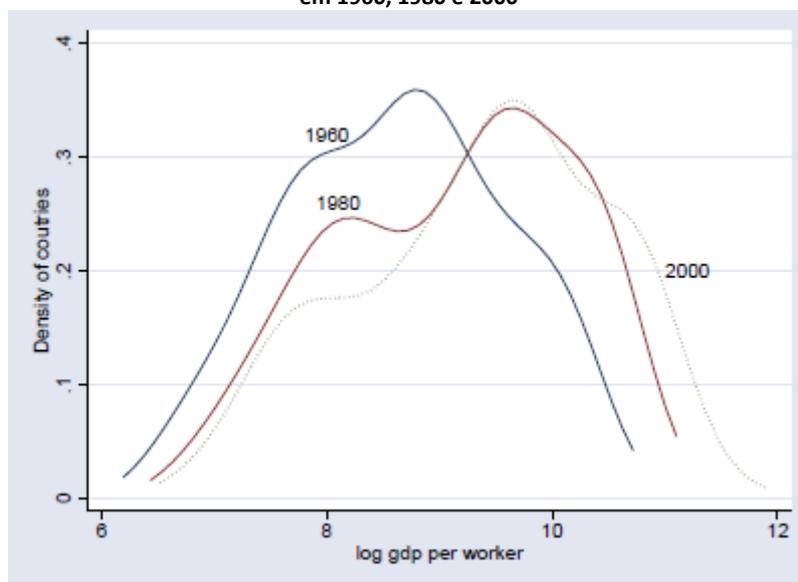
Os autores comparam esse modelo e seus resultados com a evolução da distribuição de renda por trabalhador conforme exposto em Acemoglu (2009), reproduzido abaixo. Como explicam Damsgaard e Krusell (2010), da década de 60 até os anos 2000 a distribuição de renda por trabalhador também foi se alterando, de maneira que há dois picos em 1980 e mais dispersão em 2000, enquanto a distribuição de 1960 não mostrava isso tão claramente. Apesar de isso ser não ser evidência suficiente para defender que há de fato uma distribuição bimodal, é fácil notar que a distribuição mudou ao longo do tempo, e nesse sentido os autores defendem que o modelo construído é capaz de identificar alguns fatores determinantes. Dentre esses fatores, o parâmetro γ parece ter bastante relevância principalmente na determinação da bimodalidade da distribuição de produtividades. Sobre esse parâmetro, os autores explicam que ele captura, na verdade, dois efeitos: por um lado ele mensura o quão fácil é a transferência de conhecimento na PTF do mundo para cada país. Por outro lado, ele também captura o quanto dos investimentos na PTF é internalizado em oposição ao quanto eles são tratados de maneira exógena.

Vale notar, porém, que γ não é o único parâmetro relevante na distribuição, os custos de troca são ignorados no modelo, mas seria de se esperar que isso também afete a distribuição. Como eles argumentam, uma queda nos custos de troca afetaria também a mobilidade dos trabalhadores altamente qualificados, e isso teria potencial para diminuir a convergência e ir em direção a uma distribuição bimodal, bem como mudanças na forma funcional do investimento em PTF, $H(.)$ tem potencial para alterar a distribuição.

Na Figura 1, de Acemoglu (2009), aparentemente a distribuição do produto por trabalhador no mundo teria se tornado mais bimodal. Considerando os efeitos analisados por Damsgaard e Krusell (2010) dão algumas possíveis explicações para esse fenômeno: (i) pode ter ocorrido uma queda nos transbordamentos tecnológicos (queda

em γ); (ii) diminuição nos custos de troca de e ; (iii) maior sensibilidade da TFP em relação à e ; ou (iv) dinâmica de transição.

Figura 1 - Estimativa da distribuição de países de acordo com o PIB por trabalhador (em log e ajustado pela PPP) em 1960, 1980 e 2000



Fonte: Acemoglu (2009).

2.5. Eficiência Alocativa e PTF

Como nos argumentos de Parente e Prescott (2000 *apud* GONÇALVES, 2013) e no modelo de Acemoglu, Antras e Helpman (2007 *apud* ACEMOGLU, 2009), diferenças organizacionais são causas potenciais para explicar as razões pelas quais alguns países são mais ou menos capazes de absorver tecnologia do resto do mundo e também o porquê de alguns países decidirem investir mais ou menos no desenvolvimento de novas tecnologias. No entanto, deve-se ressaltar que a PTF, enquanto resíduo, engloba tanto questões relacionadas ao uso de novas tecnologias quanto à alocação eficiente dos recursos disponíveis naquela economia. Nessa dimensão de *eficiência alocativa*, Gonçalves (2013) discute brevemente a forma como instituições e a política econômica em um dado país podem ter impactos diretos sobre a PTF.

No que se refere às instituições, Hall e Jones (1999) e Acemoglu et. al. (2001, 2002) apresentam evidências de forte relação entre instituições e a PTF. O primeiro artigo faz uso de variáveis geográficas e linguísticas como instrumento para um indicador de qualidade institucional. O segundo artigo, por sua vez, se aproveita da ideia de que as instituições evoluem lentamente. Isso permite aos autores associar a qualidade institucional com o tipo de colonização feita naquele país, dividindo entre colônias de

povoamento e colônias de exploração⁷. O instrumento para permitir a identificação dessa variável foi à taxa de mortalidade dos colonizadores nas suas respectivas colônias, partindo do pressuposto que países mais inóspitos estariam mais sujeitos a se tornarem colônias de exploração. A partir dessas instituições iniciais, os autores supõem que tais diferenças se perpetuariam no tempo, devido a grande lentidão com que ocorrem as mudanças institucionais. Essa perspectiva é vista também nos trabalhos de North (1990), que foi um dos primeiros a tentar explicar a relação entre instituições formais e informais sobre o crescimento econômico.

Apesar da relativa rigidez das instituições, Gonçalves (2013) faz questão de ressaltar que “o entorno institucional de um país não é tudo e não determina unicamente o nível de desenvolvimento possível de ser atingido” (pg.55). Como ele argumenta, as instituições servem como uma restrição às escolhas possíveis e sua eficácia, mas ainda há espaço para decisão nesse ambiente. Nesse sentido, a política econômica teria um papel importante em definir os rumos de um dado país. Para exemplificar isso, é citada a diferença entre Barbados e a Jamaica. O primeiro país teria adotado políticas macroeconômicas mais adequadas (marcadas por maior prudência fiscal e combate inflacionário) e respondido de maneira mais adequada aos choques do petróleo na década de 70. Enquanto isso a Jamaica optou por populismo macroeconômico, levando a estagnação. Em termos de política econômica, o autor termina citando algumas possíveis políticas favoráveis ao crescimento econômico, tal como abertura comercial, menor inflação e maior desenvolvimento do mercado de crédito. A desigualdade e dívida pública também parecem ter um papel importante no crescimento (uma relação negativa), porém a relação dessas variáveis com o crescimento não ainda está bem estabelecida na literatura.

⁷ Basicamente, a ideia subjacente é que nas colônias de povoamento, foram implantadas instituições favoráveis ao crescimento, marcadas pelo império da lei, divisão igualitária da terra e fortalecimento dos direitos de propriedade para grande parte da população. Já as colônias de exploração não contavam com instituições similares, pois os colonizadores não tinham o interesse em viver naquele território.

3. METODOLOGIA

3.1. O Modelo

Nossa metodologia para o cálculo da produtividade total dos fatores (PTF) parte da função de produção agregada, $Y_{i,t}$, do país i no tempo t :

$$Y_{i,t} = K_{it}^\alpha (A_{i,t} H_{i,t})^{1-\alpha} \quad (3.1)$$

Onde K é o estoque capital físico, H é o estoque capital humano, L é a força de trabalho, A é a produtividade total dos fatores, α é a participação do capital no produto e $(1 - \alpha)$ é a participação do trabalho no produto; $i = 1, \dots, N$ e $t = 0, \dots, T$.

A PTF é calculada como o resíduo de uma função de produção agregada do tipo Cobb-Douglas:

$$A_{i,t} = \left(\frac{Y_{i,t}}{K_{i,t}^\alpha H_{i,t}^{1-\alpha}} \right)^{\frac{1}{1-\alpha}} \quad (3.2)$$

Note que a tecnologia neste caso é *Horrod-neutra*⁸. Nesta especificação α é invariante no tempo e entre países. Para este trabalho, supomos que $\alpha = 0.4$.

O capital humano, H , é definido como:

$$H_{it} = e^{\phi(s_{it})} L_{it} \quad (3.3)$$

Onde s são os anos médios de escolaridade dos residentes com mais de 15 anos do país e $\phi(\cdot)$ é uma função de retorno da educação sobre o salário, cuja derivada, $\partial\phi/\partial s_{it}$ é tomado como sendo o retorno da escolaridade da regressão de salário minceriana. Nesse sentido, quanto maior a escolaridade, maior a produtividade daquele trabalhador.

Como visto anteriormente, os modelos de Nelson e Phelps (1966), Acemoglu (2009), Damsgaard e Krusell (2010) e Aiyar e Feyrer (2002) trabalham todos com algum tipo de *gap* tecnológico, dado pela diferença entre a PTF mundial (que pode ser uma fronteira tecnológica ou média mundial) e a tecnologia atualmente disponível no país. Desse modo, suporemos ao longo do trabalho que a acumulação de PTF se dará de acordo com a formulação de Damsgaard e Krusell (2010) modificada para incluir J investimentos em PTF:

$$\frac{A_{i,t+1}}{A_{i,t}} = \left(\frac{T_t}{A_{i,t}} \right)^\gamma \prod_{j=1}^J X_{j,i,t}^{\beta_j} F_i \quad (3.4)$$

⁸ Escolhemos essa especificação por ela ser consistente com a trajetória de *crescimento balanceado*. Isto é, a trajetória de crescimento consistente com os *Fatos de Kaldor*. Tal relação pode ser demonstrada pelo Teorema de Uzawa e seu respectivo corolário (ver ACEMOGLU, 2009).

Tal que T_t é a PTF mundial, F_i é um efeito fixo e $X_{j,i,t}$ é o j -ésimo investimento em PTF (de J investimentos). Nesta formulação o parâmetro γ captura a importância do *catch up* tecnológico. Os coeficientes β_j representam as respectivas constantes de cada investimento (determinante) da PTF. Pela hipótese de *catch up* tecnológico, esperaríamos que $\gamma > 0$.

Tomando o logaritmo de (3.4):

$$\begin{aligned}\Delta \ln A_{i,t+1} &= \gamma [\ln T_t - \ln A_{i,t}] + \sum_{j=1}^J \ln X_{i,t}^{\beta_j} F_i \\ \Delta \ln A_{i,t+1} &= \gamma \left[\ln \left(T_t F_i^{1/\gamma} \prod_{j=1}^J X_{j,i,t}^{\beta_j/\gamma} \right) - \ln A_{i,t} \right]\end{aligned}\quad (3.5)$$

Em tempo contínuo, ficaríamos com:

$$\frac{d \ln A_i(t)}{dt} = \gamma \left[\ln \left(T(t) F_i^{1/\gamma} \prod_{j=1}^J X_{j,i}(t)^{\beta_j/\gamma} \right) - \ln A_i(t) \right]\quad (3.6)$$

Observe que essa expressão é bastante similar à formulação da dinâmica da PTF em Aiyar e Feyrer (2002). Seguindo a terminologia dos autores, definimos:

$$A_i^*(t) = T(t) \tilde{F}_i \prod_{j=1}^J X_{j,i}(t)^{\phi_j}$$

Tal que $A_i^*(t)$ é a PTF potencial do país i e $\phi_j = \beta_j/\gamma$ é a elasticidade de cada um dos potenciais determinantes da PTF, e $\tilde{F}_i = F_i^{1/\gamma}$ é um efeito fixo. Sob essa formulação, caímos na mesma formulação que os autores supracitados. Nesta formulação, γ pode ser entendido como um coeficiente de convergência condicional.

Podemos desenvolver a expressão (3.6) da mesma forma que os autores fizeram. Primeiro multiplicando ambos os lados por $e^{\gamma t}$ e então tomando a integral entre t_1 e t_2 :

$$\begin{aligned}e^{\gamma t} \left[\frac{d \ln A_i(t)}{dt} + \gamma \ln A_i(t) \right] &= \gamma e^{\gamma t} \left[\ln \left(T(t) \prod_{j=1}^J X_{j,i}(t)^{\phi_j} \tilde{F}_i \right) \right] \\ \int_{t_1}^{t_2} e^{\gamma t} \left[\frac{d \ln A_i(t)}{dt} + \gamma \ln A_i(t) \right] dt &= \int_{t_1}^{t_2} \gamma e^{\gamma t} \left[\ln \left(T(t) \prod_{j=1}^J X_{j,i}(t)^{\phi_j} \tilde{F}_i \right) \right] dt\end{aligned}$$

Desenvolvendo o lado esquerdo da equação:

$$\begin{aligned}\int_{t_1}^{t_2} e^{\gamma t} \frac{d \ln A_i(t)}{dt} dt + \int_{t_1}^{t_2} e^{\gamma t} \gamma \ln A_i(t) dt \\ \int_{t_1}^{t_2} e^{\gamma t} \frac{d \ln A_i(t)}{dt} dt + \left([e^{\gamma t} \ln A_i(t)]_{t_1}^{t_2} - \int_{t_1}^{t_2} e^{\gamma t} \frac{d \ln A_i(t)}{dt} dt \right) \\ [e^{\gamma t} \ln A_i(t)]_{t_1}^{t_2} = e^{\gamma t_2} \ln A_i(t_2) - e^{\gamma t_1} \ln A_i(t_1)\end{aligned}$$

Portanto,

$$e^{\gamma t_2} \ln A_i(t_2) = e^{\gamma t_1} \ln A_i(t_1) + \int_{t_1}^{t_2} \gamma e^{\gamma t} \left[\ln \left(T(t) \prod_{j=1}^J X_{j,i}(t)^{\phi_j} \tilde{F}_i \right) \right] dt$$

Multiplicando ambos os lados por $e^{-\lambda t_2}$ resulta em:

$$\ln A_i(t_2) = e^{-\gamma\tau} \ln A_i(t_1) + e^{-\lambda t_2} \int_{t_1}^{t_2} \gamma e^{\gamma t} \left[\ln \left(T(t) \prod_{j=1}^J X_{j,i}(t)^{\phi_j} \tilde{F}_i \right) \right] dt$$

Tal que $\tau = (t_2 - t_1)$. Agora assumamos que a PTF mundial e seus determinantes são constantes dentro do intervalo de integração t_1 e t_2 , mas dão saltos entre os períodos (isto é, são variáveis discretas). Este pressuposto não é forte, visto que, para eliminar efeitos de ciclos de negócios, é usual trabalhar com as médias de 5 ou mais anos das variáveis em modelos de crescimento. Como T e X não variam dentro do intervalo, podemos tirá-los da integral:

$$\ln A_i(t_2) = e^{-\gamma\tau} \ln A_i(t_1) + e^{-\lambda t_2} \ln \left(T \prod_{j=1}^J X_{j,i}^{\phi_j} \tilde{F}_i \right) \int_{t_1}^{t_2} \gamma e^{\gamma t} dt$$

$$\ln A_i(t_2) = e^{-\gamma\tau} \ln A_i(t_1) + e^{-\lambda t_2} \ln \left(T \prod_{j=1}^J X_{j,i}^{\phi_j} \tilde{F}_i \right) [e^{\lambda t_2} - e^{\lambda t_1}]$$

$$\ln A_i(t_2) = e^{-\gamma\tau} \ln A_i(t_1) + (1 - e^{-\lambda\tau}) \ln T + \phi(1 - e^{-\lambda\tau}) \ln \prod_{j=1}^J X_{j,i} + (1 - e^{-\lambda\tau}) \ln \tilde{F}_i \quad (3.7)$$

Defina $a_{i,t} = \ln A_i(t_2)$, $a_{i,t-1} = \ln A_i(t_1)$, $g_{t-1} = \ln T$, $f_i = (1 - e^{-\lambda\tau}) \ln \tilde{F}_i$, $x_{j,i,t-1} = \ln X_{j,i}$, $\omega = e^{-\lambda\tau}$, $\theta = (1 - e^{-\lambda\tau})$ e $\beta_j = \phi_j(1 - e^{-\lambda\tau})$. Com isso, e adicionando um termo de erro na equação, ficamos com:

$$a_{i,t} = \omega a_{i,t-1} + \theta g_{t-1} + \sum_{j=1}^J \beta_j x_{j,i,t-1} + f_i + \varepsilon_{i,t} \quad (3.8)$$

Aiyar e Feyrer (2002), ao construir sua PTF potencial e chegar à sua versão da equação (3.8), consideram apenas um investimento: estoque de capital humano. Aqui buscamos considerar, além do capital humano, outros determinantes frequentemente usados na literatura, tal como instituições, inflação e abertura comercial e gasto do governo (ver LOKO e DIOUF, 2009).

Se for válido que

$$E[f_i] = E[\varepsilon_{i,t}] = 0 \quad (3.9)$$

$$E[\varepsilon_{i,t} f_i] = 0, \text{ para } i = 1, \dots, N \text{ e } t = 2, \dots, T \quad (3.10)$$

$$E[\varepsilon_{i,t} \varepsilon_{i,s}] = 0, \text{ para } i = 1, \dots, N \text{ e } s \neq t \quad (3.11)$$

$$E[a_{i,t} \varepsilon_{i,t}] = 0, \text{ para } i = 1, \dots, N \text{ e } t = 2, \dots, T \quad (3.12)$$

Neste caso, é possível o modelo via *first difference* GMM. Adicionando a hipótese de que

$$E[f_i \Delta a_{i,2}] = 0 \quad (3.13)$$

Para $i = 1, \dots, N$, então podemos estimar os parâmetros do modelo por meio de um GMM Sistemático (*system* GMM ou SGMM).

Note que as equações (3.9-13) implicam a existência de uma tendência comum à todos os países. Esta tendência pode ser entendida como uma PTF mundial. Uma forma de tratar ela num painel dinâmico seria removê-la da equação por meio alguma

transformação nas variáveis. Outra forma é incluí-la como uma variável no modelo. Para isso, é necessário especificar como essa tendência comum se comporta. Para especificar esse fator comum, trabalharemos com a ideia de que a PTF é dividida em dois componentes e que cada um desses componentes segue um processo autorregressivo de ordem 1⁹:

$$A_{i,t} = \bar{A}_i e^{\lambda g_t + f_{i,t}} \quad (3.14)$$

$$g_t = \rho g_{t-1} + v_t \quad (3.15)$$

$$f_{i,t} = \phi_i f_{i,t-1} + u_{i,t} \quad (3.16)$$

Tal que \bar{A}_i é a média da produtividade do país i e g_t é um fator comum a todos os países, $f_{i,t}$ é o fator específico e v_t e $u_{i,t}$ são os termos de erro. Cada um dos termos de erro segue uma distribuição normal com média zero e variância constante,

$$v_t \sim N(0, \sigma_v^2) \quad (3.17)$$

$$u_{i,t} \sim N(0, \sigma_{u_i}^2) \quad (3.18)$$

Onde σ_v^2 é a variância de v_t e $\sigma_{u_i}^2$ é a variância de $u_{i,t}$. Pressupõe-se que v_t , $u_{i,t}$ e $u_{j,t}$, $j \neq i$, são não-correlacionados entre si.

Passando a média para o lado esquerdo da equação (3.4) e tomando seu logaritmo natural, obtemos:

$$\ln A_{i,t}^* = \ln A_{i,t} - \ln \bar{A}_i = \lambda_i g_t + f_{i,t} \quad (3.19)$$

É esta equação (3.6) que será de nosso interesse estimar para obter o fator comum, e seus parâmetros podem ser estimados através de um modelo de fatores dinâmicos (*dynamic factors model* – DFM). O DFM é um caso específico da classe mais geral de modelos conhecidos como modelos de estado-espço. A estimação desse tipo de modelo é feita utilizando as recursões do filtro de Kalman, que permitem obter estimativas consistentes mesmo para séries não-estacionárias.

⁹ Uma formulação usual para modelar a PTF em modelos de ciclos reais de negócios é assumir que a mesma se relaciona com um fator autorregressivo de ordem 1. Romer (2011), ao definir um modelo base para o modelo de ciclos, coloca a PTF como sendo $A_t = \bar{A} e^{g_t + \tilde{A}_t} \Rightarrow \ln A_t = \bar{A} + g_t + \tilde{A}_t$, onde \tilde{A}_t reflete o distanciamento da tendência, sendo que esta segue um processo AR(1). Aqui não estamos trabalhando com ciclos econômicos, logo não há necessidade de separar a tendência. O que realmente muda em nossa formulação é considerar dois componentes autorregressivos ao invés de um.

3.2. Procedimento de Estimação¹⁰

3.2.1. Modelo de Fatores Dinâmicos

A estimação do fator comum se dá dentro do arcabouço de modelos de fatores dinâmicos (DFM). Aqui seguiremos uma metodologia similar a de Morales-Arias e Moura (2013), porém considerando um apenas o caso homocedástico. Seguindo essa metodologia, começamos por representar (12) numa forma de estado espaço:

$$a_t = ZF_t \quad (3.20)$$

$$F_t = TF_{t-1} + \eta_t \quad (3.21)$$

Onde $a_t = (\ln A_{1,t}^*, \dots, \ln A_{N,t}^*)'$ é um vetor $N \times 1$ com a produtividades de cada país centrada em torno da média; $F_t = (g_t, f_{1,t}, \dots, f_{N,t})'$ é um vetor $(N+1) \times 1$ de estado do modelo, contendo o componente comum e os componentes específicos; e $\eta_t = (v_t, u_{1,t}, \dots, u_{N,t})'$ é um vetor $N \times 1$ de erros do estado. Por sua vez, $Z = [\Lambda; I_N]$ é uma matriz $N \times (N+1)$ de parâmetros, onde $\Lambda = (\lambda_1, \dots, \lambda_N)'$ é um vetor $N \times 1$ com os *loadings*; e $T = \text{diag}(\rho, \phi_1, \dots, \phi_N)$ é uma matriz diagonal $(N+1) \times (N+1)$. Para uma visualização, seguem abaixo as equações (14) e (15) abertas em matrizes:

$$\begin{bmatrix} \ln A_{1,t}^* \\ \vdots \\ \ln A_{N,t}^* \end{bmatrix} = \begin{bmatrix} \lambda_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_N & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} g_t \\ f_{1,t} \\ \vdots \\ f_{N,t} \end{bmatrix} \quad (3.20a)$$

$$\begin{bmatrix} g_t \\ f_{1,t} \\ \vdots \\ f_{N,t} \end{bmatrix} = \begin{bmatrix} \rho & 0 & \cdots & 0 \\ 0 & \phi_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_N \end{bmatrix} \begin{bmatrix} g_{t-1} \\ f_{1,t-1} \\ \vdots \\ f_{N,t-1} \end{bmatrix} + \begin{bmatrix} v_t \\ u_{1,t} \\ \vdots \\ u_{N,t} \end{bmatrix} \quad (3.21a)$$

Uma vez postas no formato de estado espaço, é possível aplicar as recursões do filtro de Kalman e obter uma função de verossimilhança aproximada através da decomposição de previsão dos erros para $t = 1, \dots, T$:

$$\ln \mathcal{F}(a_t; \theta) = -0.5 \left(NT \ln 2\pi + \sum_{t=1}^T \ln |\Sigma_{a_t}| + \sum_{t=1}^T \varepsilon_t' \Sigma_{a_t}^{-1} \varepsilon_t \right) \quad (3.22)$$

Onde $\varepsilon_t = a_t - \tilde{a}_{t|t-1}$, $\tilde{a}_{t|t-1}$ é o estimador ótimo de a_t dado o conjunto de informações até $t - 1$ ¹¹, $\mathcal{F}(\cdot)$ é a função de máxima verossimilhança, $\Sigma_{a_t} = \text{Var}(a_t)$, N é o número de países, T é o número de períodos e θ é vetor de parâmetros a serem estimados. Detalhes sobre os modelos de estado espaço e as recursões do filtro de Kalman estão disponíveis no apêndice (apêndice A) deste trabalho.

¹⁰ Uma Revisão sobre os principais métodos utilizados neste trabalho estão apresentados em maiores detalhes no Apêndice A.

¹¹ Como notação geral, $x_{t|t} = E(x_t | \mathcal{J}_t)$, onde \mathcal{J} é o conjunto de informações até t , sendo t um período maior ou menor que t .

Para facilitar a interpretação do modelo, foi feita uma normalização nos coeficientes de *loadings* em relação aos EUA, de tal modo que $\lambda_i^* = \lambda_i / \lambda_{EUA}$ para $i = 2, \dots, N$.

Ainda seguindo Morales-Arias e Moura (2013), bem como Stock e Watson (2010), partimos da Análise de Componente Principais (ACP) sobre as respectivas variáveis dependentes para encontrar os valores iniciais dos coeficientes de *loading* e g_t (tomado como o primeiro componente principal). Na posse desses dados, é possível construir $f_{i,t}$ ¹² e então estimar regressões de $f_{i,t}$ sobre $f_{i,t-1}$ e de g_t sobre g_{t-1} para obter, respectivamente, ρ e ϕ_i , bem como as variâncias de g_t e $f_{i,t}$. A metodologia empregada para o modelo em primeira diferença é análoga.

3.2.2. Decomposição da Variância¹³

Neste trabalho realizamos uma decomposição da variância como forma de mensurar a contribuição relativa do fator comum e dos fatores específicos de cada país. Isso nos permite avaliar a importância do fator comum sobre cada PTF. Sendo assim, a variância de $\ln A_{i,t}^*$, a partir de (3.12), é dada por:

$$\text{Var}(Y_{i,t}) = \sum_{i=0}^2 \text{Var}(X_i) + 2 \sum_{i \neq j}^2 \text{Cov}(X_i, X_j) \quad (3.23)$$

Onde $X_1 = \hat{\lambda}_i \hat{g}_t$, $X_2 = \hat{f}_{i,t}$ e $Y_{i,t} = \ln A_{i,t}^*$. Dividindo (3.17) por $\text{Var}(Y_{i,t})$ resulta em:

$$\sum_{i=0}^2 \frac{\text{Var}(X_i)}{\text{Var}(Y_{i,t})} + 2 \sum_{i \neq j}^2 \frac{\text{Cov}(X_i, X_j)}{\text{Var}(Y_{i,t})} = 1 = R^2(Y_{i,t}; X_1, X_2) \quad (3.24)$$

Onde $R^2(Y_{i,t}; X_1, X_2)$ é simplesmente o R^2 de uma regressão das PTFs contra o fator comum e idiossincrático. Neste caso em específico, os dois elementos explicam toda a variância das PTFs por construção.

Assumindo ortogonalidade entre os fatores, uma forma possível de representar a parcela da variância explicada pelo fator comum é através da Parcela da Variância Bruta (PVB):

$$\frac{\hat{\beta}_i^2 \text{Var}(\hat{g}_t)}{\text{Var}(\ln A_{i,t}^*)} = R^2(\ln A_{i,t}^*; \hat{g}_t) \quad (3.25)$$

A PVB poderia ser entendida como o R^2 de uma regressão de cada PTF sobre o fator comum sem controlar pelo fator específico¹⁴. Como explica Gibbons et. al. (2012), essa

¹² Do fato de que $f_{i,t} = \ln A_{i,t}^* - \lambda_i g_t$.

¹³ Uma rápida revisão sobre as formas de decomposição pode ser encontrada no apêndice deste trabalho. Também ver Gibbons et. al. (2012).

¹⁴ Por conta disso, $\hat{\beta}_i$ não é necessariamente igual ao $\hat{\lambda}_i$ estimado pelo modelo de fatores dinâmicos.

estimativa mostra o quanto a variável explicativa em questão explica na *média* as flutuações na variável de interesse. Exemplos desse procedimento sendo aplicado para decompor a variância de modelos de fatores dinâmicos podem ser vistos em Morales-Arias e Moura (2010) e Kose et. al. (2008). Para uma discussão sobre esta e outras formas de decomposição da variância, ver Gibbons et. al. (2012).

3.2.3. Método dos Momentos Generalizado Sistemático¹⁵

Havíamos derivado a seguinte equação

$$a_{i,t} = \omega a_{i,t-1} + \theta g_{t-1} + \sum_{j=1}^J \beta_j x_{j,i,t-1} + e_{i,t} \quad (3.26)$$

$$e_{i,t} = f_i + \varepsilon_{i,t} \quad (3.27)$$

Com as seguintes hipóteses:

$$E[f_i] = E[\varepsilon_{i,t}] = 0 \quad (3.28)$$

$$E[\varepsilon_{i,t} f_i] = 0, \text{ para } i = 1, \dots, N \text{ e } t = 2, \dots, T \quad (3.29)$$

$$E[\varepsilon_{i,t} \varepsilon_{i,s}] = 0, \text{ para } i = 1, \dots, N \text{ e } s \neq t \quad (3.30)$$

$$E[a_{i,t} \varepsilon_{i,t}] = 0, \text{ para } i = 1, \dots, N \text{ e } t = 2, \dots, T \quad (3.31)$$

$$E[Z_{i,t} \varepsilon_{i,t}] = 0, \text{ para } i = 1, \dots, N \text{ e } t = 1, \dots, T \quad (3.32)$$

$$E[f_i \Delta a_{i,t}] = 0 \quad (3.33)$$

$$E[f_i \Delta Z_{i,t}] = 0 \quad (3.34)$$

Tal que Z é o vetor com todos os X_j e g_t . Este modelo pode ser posto dentro de um painel com $i = 1, \dots, N$ indivíduos e $t = 1, \dots, T$ períodos de tempo¹⁶. Um problema, porém, é que f_i não é observado, de modo que uma estimação via Mínimos Quadrados Ordinários levaria a estimativas inconsistentes dos parâmetros. Quando se suspeita que um ou mais variáveis explicativas esteja relacionadas com o efeito fixo, a forma usual de se trabalhar isso é transformar os dados de modo a eliminar o efeito fixo, entretanto a transformação do estimador de efeitos fixos não consegue eliminar a fonte de endogeneidade da variável dependente defasada, sendo necessárias outras formas de lidar com o problema. É nesse contexto que surge o GMM em primeira diferença (*first-difference* GMM ou FD-GMM), que busca utilizar a transformação de primeira diferença para remover o efeito fixo e então instrumentalizar as variáveis em primeiras diferenças do lado direito da equação através de defasagens das séries.

¹⁵ Para uma apresentação mais detalhada, ver o Apêndice A.

¹⁶ Sobre a razoabilidade desta hipótese em modelos de crescimento, a princípio ela pode parecer estranha, visto que esse pressuposto implicaria estacionariedade das séries utilizadas, como se o efeito fixo se compensasse com o fator autorregressivo em cada período no tempo. Apesar de parecer irrealista para nosso modelo, é possível satisfazer essa hipótese incluindo *dummies* de tempo na equação (ROODMAN, 2009; BOND ET. AL., 2001).

Um problema com o GMM em primeira diferença é que se o processo gerador de dados for próximo de um passeio aleatório, então as variáveis em primeira diferença estarão fracamente correlacionadas com as defasagens em nível, logo os instrumentos serão fracos.

O GMM Sistemico, de Arellano e Bover (1995), soluciona este problema. Sob as hipóteses (3.28-34), podemos construir os seguintes momentos:

$$E[a_{i,t-s}\Delta\varepsilon_{i,t}] = 0$$

Para $t = 3, \dots, T$ e $s \geq 2$. E

$$E[(f_i + \varepsilon_{i,t})\Delta a_{i,t-1}] = 0$$

Para $i = 1, \dots, N$ e $t = 3, \dots, T$. Essas novas restrições permitem a utilização de defasagens em primeira diferença e a série como instrumentos na equação em nível, como sugerido em Arellano e Bover (1995), isto é, formando um sistema de equações com ambas as equações em primeira diferença e em nível, onde as equações em primeira diferença são instrumentalizadas com as defasagens em nível (igual ao GMM em primeira diferença), e o segundo conjunto de equações é instrumentalizado com as defasagens da primeira diferença. Sendo assim, o GMM Sistemico terá $T - 2$ equações em primeira diferença e mais $T - 2$ equações em nível, correspondendo aos períodos $t \geq 3$ disponíveis para os quais os instrumentos são observados.

3.3. Base de Dados

Nossa base de dados considera dados anuais da Penn World Table 9.0 (PWT)¹⁷. A base traz informações anuais sobre os níveis relativos de renda, produto, insumos e produtividade para 182 países entre 1950 e 2014. As variáveis Y_{it} , K_{it} e H_{it} podem ser obtidas diretamente da base. Quanto a PTF, são disponibilizados dois indicadores, mas eles não são adequados para um tratamento em painel. Sendo assim, construiremos a PTF como em (3.2).

Da PWT, selecionamos os países com algum dado de PTF disponível entre 1970 e 2014. No período, ao menos 144 países tinham dados para PTF em algum ano. Para o modelo de fatores dinâmicos, foram selecionados 10 países escolhidos de acordo com o *ranking* de eficiência técnica criado por Pires e Garcia (2011). Os 10 países são Israel, Suíça, Canadá, Japão, Noruega, Suécia, Estados Unidos, Dinamarca, Grã Bretanha e Hong Kong.

¹⁷ Versão pós-correções de 19 de agosto de 2016.

Sobre as variáveis, a PWT 9.0 disponibiliza três formas de calcular o PIB Real: (i) pela ótica do dispêndio ($RGDP^e$); (ii) pela ótica da produção ($RGDP^o$); e (iii) pelas contas nacionais ($RGDP^{NA}$). As duas primeiras métricas diferem entre si na forma como são deflacionadas. A variável $RGDP^e$ é deflacionada utilizando a paridade do poder de compra (PPP) para os bens finais para todas as quantidades produzidas, enquanto a variável $RGDP^o$ usa um PPP próprio para as exportações e importações. Com isso, as duas variáveis não serão iguais, e sua diferença é devida aos termos de troca, grau de abertura da economia, saldo da balança comercial e preço relativo dos bens transacionáveis. Já a variável $RGDP^{NA}$ difere das outras duas pela forma como é calculado sua taxa de crescimento para realizar extrapolações e interpolações a partir dos anos base (*benchmark*) dos preços de referência¹⁸.

Para calcular o $RGDP^e$ e $RGDP^o$ são considerados múltiplos *benchmarks* dos dados de preços e preços finais da ICP (*International Comparision Program*) para categorias de consumo, investimento e gastos do governo para cada país relativamente aos Estados Unidos¹⁹. A partir desses anos são então feitas interpolações e extrapolações a partir das tendências de preços das contas nacionais. Para que essas variáveis sejam comparáveis entre países e no tempo, é computada a seguinte taxa de crescimento para o cálculo do produto nos anos fora do ano base:

$$\frac{RGDP_{j,t}}{RGDP_{j,t-1}} = \left[\left(\frac{\Pi'_{t-1} y_{j,t}}{\Pi'_{t-1} y_{j,t-1}} \right) \left(\frac{\Pi'_t y_{j,t}}{\Pi'_t y_{j,t-1}} \right) \right]^{0.5} \quad (3.35)$$

Onde Π é o vetor de preços de referência e y é o vetor de quantidades.

Por sua vez, a variável $RGDP^{NA}$ utiliza o crescimento das contas nacionais como base (FEENSTRA *et al.*, 2013). Vale notar que, por conta dessas diferenças no cálculo das taxas de crescimento de cada série, a priori não há qualquer razão para acreditar que tais séries serão próximas, e de fato é possível encontrar exemplos em que há fortes discrepâncias entre as séries²⁰. Essas diferenças se devem à forma como o produto é deflacionado. Enquanto $RGDP^e$ e $RGDP^o$ consideram a PPP computada a partir do *benchmark* da ICP, $RGDP^{NA}$ considera as taxas de inflação relativas das contas nacionais²¹. Essa diferença surge então porque a inflação e a PPP medem coisas

¹⁸ Todas as três medidas de PIB Real permitem comparações entre países e ao longo do tempo.

¹⁹ O processo de agregação para os bens finais é, então, possível para todos os países e anos em que há dados coletados pela ICP (1970, 1975, 1980, 1985, 1996, 2005 e 2011). Para bens comercializados, os preços usados são versões ajustadas obtidas de Feenstra e Romalis (2012) (FEENSTRA *et al.*, 2013).

²⁰ Um exemplo é o Brasil na década de 90, em que se observa uma grande discrepância entre $RGDP^e$ e $RGDP^{NA}$.

²¹ Relativo à inflação dos Estados Unidos.

distintas. A inflação considera apenas a mudança nos preços domésticos, levando em conta apenas as parcelas do orçamento doméstico em conta. Em contraste, a PPP busca comparar os preços entre países, e naturalmente levam em conta as parcelas do orçamento de vários países. Como tal, discrepâncias surgem quando há discrepâncias significativas entre os orçamentos que são levados em consideração por cada indicador (FEENSTRA *et al.*, 2013).

Devido a essas diferenças metodológicas, temos que cada uma dessas variáveis é adequada para tratar de acordo com a questão que está sendo tratada. A variável $RGDP^e$ é considerada uma medida de mudanças nos padrões de vida, enquanto que a variável $RGDP^o$ é uma medida de mudanças nas capacidades produtivas. A terceira medida, $RGDP^{NA}$, por sua vez, é uma medida mais apropriada para avaliar o crescimento econômico de um país.

Portanto, neste trabalho a variável de produto mais adequada é a variável $RGDP^{NA}$, que é mais indicada para tratar questões relativas de crescimento econômico.

Para o estoque de capital físico, a PWT 9.0 disponibiliza duas medidas de capital físico, ck e rk^{na} , que diferem entre si no fato de que a primeira está calculada a PPP corrente (em milhões de US\$ de 2011) e a segunda está a preços nacionais constantes de 2011 (em milhões de US\$ de 2011). Neste trabalho utilizaremos rk^{na} para representar o estoque de capital físico. A forma de cálculo do estoque de capital físico difere do usualmente assumido nos modelos de crescimento, que aplica o método do inventário perpétuo (PIM) para o estoque de capital físico agregado e depreciação constante²². Na PWT 9.0 há a diferença de que são consideradas taxas de depreciação diferentes para os diferentes ativos que compõem o estoque de capital. Além disso, apenas as taxas de depreciação de cada ativo individual é que é constante, a taxa de depreciação do estoque de capital como um todo pode variar tanto no tempo quanto entre países, fornecendo maior realismo ao indicador. Nesse caso, são considerados 6 tipos de ativos (estruturas, equipamentos de transporte, computadores, equipamentos de comunicação, softwares e outras máquinas e ativos). Cada ativo segue a seguinte lei de movimento:

$$K_{a,i,t} = (1 - \delta_a)K_{a,i,t-1} + I_{a,i,t} \quad (3.36)$$

Tal que o a denota o ativo, i o país e t o tempo. Assim, cada ativo deprecia a uma taxa fixa, enquanto a depreciação total é computada como $\delta_{i,t} = \sum_{a=1}^6 P_{a,i,t} \delta_{a,i,t} K_{a,i,t-1}$,

²² Geralmente é comum partir do pressuposto de que o estoque de capital físico é descrito pela seguinte equação de movimento: $K_t = I_t + (1 - \delta)K_{t-1}$, onde δ é constante.

onde P é o deflator do ativo. Assim, a taxa de depreciação total varia tanto no tempo quanto entre países (INKLAAR E TIMMER, 2013).

Após a construção de cada série de ativos, rk^{na} é construído pela agregação das séries de ativos a partir de um índice de Törnqvist das taxas de crescimento dos ativos individuais, dado por:

$$\Delta \log RK_{i,t}^{NA} = \sum_a \frac{1}{2} \left(\frac{P_{a,i,t} K_{a,i,t}}{\sum_a P_{a,i,t} K_{a,i,t}} + \frac{P_{a,i,t-1} K_{a,i,t-1}}{\sum_a P_{a,i,t-1} K_{a,i,t-1}} \right) \Delta \log K_{a,i,t} \quad (3.37)$$

Adicionalmente, na PWT o estoque de capital inicial é calculado a partir da razão capital-produto:

$$K_0 = (Y_0)(ky) \quad (3.38)$$

Onde Y_0 é o produto inicial e ky é a razão capital-produto. Portanto, o cálculo de K_0 consiste em utilizar o produto no ano inicial e multiplicar por um valor assumido para ky ²³.

A PWT 9.0 também disponibiliza uma medida de capital humano e dados sobre a força de trabalho. Nesse trabalho consideramos $L_{i,t}$ como sendo o número de pessoas empregadas (em milhões). Para a variável de capital humano, é utilizado um indicador de capital humano baseado nos anos de estudo e seu respectivo retorno (hc). Esse índice é calculado por Barro e Lee (2013), bem como de Cohen e Leker (2014) e é computado igual em (3.3). Nesse caso,

$$hc_{i,t} = e^{\phi(s_{i,t})} \quad (3.39)$$

Onde $s_{i,t}$ são os anos médios de escolaridade para pessoas com 15 anos ou mais. Essa medida é construída considerando diferentes grupos de idade, sexo e níveis de ensino. Assim,

$$s_{i,t} = \sum_{a=1}^A l_{i,t}^a s_{i,t}^a \quad (3.40)$$

$$s_{i,t}^a = \sum_{j=1}^J h_{i,t}^a \mathcal{D}_{i,t}^a \quad (3.42)$$

Onde $l_{i,t}^a$ é a parcela da população pertencente ao grupo de idade a e $s_{i,t}^a$ é o número o nível de escolaridade do grupo a . Essa última variável é definida por (3.42), onde $h_{i,t}^a$ representa a fração de a com nível de educação j e $\mathcal{D}_{i,t}^a$ é a duração de anos de cada nível de educação. Para o índice são considerados quatro níveis de estudo (sem estudo, ensino primário, ensino secundário e ensino terciário/superior), e os grupos de idade definidos em cortes de 5 anos a partir dos 15 anos de idade.

²³ Os valores dos ativos de informação e tecnologia de comunicação (ICT) têm valor inicial zero. Para os demais ativos, os ativos de estrutura têm uma razão inicial de 2.2, os equipamentos de transporte têm uma razão inicial de 0.1 e os demais ativos e máquinas têm razão inicial de 0.3.

Por sua vez, $\phi(s_{i,t})$ é definido como

$$\phi(s_{i,t}) = \begin{cases} (0.134)s_{i,t} & \text{se } s_{i,t} \leq 4 \\ (0.134)4 + (0.101)(s_{i,t} - 4) & \text{se } 4 < s_{i,t} \leq 8 \\ (0.134)4 + (0.101)4 + (0.068)(s_{i,t} - 8) & \text{se } s_{i,t} > 8 \end{cases} \quad (3.43)$$

Conhecidos $Y_{i,t}$, $h_{i,t} = e^{\phi(s_{i,t})}$, $L_{i,t}$ e $K_{i,t}$, a série de PTF de cada país é calculada como um resíduo. Para isso, basta isolar $A_{i,t}$ igual em (3.2).

Para a estimação do painel dinâmico, utilizamos dados referentes à inflação, indicador do grau de abertura comercial, capital humano, gasto do governo como proporção do PIB e diferentes variáveis institucionais. Alguns desses dados estão disponíveis nas Penn World Table: o estoque de capital humano, já apresentado, o grau de abertura comercial, calculado como a soma das exportações e importações sobre o PIB e o gasto do governo como proporção do PIB. As parcelas da exportação, importação e gasto do governo no PIB estão disponíveis na própria PWT 9.0, todas calculadas a PPP corrente, tomando como base o PIB pela ótica da produção ($CGDP_0$). Por sua vez, os dados de inflação foram obtidos através dos Indicadores de Desenvolvimento Mundial (World Development Indicators), compilados pelo Banco Mundial. Essa base de dados inclui 217 países, com dados de 1960 a 2015.

Para as variáveis institucionais, utilizamos três diversas fontes diferentes. Ao todo utilizamos três indicadores de instituições. O primeiro é o índice de políticas governamentais anti-desvio (GADP), de Hall e Jones (1999) e disponibilizado em Morier e Teles (2011). O indicador foi construído por uma firma especializada em prover avaliação de risco para investidores internacionais, a *Political Risk Services*. Sua base de dados, quando usada pelos autores, contava com avaliações de 130 países, incluindo 24 categorias.

A segunda variável utiliza os mesmos dados que McGuiness (2007), que reúne um conjunto de 20 variáveis provenientes dos artigos de Botero et. al. (2004), Gleaser et. al. (2004) e Djankov et. al. (2004), bem como do Projeto Polity IV. Abaixo segue uma tabela com uma breve descrição das variáveis e suas respectivas fontes.

Tabela 2 - Variáveis utilizadas na construção do segundo indicador de instituições.

Variáveis	Definição e fonte
Government Effectiveness	Mensura a qualidade da provisão de serviços públicos. World Governance Index – World Bank.
Executive Constraints	Mensura a extensão das restrições institucionais no poder do chefe do executivo. Polity IV.
Expropriation Risk	Risco de confisco e nacionalização forçada de propriedade. Gleaser et. al. (2004).
Corruption	Índice de percepção da corrupção para 1999. Djankov et. al. (2002).
Civil Rights	Combinação de duas medidas de direitos civis, uma de Djankov et. al. (2002) e

Variáveis	Definição e fonte
	outra de Botero et. al. (2004).
Property Rights	Índice de direitos de propriedade dos cidadãos. Djankov et. al. (2002).
Court Index	Combinação de dois índices de formalismo tribunal – coleção de cheques devolvidos e despejo de inquilinos não-pagantes. Botero et. al. (2004).
Alternative contract Index	Mensura a existência e custos de formas alternativas às formas de contrato padrão. Botero et. al. (2004).
Male (Female) Participations	Taxa de participação do homem (mulher) como porcentagem da população total entre 15 e 64 anos. Botero et. al. (2004).
Minimum Age	Idade mínima para que uma criança possa ser empregada como aprendiz ou num emprego de tempo integral, excluindo empregos insalubres, rurais, noturnos e em negócios familiares. Botero et. al. (2004).
Minimum Wage	Salário Mínimo. Botero et. al. (2004).
Labour Index	Índice de proteção do trabalho e leis trabalhistas. Média de (i) contratos de emprego alternativos; (2) custos da hora extra; (3) custo de demitir empregados; e (4) processos de demissão. Botero et. al. (2004).
Response	Responsividade do sistema de saúde. Djankov et. al. (2002).
ISO ave	Combinação do ISO ambiental e padrão de qualidade. Djankov et. al. (2002).
Sick Index	Mensuração do nível de enfermidades e benefícios de saúde. Botero et. al. (2004).
Population at risk of malaria	Porcentagem da população sob o risco da Malária em 1994. Gleaser et. al. (2004).
% Population in temperature zone	Porcentagem da população de um país na zona de temperatura de Koeppan-Geiger em 1995. Gleaser et. al. (2004).
Social Benefits	
Industrial Index	Mensura a proteção às leis de relações coletivas como média de: (i) poder dos sindicatos; e (ii) disputas coletivas. Botero et. al. (2004).
Social Security Index	Mensura os benefícios da seguridade social como média de: (i) benefícios por morte, idade ou invalidez; (ii) benefícios por doença e saúde; (iii) seguro desemprego. Botero et. al. (2004).

Por fim, o último indicador é calculado a partir das estimativas de (i) Controle da Corrupção; (ii) Efetividade do Governo; (iii) Estabilidade Política e Ausência de Violência/Terrorismo; (iv) Qualidade Regulatória; (v) Império da Lei; e (vi) Voz e *Accountability*. Essas estimativas foram produzidas por Kaufmann, Aart e Mastruzzi (2010) e estão disponíveis no *World Governance Indicators*. O projeto reúne dados para 215 países entre 1996 e 2014.

Também utilizamos um indicador de liberdade econômica, calculada pela *Heritage Foundation*. O índice é composto por dez indicadores: propriedade privada, liberdade de não ter corrupção (*freedom from corruption*), liberdade fiscal, gasto do governo, liberdade de negócios (*business freedom*), liberdade de trabalho, liberdade monetária, liberdade de comércio, liberdade de investimento e liberdade financeira. Os dados estão disponíveis para 186 países desde 1995 até 2016.

4. RESULTADOS

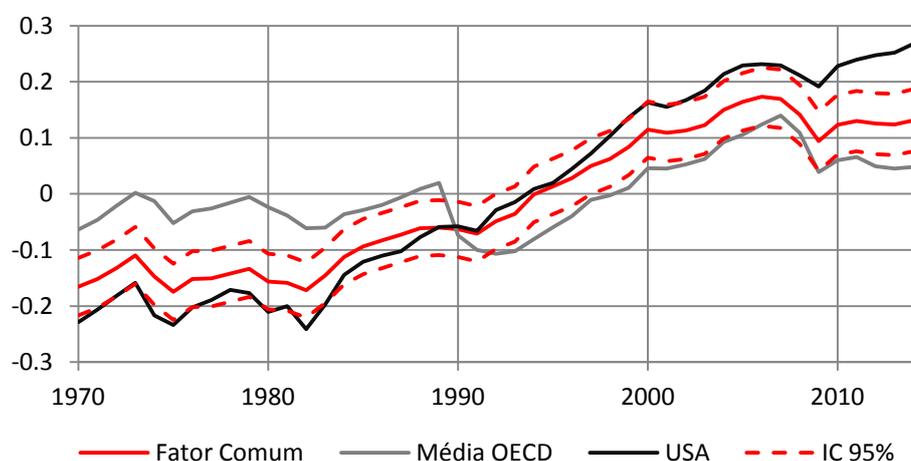
Para considerar o efeito de *catch up*, utilizamos uma variável que é definida como sendo o *gap* entre a PTF de cada país e a PTF mundial. Escolhemos três candidatos para representar a PTF mundial: média da OCDE (geométrica), PTF dos Estados Unidos e uma estimativa da PTF mundial. Esta última consiste no fator comum de um modelo de fatores dinâmicos para os 10 países com maior nível de eficiência técnica, segundo ranking de Pires e Garcia (2011). A correlação entre as variáveis está apresentada na tabela abaixo. É possível notar que todas as variáveis estão bastante correlacionadas entre si, em especial, há forte correlação entre a PTF americana e o fator comum. Abaixo também se encontra o gráfico com as variáveis centradas em torno de suas respectivas médias. Uma questão é se essas variáveis se comportam de forma parecida com o que esperaríamos da PTF mundial, segundo a teoria econômica.

Tabela 3 - Correlações entre *proxies* da PTF mundial

Correlações	Média OCDE	Estados Unidos	Fator Comum
Média OCDE	1	-	-
Estados Unidos	0.737	1	-
Fator Comum	0.765	0.992	1

Fonte: cálculo dos autores.

Figura 2 - Séries de *proxies* da PTF mundial entre 1970 e 2014 (em log e centradas em torno de suas médias)



Fonte: Penn World Table 9.0 e cálculo dos autores.

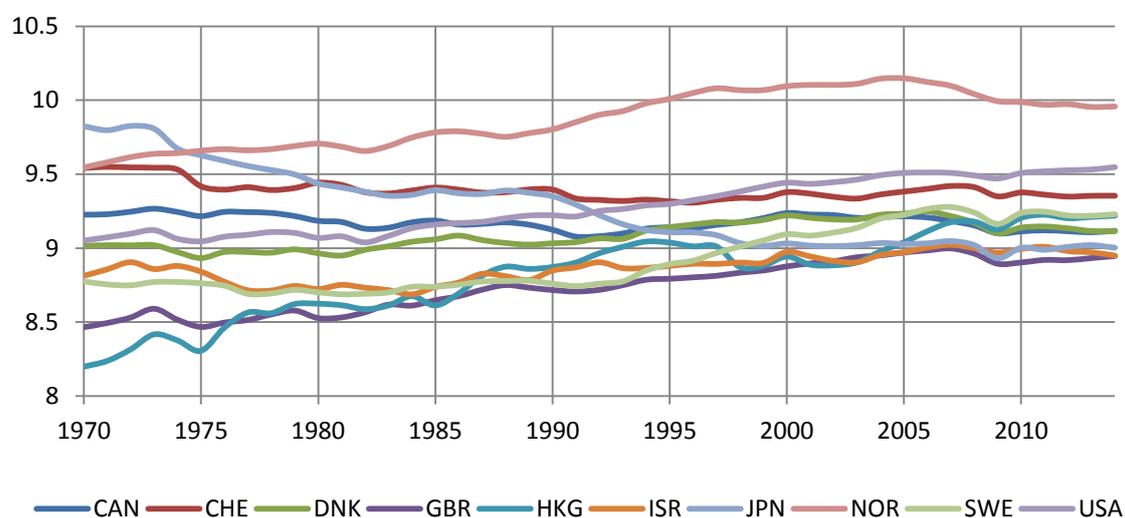
4.1. Análise Descritiva

De 1950 (início das séries da Penn World Table) até 70, a produtividade de praticamente todos os países selecionados (quando há dados) é ascendente em maior ou menor grau, com destaque para o Japão. De 1970 em diante (período coberto pela nossa

amostra), há um misto entre trajetórias ascendentes e estagnação. Noruega, Dinamarca, Hong Kong, Reino Unido e Estados Unidos apresentam trajetórias ascendentes entre 1970 e 2014. Para os demais, o período foi marcado por estagnação em suas respectivas PTFs. Entretanto, especificamente para a Suécia é possível notar uma leve trajetória ascendente após 1993. A PTF do Japão calculada a partir dos dados da Penn World Table 9.0 apresenta clara trajetória descendente, enquanto a mesma calculada a partir dos dados da Penn World Table 8.0 possui uma trajetória praticamente estagnada. Essa diferença entre as bases de dados parece ter tido um efeito significativo apenas sobre a PTF do Japão.

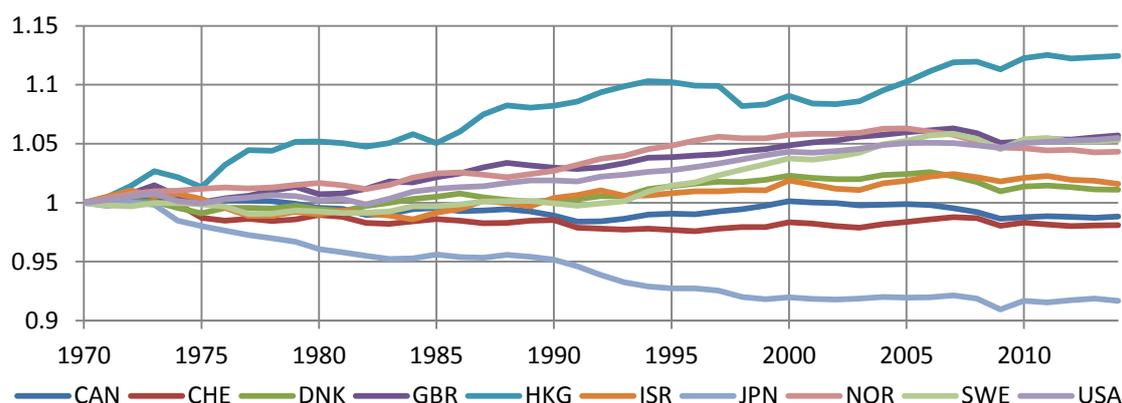
Em todo caso, apesar das diferenças de produtividade, é possível notar algumas características comuns a todas as PTFs, como, por exemplo, o fato de que durante a crise de 2008-2009 a PTF de todos os países caiu, com alguns casos de recuperação e outros não (caso da Noruega). Outra coisa a se notar é que as PTFs partem de níveis distintos, quando as séries são postas relativamente ao seu ponto de partida, podemos observar que Hong Kong, por exemplo, apresenta um ritmo de crescimento acelerado em sua PTF. Não só isso, exceto pela Suíça, Canadá e Japão, todos os demais países apresentavam uma PTF maior nos anos mais recentes relativamente a 1970.

Figura 3 - PTFs de países selecionados entre 1970 e 2014 (em log).



Fonte: Penn World Table (PWT) 9.0.

Figura 4 - Evolução da PTF de países selecionados, entre 1970 e 2014.



Fonte: Penn World Table (PWT) 9.0 e cálculo dos autores. Obs: 1970 = 1.

Abaixo estão as correlações entre cada série. As correlações indicam o que já era visível nos gráficos: as séries do Canadá, Suíça e Japão destoam do resto, apresentando correlação baixa ou mesmo negativa em relação às demais séries. Em compensação, as demais séries estão bem correlacionadas entre si. Em outras palavras, as séries de PTFs com maior eficiência técnica, em geral, andam juntas, mas há exceções.

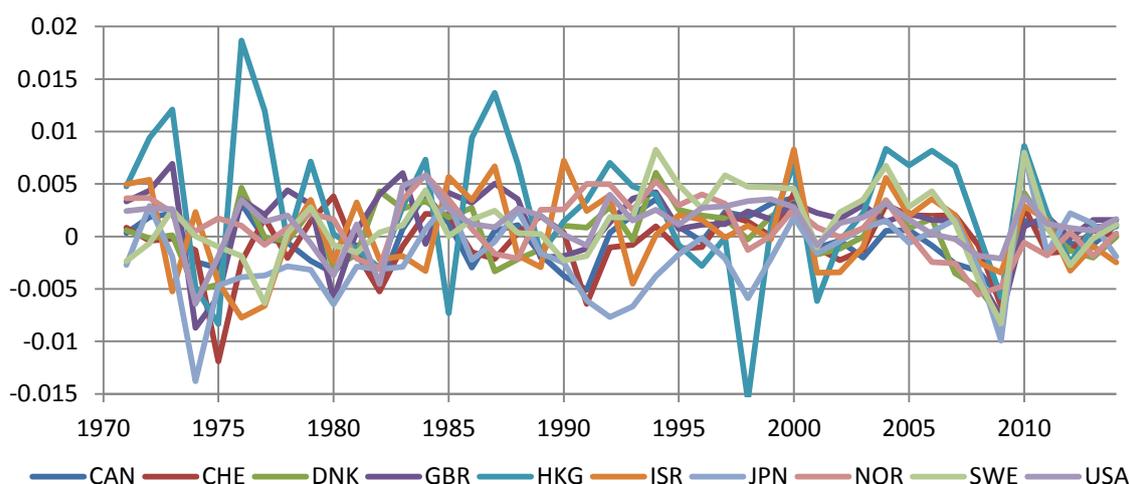
Tabela 4 - Correlações entre as PTFs dos países selecionados.

Correlações	USA	JPN	CHE	GBR	ISR	SWE	CAN	HKG	DNK	NOR
USA	1									
JPN	-0.8949	1								
CHE	-0.5042	0.7831	1							
GBR	0.9748	-0.9195	-0.5687	1						
ISR	0.8608	-0.644	-0.1836	0.8077	1					
SWE	0.9621	-0.7994	-0.3249	0.9124	0.8829	1				
CAN	-0.3129	0.5044	0.6132	-0.357	-0.1982	-0.1743	1			
HKG	0.8815	-0.9273	-0.7194	0.911	0.6746	0.7701	-0.628	1		
DNK	0.8894	-0.8189	-0.4499	0.9102	0.7745	0.8501	-0.109	0.7289	1	
NOR	0.91	-0.928	-0.6492	0.9336	0.7472	0.83	-0.2649	0.8372	0.9439	1

Fonte: cálculo dos autores.

Outra forma de tentar verificar similaridades entre as séries de PTF é observando suas respectivas taxas de crescimento, apresentadas no gráfico abaixo. Durante a maior parte do tempo quando uma série apresenta alta taxa de crescimento, as demais também apresentam crescimento, ainda que em magnitudes diferentes. É válido notar, entretanto, que há algumas exceções: por exemplo, entre 1975 e 1980 é possível notar que, enquanto a PTF de vários países crescia (em especial Hong Kong), as PTFs sueca e israelense apresentaram crescimento negativo.

Figura 5 - Taxas de crescimento das PTFs (países selecionados) entre 1970 e 2014.



Fonte: Penn World Table (PWT) 9.0 e cálculo dos autores.

4.2. Resultados: Modelo de Fatores Dinâmicos

As estimativas do modelo de fatores dinâmicos se encontram na Tabela 5. Uma primeira coisa que chama atenção são os altos coeficientes autorregressivos das equações do fator comum e fatores específicos. Em específico, o fator comum não é estacionário²⁴. Os coeficientes de *loading* estão normalizados em relação aos Estados Unidos, e mostram a sensibilidade das PTFs em relação ao fator comum, permitindo uma comparação entre países. Desse modo, o Japão e Suíça são os países menos sensíveis ao fator comum, enquanto Hong Kong é o país mais sensível da amostra. Os Estados Unidos, por sua vez, fica atrás de vários outros países: é a 4^o variável mais sensível do modelo. Vale notar, porém, que o *loading* da produtividade japonesa não é significativo a 10% de confiança (porém chega próximo, visto o p-valor de 0.14), colocando em questão a validade do coeficiente. Apesar disso, isso não significa necessariamente que seu valor seja zero, o *loading* do Japão é significativo quando utilizamos os dados da Penn World Table 8.0. Essa diferença se deve provavelmente pelo fato de que há diferenças significativas entre a PTF Japonesa calculada a partir dos dados da versão 8 e 9. Apesar disso, as demais PTFs não sofreram grandes modificações e os resultados e conclusões tiradas aqui são robustas às duas versões da base de dados.

²⁴ O teste de Dickey-Fuller Aumentado sem constante e com 1 *lag* é incapaz de rejeitar a hipótese de raiz unitária para o fator comum ($\tau = -0.777$).

Uma primeira pergunta a se fazer é o que significa um país ser mais ou menos sensível ao fator comum. Os modelos apresentados em Acemoglu (2009) (modelo *benchmark*), Nelson e Phelps (1966), Aiyar e Feyrer (2002) e Damsgaard e Krusell (2010) trabalham com a noção de uma PTF mundial, que poderia ser definida tanto como uma fronteira tecnológica ou um agregado de todas as PTFs combinadas de alguma forma (ao trabalhar com uma PTF mundial endógena, Acemoglu (2009) considera a média aritmética de todas as PTFs, por exemplo). Nesses modelos é pressuposto que o canal de comunicação entre as diferentes PTFs é apenas via difusão tecnológica, uma vez que tais modelos desconsideram a existência de outras interações entre países. Além desse canal, o modelo de fatores dinâmicos captura os choques comuns a todos os países, tal como um choque de petróleo ou uma crise. Sendo assim, é de se esperar que o fator comum e a sensibilidade de cada país em relação ao fator comum reflitam em alguma medida a distribuição de variáveis relacionadas à tecnologia e choques de produtividade. Pelo gráfico do fator comum, fica aparente que os choques de 1973 e 1979 impactaram o fator comum: há um vale em 1975 e outra queda a partir de 1979. Há outro choque visível na série, no período dado pela crise de 2008-2009: há uma queda brusca no período, seguido por uma leve recuperação.

Tabela 5 – Estimativas do modelo de fatores dinâmicos

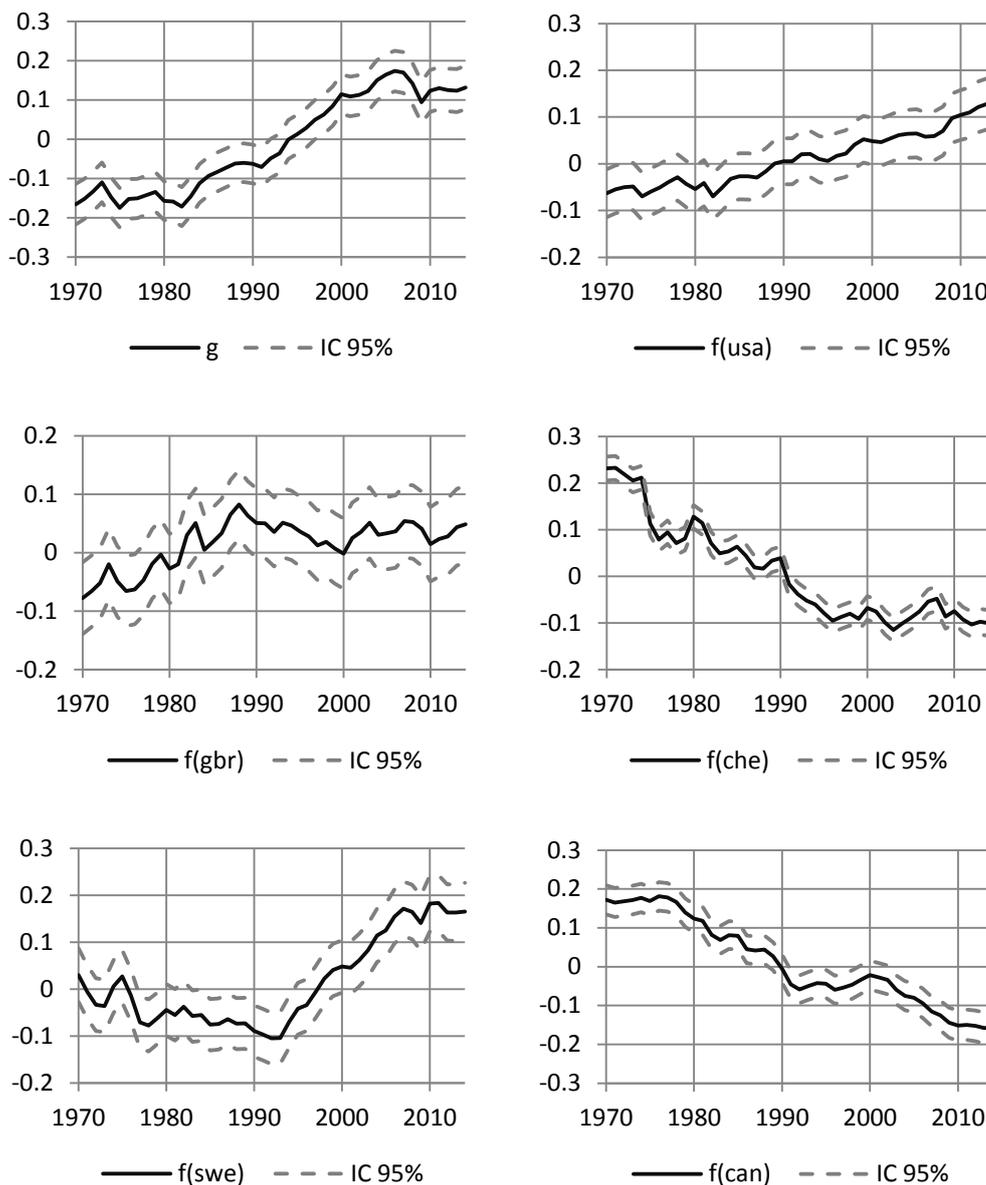
Modelo de Fatores Dinâmicos					
	λ_i^*	ρ	ϕ_i	σ_v^2	σ_u^2
g_t	-	0.9896181***	-	0.0004243***	-
	-	(0.0139428)	-	(0.0001304)	-
USA	1	-	1.018787***	-	0.0001793**
	(constrained)	-	(0.0603577)	-	(0.0000635)
JPN	0.4772252	-	0.9969237***	-	0.0014349***
	(0.3237045)	-	(0.004331)	-	(0.0003118)
CHE	0.5007337**	-	0.9866081***	-	0.0006383***
	(0.2279851)	-	(0.0183165)	-	(0.0001418)
GBR	1.197533***	-	0.936609***	-	0.0004158***
	(0.2266115)	-	(0.1198689)	-	(0.0001143)
ISR	0.5521987***	-	0.8020176***	-	0.0009759***
	(0.1871239)	-	(0.1057241)	-	(0.0002122)
SWE	1.107508***	-	0.9778699***	-	0.000537***
	(0.2383259)	-	(0.03178)	-	(0.0001365)
CAN	0.7299003***	-	0.9945431***	-	0.0002629***
	(0.1606663)	-	(0.0076357)	-	(0.0000698)
HKG	1.626537***	-	0.9730336***	-	0.002701***
	(0.476927)	-	(0.0386779)	-	(0.0006141)
DNK	0.8601812***	-	0.954034***	-	0.0003133***

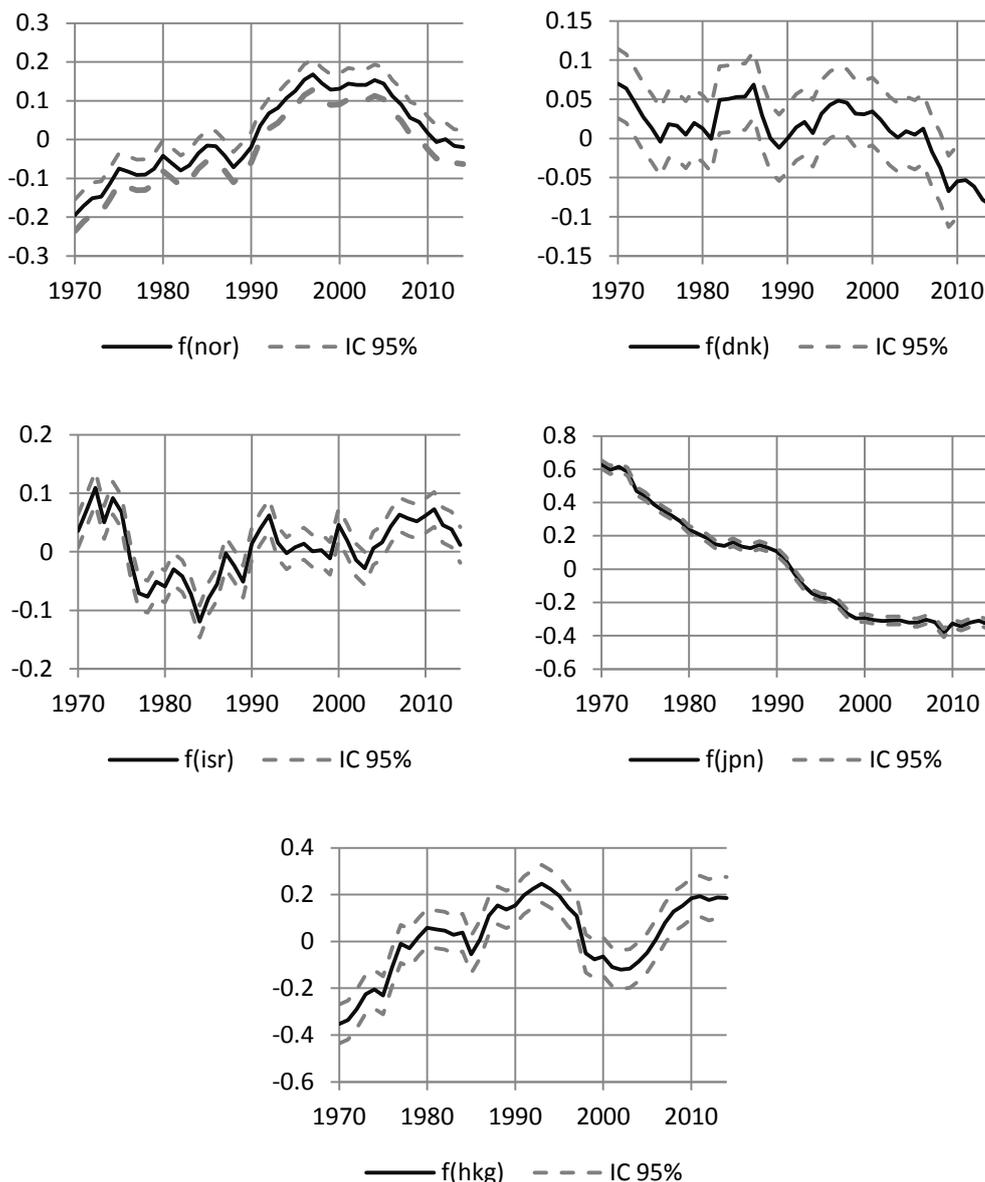
	(0.2279887)	-	(0.078188)	-	(0.0000849)
NOR	0.7932897***	-	0.9829859***	-	0.000532***
	(0.2042891)	-	(0.0217913)	-	(0.0001222)
Amostra	1970-2014				

Erros-padrão em parêntesis; *** p<0.01, ** p<0.05, * p<0.1.

Fonte: cálculo dos autores.

Figura 6 - Fator comum e Fatores específicos.





Fonte: cálculo dos autores.

Quanto à relação entre os *loadings* e os transbordamentos tecnológicos, a princípio, se um país depende muito do fator comum (refletido por um *loading* alto), então seria de se esperar que o país esteja mais sujeito a transbordamentos tecnológicos, enquanto os países com baixo *loading* estariam menos sujeitos a isso – isto porque seria de se esperar que um país menos avançado em termos tecnológicos tenha mais espaço para absorver da PTF mundial (efeito de *catch up*). Vale notar também que o peso da média ponderada dada pelo fator comum sugere que quanto menor o *loading*, maior será o peso daquele país na composição do fator comum. Se esta estimativa captura bem a PTF mundial, então um país tecnologicamente avançado deveria ter um *loading* pequeno.

Dito isto, um exercício interessante que pode ser feito é comparar a ordenação dada pelos *loadings* com algumas variáveis de produtividade e tecnologia. Para isso, selecionamos alguns possíveis candidatos: (i) gasto em P&D como proporção do PIB; (ii) número de pesquisadores empregados em atividades de P&D; (iii) capital humano²⁵; (iv) PTF; (v) taxa de crescimento da PTF; e (v) um índice de mudança de eficiência tecnológica. As duas primeiras medidas estão disponíveis nos principais indicadores de ciência e tecnologia da OECD, com dados disponíveis a partir de 1981. As medidas de capital humano e PTF estão disponíveis na própria Penn World Table 9.0, utilizada neste trabalho. Finalmente, a mudança de eficiência tecnológica foi calculada por Pires e Garcia (2011), que construíram o indicador decompondo a PTF através de uma abordagem de fronteira estocástica para 75 países entre 1950 e 2000, e a partir disso fizeram um *ranking* para 1995-2000, com dados referentes às taxas médias anuais entre 1996 e 2000²⁶. Os dados para as cinco variáveis e nossos *loadings* estão na tabela abaixo.

Tabela 6 - Indicadores de tecnologia e produtividade vs *loadings*

	ISR	CHE	CAN	JPN	NOR	SWE	USA	DNK	GBR	HKG
λ_{DFM}^* ^a	0.552	0.501	0.730	0.477	0.793	1.108	1.000	0.860	1.198	1.627
Gasto em P&D (% do PIB) ^b	3.832	2.595	1.868	3.165	1.585	3.471	2.598	2.426	1.692	0.633
Pesquisadores em P&D (por mil trabalhadores) ^b	7296. 25	3380. 07	3947. 39	5171. 09	4809. 63	5233. 49	3661. 55	5061. 98	3503. 23	2127. 68
Índice de Capital Humano ^c	3.148	3.450	3.314	3.200	3.220	3.097	3.461	3.144	3.268	2.606
PTF (em log) ^c	8.867	9.390	9.176	9.273	9.872	8.920	9.280	9.087	8.742	8.821
PTF (Tx. de crescimento) ^c	0.003	-0.004	-0.002	-0.019	0.009	0.010	0.011	0.002	0.011	0.023
Eficiência Técnica (US\$) ^d	0.811	0.872	0.770	0.899	0.740	0.778	0.955	0.761	0.820	0.763

Fontes:

^a Cálculo dos autores.

^b World Bank, World Development Indicators, January 2016, valor médio 1996-2011;

^c Penn World Table 9.0, valor médio 1970-2014;

^d Pires e Garcia (2011), eficiência técnica sem paridade do poder de compra.

Como dito, seria de se esperar que os países com menor sensibilidade em relação ao fator comum fossem mais tecnologicamente avançados, sob a noção de que eles estariam mais próximos da fronteira e, portanto, seriam menos capazes de absorver da

²⁵ Seguindo a hipótese de Nelson-Phelps, e evidência de Aiyar e Feyrer (2002), um maior nível de capital humano facilita a adoção de novas tecnologias, então seria de se esperar que esta variável esteja associada ao nível de desenvolvimento tecnológico de um país.

²⁶ Pires e Garcia (2011) constroem dois *rankings* de eficiência técnica: um usando o PPP ajustado pela paridade de poder de compra (PPP) e outro não. Isto porque ao ser ajustado pelo PPP, o *rank* reflete a eficiência na produção de *nontradables*, enquanto sua versão ajustada reflete a produtividade avaliada aos preços dos *tradables*.

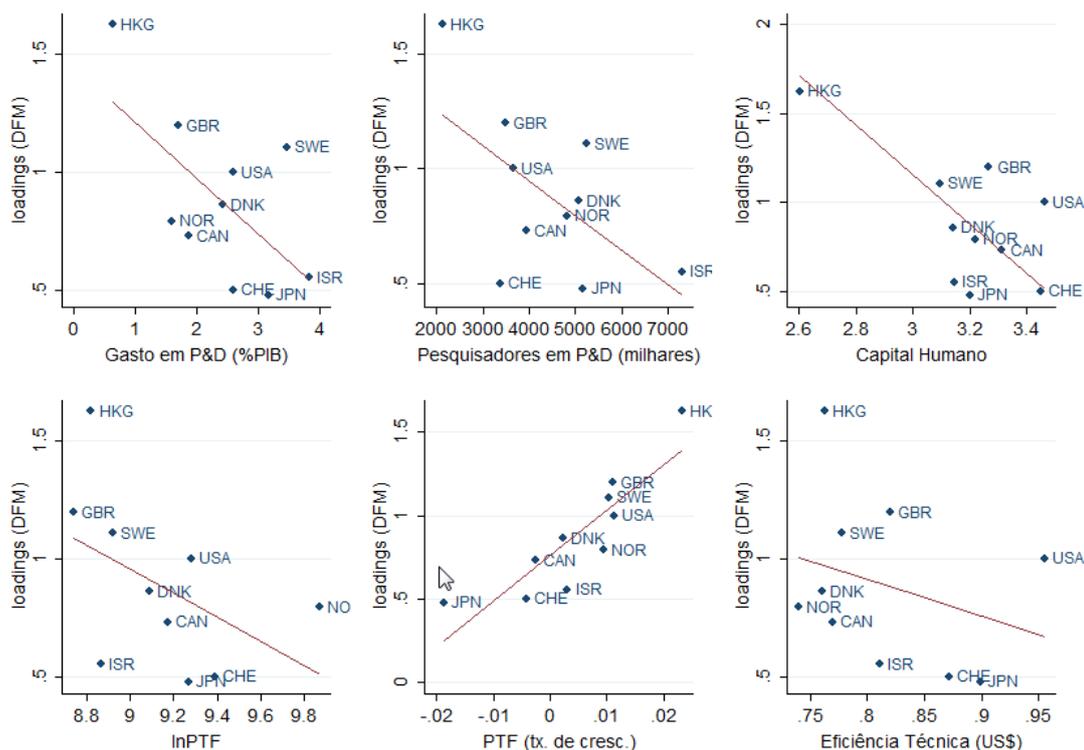
mesma. Como resultado, isso se refletiria numa menor sensibilidade em relação à PTF mundial. Com o objetivo de verificar a razoabilidade dessa interpretação, reunimos na Tabela 7 as correlações entre nossos *loadings* e as variáveis selecionadas. Os gráficos da Figura 7 ilustram tais relações. Em linha com o esperado, boa parte das variáveis selecionadas apresenta correlação negativa com os *loadings*, em especial as variáveis de Gasto em P&D e Pesquisadores em P&D. Uma exceção, porém, é a correlação com a taxa de crescimento da PTF, que apresenta forte correlação positiva com os *loadings*. Essas evidências seguem bem a linha do que esperaríamos para o comportamento dos *loadings*. Pelo lado das variáveis com correlação negativa, isso sinaliza que os países com baixo λ_i^* são justamente os países mais tecnologicamente avançados, como seria esperado pela teoria. Além disso, a forte correlação positiva com a taxa de crescimento da PTF também corrobora com o que se esperaria: países com menos tecnologia (λ_i^* alto) desfrutam de mais espaço para realizar o *catch up* tecnológico, levando a maiores taxas de crescimento de suas respectivas PTFs. É importante ressaltar, porém, que essas associações devem ser vistas com cuidado, uma vez que selecionamos um número bastante limitado de países para incluir no modelo de fatores dinâmicos, o que dificulta uma investigação mais rigorosa neste sentido. Para verificar se esse mesmo padrão se repete com outro conjunto de países, estimamos o modelo de fatores dinâmicos utilizando os países do G7, os resultados podem ser vistos no apêndice (apêndice B) deste trabalho. Como se pode observar, as correlações se repetem, exceto, porém, para a taxa de crescimento da PTF, que chama atenção por ser negativamente correlacionada neste caso.

Tabela 7 - Correlações entre *loadings* e indicadores de tecnologia e produtividade

Correlações	λ_{DEM}^*
Gasto em P&D (% do PIB)	-0.6302
Pesquisadores em P&D (por mil trabalhadores)	-0.5958
Índice de Capital Humano	-0.6548
PTF (em log)	-0.4799
PTF (Tx. de crescimento)	0.8659
Eficiência Técnica (US\$)	-0.3014

Fonte: cálculo dos autores.

Figura 7 - Variáveis selecionadas e loadings (DFM)



Fonte: cálculo dos autores e dados da OCDE, World Bank, PWT 9.0 e Pires e Garcia (2011).

Por fim, a Tabela 8 mostra os resultados para a decomposição da variância. Nós encontramos que o fator global explica, em média, 71.71% da produtividade dos países considerados. É importante ressaltar, porém, que há alguma heterogeneidade entre os países: enquanto o fator comum explica mais de 70% da variação de produtividade da maioria dos países, há casos em que o fator comum explica muito pouco, como é o caso da Suíça e Canadá. A correlação entre os coeficientes e o R^2 é positiva, mas não muito alta (0.39), ou seja, o fator comum é importante e explica em geral mais dos países distantes da fronteira, mas há exceções.

Tabela 8 - Decomposição da variância

$$PVB = \hat{\beta}_i^2 Var(g_t) / Var(\ln A_{i,t}^*)$$

	ISR	CHE	CAN	JPN	NOR	SWE	USA	DNK	GBR	HKG
$\ln A_{i,t}^*$	74.88%	22.97%	05.85%	78.62%	87.42%	92.62%	98.42%	86.79%	96.07%	73.42%

Fonte: cálculo dos autores.

4.3. Recuperando a constante do Fator Comum

Para facilitar a estimação do modelo de fatores dinâmicos, havíamos estimado o componente comum a partir das variáveis centralizadas em torno da média. Sendo assim, nossa estimativa do fator comum também não possui constante. Essa variável, porém, pode ser recuperada. Observe que o fator comum é identicamente igual a

$$g_t = \frac{\sum_{i=0}^{10} (\ln A_{i,t} - \ln \bar{A}_i - \ln f_{i,t})}{\sum_{i=0}^{10} \lambda_i 10}$$

Portanto somar g_t com $10^{-1} \sum_{i=1}^{10} (\ln \bar{A}_i / \lambda_i)$ para obter o fator comum não-centralizado em torno da média temporal. Com isso podemos construir os *gaps* de acordo com cada uma das nossas *proxies*. Na tabela abaixo se encontram as médias mundiais²⁷ de cada um dos *gaps* construídos. Uma primeira coisa que chama atenção é o tamanho do *gap* calculado pelo fator comum. Isto porque a média do fator comum calculada da forma acima é muito maior do que qualquer uma das PTFs. Isso em si não é um problema, visto que uma das formas de se entender a PTF mundial é como uma fronteira tecnológica. Apesar disso, os três *gaps* são fortemente correlacionados entre si e possuem trajetórias parecidas, como se pode ver nos gráficos abaixo. Num primeiro momento há uma grande divergência entre a média mundial e a PTF mundial, seguido por convergência a partir dos meados da década de 90.

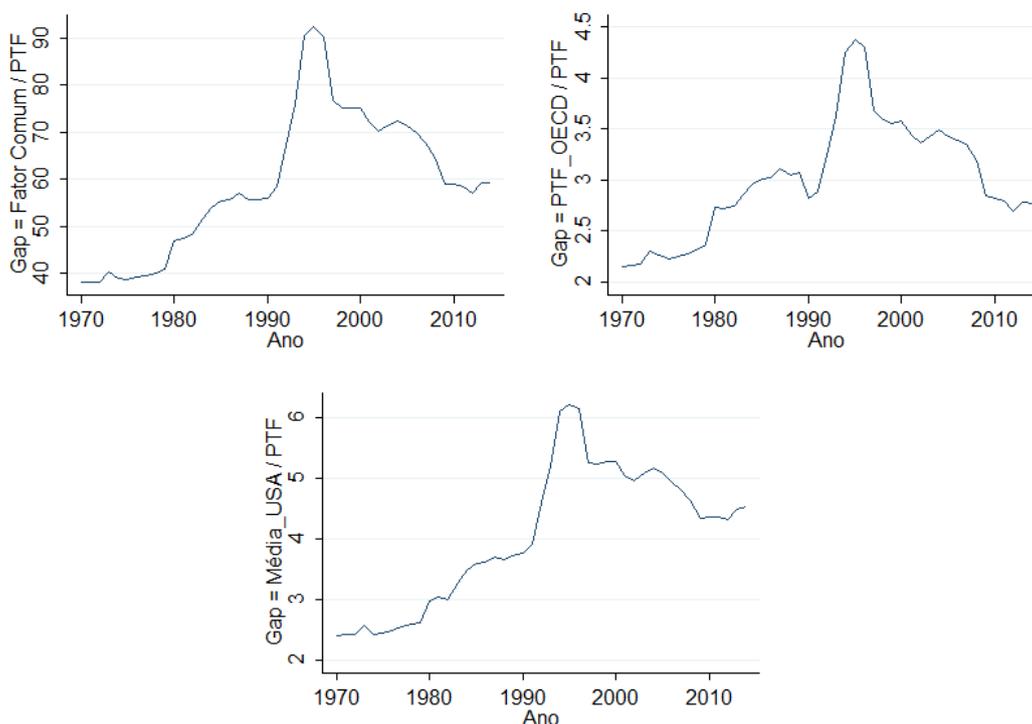
Tabela 9 - Estatísticas para os *gaps*

Variável	Média	Desvio-padrão	Min	Max
Gap(Fator Comum)	59.2007	14.94227	38.10699	92.39375
Gap(OCDE)	3.011525	0.570094	2.154008	4.379024
Gap(USA)	4.043666	1.134121	2.392994	6.209536
Correlações	Gap(Fator Comum) Gap(OCDE) Gap(USA)			
Gap(Fator Comum)	1			
Gap(OCDE)	0.9795	1		
Gap(USA)	0.9856	0.9367	1	

Médias temporais da média mundial de cada *gap*.

Fonte: cálculo dos autores.

²⁷ Utilizando os países da PWT 9.0 com dados de PTF. Ao há 144 países com dados para algum ano entre 1970 e 2014.

Tabela 10 - Estimativas para o *gap* entre PTF mundial e PTFs. Médias temporais dos *gaps* médios.

Fonte: cálculo dos autores.

4.4. Resultados: GMM Sistêmico

Nesta seção buscamos estimar os parâmetros das equações (3.26-27) para um painel de 144 países entre 1970 e 2014. Como nem todos os países possuem dados disponíveis para todas as variáveis, optamos por utilizar a transformação de desvios ortogonais ao invés de primeira diferença. Seja x uma variável qualquer, a transformação de desvios ortogonais é dada por:

$$x_{i,t+1}^{\perp} = c_{i,t} \left(x_{i,t} - \frac{1}{T_{i,t}} \sum_{s>t} w_{i,s} \right)$$

Isto é, ao invés de se subtrair a média temporal da variável, como ocorre na transformação de efeitos fixos, esta considera na soma apenas as observações futuras disponíveis. $T_{i,t}$ é o número dessas observações e $c_{i,t} = \sqrt{T_{i,t}/(1 + T_{i,t})}$. Esta transformação é recomendada quando há *missing*, visto que a primeira diferença pode exacerbar os *gaps* presentes na base de dados. Outra vantagem dessa transformação é que ela mantém as variáveis i.i.d, caso a variável original já fosse i.i.d. (ver ROODMAN, 2009).

Sendo assim, iremos primeiro verificar as estimativas utilizando diferentes indicadores institucionais. Uma vez selecionado um desses indicadores, iremos analisar

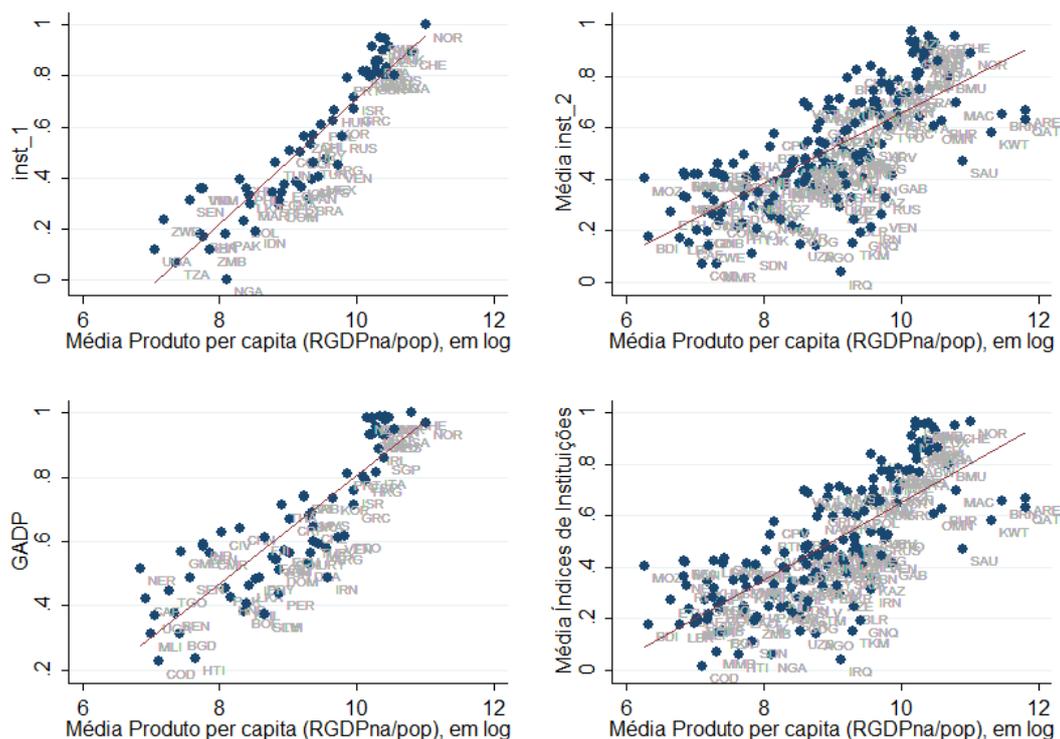
como esses resultados ficam para diferentes *proxies* para a PTF mundial: com o fator estimado na seção anterior, com a PTF americana e com a média das PTFs dos países da OCDE.

4.4.1. Instituições e Produtividade

Para mensurar o impacto das instituições, utilizamos três possíveis indicadores: (i) o índice de políticas governamentais anti-desvio (GADP), de Hall e Jones (1999) e disponível em Morier e Teles (2011), que é constante no tempo, mas diferente para cada país; (ii) um índice de instituições a partir dos dados de Botero et. al (2004), Gleaser et. al. (2004) e Djankov et. al. (2004). Para construção deste índice seguimos a formatação sugerida por McGuinness (2007), que reúne diferentes variáveis em cinco grupos (governo político, direitos do cidadão, regulamentação do trabalho, benefícios de saúde e benefícios sociais). A partir dessa separação, combinamos as variáveis de cada subgrupo por meio de análise de fatores dinâmicos, a fim de reduzir a dimensionalidade do vetor de cada grupo, e então combinamos os fatores via componentes principais. O resultado final foi normalizado para variar entre 0 e 1 (*inst_1*); e (iii) um índice construído a partir dos índices de voz e accountability, estabilidade política e ausência de violência, efetividade do governo, qualidade regulatória, controle da corrupção e liberdade econômica. Os 6 primeiros indicadores são calculados pela Worldwide Governance Indicators (WGI), e o último pela Herritage Foundation. Essas variáveis foram combinadas via análise de componentes principais e o indicador final foi normalizado para variar entre 0 e 1 (*inst_2*) Diferente dos demais, este varia ao longo do tempo, mas só possui dados a partir de 1996 e 1995. Na posse desses três indicadores, renormalizamos a variável GADP para variar entre 0 e 1 e então tomamos a média dos três indicadores de instituições (*Minst*).

A Figura 8 mostra cada variável em relação à média do produto per capita (em log) de cada país. Como se nota, todos os indicadores estão fortemente relacionados com o nível de renda de cada país. O indicador *inst_2* é o que consegue reunir o maior conjunto de países, além do *Minst*, que é a média das três variáveis.

Figura 8 - Correlação entre Índices de Instituições e Produto per capita (em log).



Fonte: cálculo dos autores.

Na Tabela 11 se encontram as estimativas do SGMM com os determinantes sugeridos pela literatura. Estimamos o mesmo modelo quatro vezes, modificando apenas a variável institucional usada, para cada uma consideramos o modelo com e sem a interação entre o fator comum e pertencer ou não à OCDE. Na tabela estão as estimativas para o SGMM em um estágio e robustas à heterocedasticidade e autocorrelação serial. Ao considerar os instrumentos, todas as variáveis que não a própria variável dependente foram colapsadas (isto é, foi criado um instrumento para cada variável e lag, ao invés de ter um instrumento para cada período, variável e lag). Isto pode diminuir a eficiência, mas impede um excesso de instrumentos, que seria problemático em painéis pequenos. As *dummies* de tempo estão presentes em todos os modelos, mas foram omitidas da tabela.

Além da própria PTF, as variáveis de inflação, abertura comercial, capital humano e gasto do governo são considerados pré-determinados, mas não necessariamente estritamente exógenos, nos permitindo usar os *lags* dessas variáveis como instrumentos também. Em todos os 8 modelos o teste de Arellano-Bond aponta para presença de autocorrelação de primeira ordem e de segunda ordem em alguns

casos, no termo de erro idiossincrático, $v_{i,t}$. Sendo assim, apenas as defasagens da PTF a partir de $t - 4$ foram usadas. Além disso, as variáveis $inst_2$ e $Minst$ variam no tempo, portanto também podemos tentar considerar suas defasagens como instrumentos. Neste caso, também utilizamos defasagens somente a partir de $t - 4$, visto que provavelmente essas variáveis contém algum tipo de erro de mensuração, invalidando o segundo e terceiro *lag* (ver BOND, HOEFFLER e TEMPLE, 2001).

A primeira coisa que chama atenção é o fato de que apenas a variável GADP é significativa a 10% nos modelos (1) e (2). Visto que as variáveis GADP e $inst_1$ são constantes no tempo e que $inst_2$ e $Minst$ variam apenas a partir de 1995, talvez não surpreenda que o GMM Sistêmico, utilizando apenas variáveis defasadas como instrumentos, seja incapaz de limpar a endogeneidade presente na relação entre produtividade e instituições. Em todo caso, o coeficiente de GADP é positivo, sugerindo que melhores instituições levam a produtividades maiores.

Em relação às demais variáveis, em todos os modelos o fator comum é significativo e nos modelos sem interação os testes de Wald não conseguem rejeitar a hipótese de que $\omega + \theta = 1$, que seria previsto pelo modelo teórico. As estimativas também sugerem que países mais ou menos desenvolvidos possuem ω s diferentes. O coeficiente para os países que não pertencem à OECD é sempre positivo, enquanto o coeficiente dos países da OECD é negativo ou não-significativo. Isto também está em concordância com o que esperaríamos: o efeito de *catch up*, capturado por ω , é maior para os países menos desenvolvidos. Seguindo a formulação de Aiyar e Feyrer (2002), isso implicaria que o coeficiente de convergência condicional, γ , é de, em média, 1.13%, bem menor do que o encontrado pelos referidos autores.

Por sua vez os demais determinantes não são significativos na maioria dos casos, mas quando o são, vão na direção esperada. A inflação, pelo modelo (1) e (2), impacta negativamente a acumulação de PTF; Isto sugere a importância política econômica para o crescimento, ainda que o gasto do governo não tenha sido significativo. Por sua vez, o coeficiente de abertura comercial é positivo e significativo nos modelos (3) e (4), sugerindo que países mais abertos ao comércio internacional possuem maior produtividade; e, por fim, o capital humano só é significativo a 10% no modelo (3), mas o coeficiente dos modelos (5) e (6) são significativos a 11% (p-valores de, respectivamente, 0.103 e 0.105). Estes três coeficientes são similares em magnitude e positivos, sugerindo a validade da hipótese de Nelson-Phelps de que o capital humano,

além de afetar o produto diretamente, também o afeta através do progresso técnico. É interessante notar que a média desses coeficientes, quando significativos, implica $\phi_{hc} = 4.64$, ou seja, um aumento de 1% no estoque de capital humano levaria a um aumento de aproximadamente 4.6% no produto potencial. Aiyar e Feyrer (2002) encontram um ϕ igual a 4.41, próximo ao encontrado aqui.

Tabela 11 - GMM Sistêmico em um estágio para diferentes índices de instituições

SGMM em um estágio	GADP		Inst_1		inst_2		Minst	
Variáveis	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fator comum	0.062*** (0.013)		0.076*** (0.021)		0.046** (0.022)		0.052*** (0.017)	
Fator comum se OCDE=0		0.067*** (0.014)		0.083*** (0.022)		0.060*** (0.020)		0.052*** (0.015)
Fator comum se OCDE=1		-0.370** (0.168)		-0.206 (0.168)		-0.523 (0.377)		-0.334** (0.150)
PTF	0.921*** (0.040)	0.951*** (0.038)	0.918*** (0.042)	0.910*** (0.039)	0.966*** (0.042)	0.950*** (0.044)	0.950*** (0.036)	0.951*** (0.033)
OCDE	-0.014 (0.062)	5.243** (2.059)	-0.003 (0.049)	3.468* (2.125)	-0.02124 (0.085)	6.996 (4.553)	0.069 (0.061)	4.663** (1.817)
$y_{i,0}$	-0.006 (0.038)	-0.018 (0.037)	-0.037 (0.031)	-0.024 (0.029)	-0.04376 (0.036)	-0.039 (0.035)	-0.016 (0.025)	-0.020 (0.025)
Inflação	-0.022* (0.013)	-0.024** (0.013)	-0.014 (0.012)	-0.016 (0.012)	0.000445 (0.021)	0.008 (0.022)	-0.001 (0.010)	-0.003 (0.011)
Abertura Comercial	0.004 (0.036)	0.036 (0.035)	0.048* (0.025)	0.049** (0.023)	-0.00926 (0.074)	0.028 (0.068)	-0.019 (0.035)	0.004 (0.033)
Capital Humano	0.035 (0.112)	-0.116 (0.016)	0.227* (0.118)	0.129 (0.145)	0.2956926 (0.180)	0.264 (0.161)	0.067 (0.094)	0.072 (0.102)
Gasto do Governo	-0.026 (0.032)	-0.03411 (0.031)	0.008 (0.034)	0.020 (0.034)	0.022388 (0.047)	0.024 (0.046)	0.032 (0.031)	0.039 (0.030)
GADP	0.158* (0.083)	0.140* (0.081)						
inst_1			0.005 0.050	0.012 (0.044)				
inst_2					-0.05 (0.045)	-0.027 (0.035)		
Minst							0.012 (0.031)	0.013 (0.018)
T (média)	7.6	7.6	7.54	7.54	2.99	2.99	6.51	6.51
N	68	68	59	59	99	99	99	99
TN	517	517	445	445	296	296	644	644
Instrumentos	66	67	66	67	52	57	71	76
$H_0: \omega + \theta = 1$	0.6669	-	0.8534	-	0.7623	-	0.9279	-

Todas as variáveis estão defasadas (exceto variáveis constantes no tempo) e em logaritmos (exceto dummies).

Dummies de tempo foram omitidas.

(1), (3), (5), (7) = modelos sem interação entre fator comum e dummy OCDE;

(2), (4), (6), (8) = modelos com interação entre fator comum e dummy OCDE.

Fonte: cálculo dos autores.

Segue abaixo o mesmo conjunto de dados estimado pelo SGMM em dois estágios. Como no SGMM em dois estágios há um viés negativo nos desvios-padrão, aplicamos também a correção de Windmeijer (2005), que corrige este viés. Em geral os resultados permanecem os mesmos, com exceção da variável de capital humano no

modelo (3) e a interação do fator comum com a variável OCDE = 1 no modelo (6). Agora a variável de capital humano deixa de ser significativa em todos os modelos, enquanto a interação se torna significativa. O p-valor da variável de capital humano é igual a 0.176 agora.

Tabela 12 - GMM Sistemico em dois estágios para diferentes índices de instituições

SGMM em dois estágios	GADP		Inst_1		inst_2		Minst	
Variáveis	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fator comum	0.061*** (0.013)		0.076** (0.024)		0.042** (0.0190)		0.050*** (0.014)	
Fator comum se OCDE=0		0.066*** (0.012)		0.079** (0.027)		0.054*** (0.015)		0.055*** (0.012)
Fator comum se OCDE=1		-0.350** (0.160)		-0.196 (0.195)		-0.686* (0.401)		-0.317** (0.159)
PTF	0.923*** (0.036)	0.942*** (0.035)	0.915*** (0.045)	0.904*** (0.042)	0.960*** (0.049)	0.950*** (0.041)	0.943*** (0.031)	0.952*** (0.028)
OCDE	-0.013 (0.060)	4.996** (1.918)	0.017 (0.053)	3.297 (2.506)	-0.010 (0.082)	8.933* (4.896)	0.053 (0.056)	4.489** (1.930)
$y_{i,0}$	-0.005 (0.039)	-0.015 (0.033)	-0.023 (0.033)	-0.014 (0.028)	-0.039 (0.035)	-0.032 (0.030)	-0.013 (0.021)	-0.025 (0.024)
Inflação	-0.025* (0.014)	-0.028* (0.014)	-0.012 (0.011)	-0.016 (0.013)	-0.060 (0.043)	0.003 (0.019)	-0.005 (0.010)	-0.001 (0.010)
Abertura Comercial	-0.002 (0.035)	0.012 (0.035)	0.049** (0.023)	0.050** (0.024)	-0.006 (0.021)	0.062 (0.061)	-0.002 (0.032)	0.020 (0.031)
Capital Humano	0.046 (0.093)	-0.031 (0.098)	0.169 (0.123)	0.107 (0.119)	0.033 (0.069)	0.119 (0.171)	0.019 (0.103)	0.059 (0.115)
Gasto do Governo	-0.020 (0.033)	-0.011 (0.033)	0.001 (0.036)	0.010 (0.038)	0.198 (0.192)	0.005 (0.049)	0.012 (0.031)	0.029 (0.029)
GADP	0.140* (0.081)	0.139* (0.081)						
inst_1			-0.008 (0.061)	0.015 (0.046)				
inst_2					-0.060 (0.043)	-0.039 (0.033)		
Minst							0.022 (0.035)	0.010 (0.021)
T (média)	7.6	7.6	7.54	7.54	2.99	2.99	6.51	6.51
N	68	68	59	59	99	99	99	99
TN	517	517	445	445	296	296	644	644
Instrumentos	66	67	66	67	52	57	71	76
$H_0: \omega + \theta = 1$	0.6709	-	0.801	-	0.9623	-	0.7834	-

Todas as variáveis estão defasadas (exceto variáveis constantes no tempo) e em logaritmos (exceto dummies).

Dummies de tempo foram omitidas.

(1), (3), (5), (7) = modelos sem interação entre fator comum e dummy OCDE;

(2), (4), (6), (8) = modelos com interação entre fator comum e dummy OCDE.

Fonte: cálculo dos autores.

4.4.2. Proxies para a PTF mundial

Qualitativamente não há muita diferença entre usar o fator comum estimado via modelo de fatores dinâmicos, a PTF dos Estados Unidos ou a PTF média dos países da OCDE. As variáveis que eram significativas antes continuam (exceto no modelo (6)) e

em magnitudes similares. Entretanto, a magnitude do coeficiente θ muda de acordo com a variável escolhida, de modo que o fator estimado apresenta menos coeficientes e a PTF média da OECD apresenta os maiores coeficientes, os coeficientes da PTF americana ficam em um meio termo entre as duas. Os γ s implicados pelos fatores dos modelos (1), (3) e (5) são, respectivamente, 0.56%, 0.72% e 0.82%. Entretanto, os γ s implicados pelo coeficiente autorregressivo da PTF são todos em direção a um $\gamma = 0.71\%$. Como os testes de Wald não nos permitem rejeitar que $\theta + \omega = 1$, não há base para dizer que há alguma diferença significativa.

Tabela 13 - GMM Sistêmico em um estágio para diferentes *proxies* para PTF mundial.

SGMM em um estágio Variáveis	Fator Comum (estimado)		PTF (USA)		PTF (Média OCDE)	
	(1)	(2)	(3)	(4)	(5)	(6)
PTF	0.921*** (0.039)	0.941*** (0.038)	0.921*** (0.039)	0.940*** (0.038)	0.921*** (0.039)	0.933*** (0.038)
OCDE	-0.014 (0.062)	5.242** (2.059)	-0.014 (0.062)	2.823** (1.172)	-0.014 (0.062)	7.687*** (1.963)
$y_{i,0}$	-0.006 (0.038)	-0.018 (0.035)	-0.006 (0.038)	-0.017 (0.035)	-0.006 (0.038)	-0.012 (0.035)
GADP	0.159* (0.083)	0.140* (0.081)	0.159* (0.083)	0.140* (0.081)	0.159* (0.083)	0.163** (0.079)
Inflação	-0.022* (0.013)	-0.024* (0.013)	-0.022* (0.013)	-0.025* (0.013)	-0.022* (0.013)	-0.01858 (0.012)
Abertura Comercial	0.004 (0.036)	0.020 (0.036)	0.004 (0.036)	0.020 (0.036)	0.004 (0.036)	0.009 (0.034)
Capital Humano	0.035 (0.112)	-0.033 (0.116)	0.035 (0.112)	-0.031 (0.116)	0.035 (0.112)	-0.031 (0.111)
Gasto do Governo	-0.026 (0.032)	-0.016 (0.031)	-0.026 (0.032)	-0.016 (0.031)	-0.026 (0.032)	-0.018 (0.031)
Fator Comum	0.062*** (0.013)					
Fator Comum se OECD = 0		0.067*** (0.014)				
Fator Comum se OECD = 1		-0.370** (0.168)				
PTF (USA)			0.080*** (0.017)			
PTF (USA) se OECD = 0				0.085*** (0.018)		
PTF(USA) se OCDE =1				-0.218* (0.122)		
PTF (OCDE)					0.090*** 0.018	
PTF (OCDE) se OCDE = 0						0.0897*** (0.019)
PTF (OCDE) se OCDE = 1						-0.762*** (0.210)
T (média)	7.6	7.6	7.6	7.6	7.6	7.6
N	68	68	68	68	68	68
TN	517	517	517	517	517	517
Instrumentos	66	67	66	67	66	67
$H_0: \omega + \theta = 1$	0.667	-	0.975	-	0.764	-

Todas as variáveis estão defasadas (exceto variáveis constantes no tempo) e em logaritmos (exceto

SGMM em um estágio	Fator Comum (estimado)	PTF (USA)	PTF (Média OCDE)
-----------------------	---------------------------	--------------	---------------------

dummies).

Dummies de tempo foram omitidas.

(1), (3), (5) = modelos sem interação entre fator comum e dummy OCDE;

(2), (4), (6) = modelos com interação entre fator comum e dummy OCDE.

Erros-padrão em parêntesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Fonte: cálculo dos autores.

Também podemos ver a razoabilidade do modelo comparando o coeficiente autorregressivo, ω , estimado pelo SGMM com as estimativas via Mínimos Quadrados Ordinários e Efeitos Fixos, uma vez que o primeiro tende a superestimar ω e o último subestima ω (ROODMAN, 2009; BOND, HOEFFLER e TEMPLE, 2001). A tabela com os coeficientes se encontra abaixo, em todos os casos os coeficientes estimados ficam dentro dos limites.

Tabela 14 - Limites 'razoáveis' para o coeficiente autorregressivo do GMM Sistemico (limites dados pelas estimativas do MQO e Efeitos Fixos (FE))

	Fator Comum (estimado)		PTF (USA)		PTF (Média OCDE)	
	(1)	(2)	(3)	(4)	(5)	(6)
MQO	0.954	0.965	0.954	0.965	0.954	0.961
SGMM	0.921	0.941	0.921	0.940	0.921	0.933
Efeitos Fixos	0.791	0.814	0.791	0.813	0.791	0.804

Fonte: cálculo dos autores.

Por fim, na tabela se encontram os mesmos modelos, porém estimados através do SGMM em dois estágios e com a correção de Windmeijer (2005). As estimativas não continuam praticamente iguais ao do SGMM em um estágio e os resultados discutidos se mantêm.

Tabela 15 - GMM Sistemico em um estágio para diferentes *proxies* para PTF mundial.

SGMM em dois estágios	Fator Comum (estimado)		PTF (USA)		PTF (Média OCDE)	
Variáveis	(1)	(2)	(3)	(4)	(5)	(6)
PTF	0.923*** (0.036)	0.942*** (0.035)	0.923*** (0.036)	0.942*** (0.035)	0.923*** (0.036)	0.940*** (0.034)
OCDE	-0.013 (0.060)	4.996** (1.918)	-0.013 (0.060)	2.688** (1.051)	-0.013 (0.060)	7.015*** (1.998)
$y_{i,0}$	-0.005 (0.039)	-0.015 (0.033)	-0.005 (0.039)	-0.015 (0.033)	-0.005 (0.039)	-0.018 (0.033)
GADP	0.140* (0.081)	0.139* (0.081)	0.140* (0.081)	0.137* (0.080)	0.140* (0.081)	0.148* (0.078)
Inflação	*-0.025 (0.014)	*-0.028 (0.014)	*-0.025 (0.014)	*-0.028 (0.014)	*-0.025 (0.014)	-0.022 (0.014)
Abertura Comercial	-0.002 (0.035)	0.012 (0.035)	-0.002 (0.035)	0.012 (0.034)	-0.002 (0.035)	0.006 (0.032)
Capital	0.046 (0.093)	-0.031 (0.098)	0.046 (0.093)	-0.028 (0.098)	0.046 (0.093)	-0.011 (0.089)
Humano	-0.020 (0.033)	-0.011 (0.033)	-0.020 (0.033)	-0.011 (0.033)	-0.020 (0.033)	-0.015 (0.031)
Gasto do Governo	0.061*** (0.013)					
Fator Comum		0.066***				

SGMM em dois estágios	Fator Comum (estimado)		PTF (USA)		PTF (Média OCDE)	
se OECD = 0		(0.012)				
Fator Comum		-0.350**				
se OECD = 1		(0.160)				
PTF (USA)			0.078***			
			(0.016)			
PTF (USA)				0.081***		
se OECD = 0				(0.016)		
PTF(USA)				-0.207*		
se OCDE =1				(0.114)		
PTF (OCDE)					0.082***	
					(0.017)	
PTF (OCDE)						0.083***
se OCDE = 0						(0.017)
PTF (OCDE)						-0.694***
se OCDE = 1						(0.221)
T (média)	7.6	7.6	7.6	7.6	7.6	7.6
N	68	68	68	68	68	68
TN	517	517	517	517	517	517
Instrumentos	66	67	66	67	66	67
$H_0: \omega + \theta = 1$	0.671	-	0.966	-	0.900	-

Todas as variáveis estão defasadas (exceto variáveis constantes no tempo) e em logaritmos (exceto dummies).

Dummies de tempo foram omitidas.

(1), (3), (5) = modelos sem interação entre fator comum e dummy OCDE;

(2), (4), (6) = modelos com interação entre fator comum e dummy OCDE.

Erros-padrão em parêntesis; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Fonte: cálculo dos autores.

5. CONSIDERAÇÕES FINAIS

Nós investigamos os fatores que afetam a produtividade total dos fatores. Para isso, primeiro estimamos um modelo de fatores dinâmico para as PTFs dos 10 países mais desenvolvidos tecnologicamente, utilizando o índice de eficiência técnica de Pires e Garcia (2011). Desse modelo conseguimos extrair um componente comum e outro idiossincrático. Uma primeira coisa que notamos das estimativas é que há certa heterogeneidade nas séries de produtividade, mesmo todos os países possuindo, em tese, elevado grau de eficiência técnica. Apesar disso, é possível observar alguns comportamentos comuns, tal como a queda de produtividade generalizada em todos os países durante a crise de 2008-2009. Também é possível notar certa sincronia entre as taxas de crescimento desses países.

A presença desses choques “comuns” e a ênfase dada na literatura de crescimento econômico à presença de *spillovers* tecnológicos, modelado a partir da suposição de uma PTF mundial comum a todos os países, nos motivou a possibilidade de decompor a produtividade em um fator comum e outro idiossincrático, com o objetivo de entender melhor como as produtividades se relacionam com essa tecnologia mundial.

Os resultados do modelo de fatores dinâmicos nos dão pistas sobre como cada PTF se relaciona com o fator comum e nos serve de base para entender melhor como e quão importante são os *links* existentes entre as PTFs. Disso, uma primeira coisa a se notar é que a sensibilidade de cada país em relação ao fator comum, capturado pelos coeficientes de *loading*, parece ser inversamente proporcional ao nível de desenvolvimento tecnológico do país, como mostrado pela correlação negativa entre os *loadings* e diferentes variáveis relacionadas à ciência e tecnologia. Também se nota que alguns desses países são guiados principalmente pelo fator específico. A interpretação econômica dessa relação é intuitiva: os países mais avançados tecnologicamente estão mais próximos da fronteira tecnológica, portanto sua capacidade de absorver a tecnologia da fronteira é menor do que a de um país mais distante. Isto é, eles dependem menos da tecnologia mundial (PTF mundial), que, por sua vez, se reflete numa menor dependência em relação à parte comum das PTFs. Se for possível descrever essa relação a partir da existência de uma tecnologia mundial, então seria esperado que o componente comum da PTF contivesse essa tecnologia mundial ou estivesse relacionado com a mesma, refletindo o comportamento.

Para tentar validar esses argumentos, inserimos nossa estimativa do fator comum em um painel dinâmico com 144 países entre 1970 e 2011, estimados por meio do GMM Sistêmico (SGMM). Os resultados apontam para similaridades entre nosso fator e outras *proxies* para a PTF mundial, tal como a PTF dos Estados Unidos e a média das PTFs da OCDE. Não só isso, mesmo notando certa sensibilidade das estimativas de outros determinantes da PTF em relação à inserção de diferentes variáveis institucionais, a relação entre nossas *proxies* para a PTF mundial se mostraram robustos. Em todos os casos analisados podemos notar que, ao não rejeitar a hipótese de que $\theta + \omega = 1$, temos algum suporte para dizer que a relação entre PTF e tecnologia mundial se dá em termos relativos, por meio do *gap* da PTF mundial com o nível de PTF do país, como a teoria supõe. Também observamos que os países que não são da OCDE são sistematicamente mais sensíveis à PTF mundial do que os países da OCDE, reforçando as evidências do efeito de *catch up* tecnológico. Seguindo a formatação teórica de Aiyar e Feyrer (2002), nossas estimativas implicam uma taxa de convergência condicional entre 0.56 e 1.13% entre as PTFs dos países, relativamente menor do que encontrado pelos autores (3%). Apesar disso, os resultados para a influência do capital humano na produtividade total dos fatores são similares, em que nós encontramos uma elasticidade de aproximadamente 4,6 entre a PTF potencial e o capital humano, em concordância com o encontrado por Aiyar e Feyrer (2002), que encontram uma elasticidade de aproximadamente 4,5.

Em geral os demais determinantes da PTF não foram significativos, em especial dois de nossos indicadores de instituições. Talvez haja alguma má especificação ou então o SGMM não foi capaz de remover a endogeneidade presente nessas variáveis de instituições, o que pode ser o caso, uma vez que elas apenas apresentam pouca ou nenhuma variação no tempo, inviabilizando que utilizemos suas defasagens como instrumentos. Apesar dessas restrições, o indicador GADP para instituições foi significativo nos modelos e positivo, indicado que países com boas instituições tendem a ter PTFs maiores. Abertura comercial e capital humano também parecem estar relacionadas com maiores níveis de PTF, enquanto a altos níveis de inflação são deletérios ao acúmulo de PTF. Em nosso modelo não pudemos dizer nada sobre a relação entre a PTF e os gastos do governo, visto que esta variável não foi significativa em nenhuma de nossas especificações.

Por fim, é importante enfatizar as limitações deste trabalho. Uma primeira dificuldade é o escopo: nosso modelo de fatores dinâmicos utiliza apenas 10 países.

Para dar maior generalidade aos resultados obtidos aqui, pode ser interessante aplicar alguma modelagem de análise fatorial dinâmica para um grupo maior de países. Além disso, nosso SGMM, em diferentes especificações, não obteve muito êxito em identificar outros determinantes da PTF, o que pode indicar a necessidade de incluir mais determinantes ou novos instrumentos, não se resumindo apenas à utilizar defasagens de variáveis como instrumentos. Finalmente, também seria interessante verificar a robustez desses resultados a partir de formas de mensuração da PTF mais flexíveis. A construção assumindo uma função de produção do tipo Cobb-Douglas, apesar de comum, tem sido alvo de críticas por vários motivos, como, por exemplo, o fato de que essa metodologia assume que as unidades de produção são eficientes e ausência de efeitos cruzados entre os fatores (AMADOR E COIMBRA, 2007; DANQUAH et. al. 2011).

6. REFERÊNCIAS

- [1]. ACEMOGLU, D. **Introduction to Modern Economic Growth**. Cambridge: The MIT Press, 2009.
- [2]. ACEMOGLU, D.; ANTRAS, P.; HELPMAN, E. Contracts and Technology Adoption. **American Economic Review**, v. 97, p. 916-943, 2007.
- [3]. ACEMOGLU, D.; ROBINSON, J. Political Losers as a Barrier to Economic Development. **American Economic Review**, v. 90, p. 126-130, 2000.
- [4]. ACEMOGLU, D.; ZILIBOTTI, F. Productivity Differences. **Quarterly Journal of Economics**, v. 116, p. 563-606, 2001.
- [5]. AIYAR, SHEKHAR S.; FEYRER, J. A contribution to the empirics of total factor productivity, **Dartmouth College Working Paper**, august, 2002.
- [6]. AGHION, P.; HOWITT, P. A Model of Growth Through Creative Destruction. **Econometrica**, v. 60, p.323-52. 1990.
- [7]. AMADOR, J.; COIMBRA, C. Total Factor Productivity Growth in the G7 Countries: different or alike? **No. w200709**. 2007.
- [8]. ARELLANO, M; BOND, S. Some tests specifications for panel data: Monte Carlo evidence and an application to employment equations. **Review of Economic Studies**. v. 58, p. 277-97, 1991.
- [9]. ARELLANO, M; BOVER, O. Another look at the instrumental variables estimation of error-components models. **Journal of Econometrics**, v. 68, p. 29-51, 1995.
- [10]. BAI, J.; S. NG, Determining the Number of Factors in Approximate Factor Models, **Econometrica**, v.70, p. 191-221, 2002.
- [11]. BAI, J., AND S. NG, Determining the Number of Primitive Shocks in Factor Models. **Journal of Business and Economic Statistics**, v. 25, p. 52-60, 2007.
- [12]. BANERJEE, Anindy; MARCELLINO, M.; MASTEN, Igor. Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. **Emerald Group Publishing Limited**, v.3, 2008.
- [13]. BARRO, R. J.; LEE, J. W. A new data set of educational attainment in the world, 1950–2010. **Journal of development economics**, v. 104, p. 184-198, 2013.
- [14]. BASU, S., WEIL, D. Appropriate Technology and Growth. **Quarterly Journal of Economics**, v 113, p. 1025–1054, 1998.

- [15]. BLUNDELL, R; BOND, S. Initial conditions and moment restrictions in dynamic panel data models. **Journal of Econometrics**, v. 87, v. 11-143, 1998.
- [16]. BOND, S. R.; HOEFLER, A.; TEMPLE, J. RW. GMM estimation of empirical growth models. **Discussion Paper No. 2048**, Centre for Economic Policy Research. 2001.
- [17]. BOTERO, C.; DJANKOV, S.; LA PORTA, R.; LOPEZ-DE-SILIANES, F.; SHLEIFER, A. The Regulation of Labour. **The Quarterly Journal of Economics**, Vol. 119, pp. 1339-1382. 2004.
- [18]. COHEN, D.; LEKER, L. Health and Education: Another Look with the Proper Data. **mimeo** Paris School of Economics, 2014.
- [19]. DANQUAH, MICHAEL; MORAL-BENITO, ENRIQUE; OUATTARA, BAZOUMANA. TFP Growth and its Determinants: Nonparametrics and Model Averaging. Banco de Espana **Working Paper No. 1104**. April 5, 2011.
- [20]. DASGUPTA, SANJOY. Spectral Decomposition [PDF document]. **Lecture**. 2008. Retrieved from Lecture Notes Online Web site: <http://cseweb.ucsd.edu/~dasgupta/291/lec7.pdf>. Acesso em: 10/06/2015.
- [21]. DAMSGAARD, e. F.; KRUSELL, P. The World Distribution of Productivity: Country TFP Choice in a Nelson-Phelps Economy. **No. w16375**. National Bureau of Economic Research, 2010.
- [22]. DJANKOV, S.; LA PORTA, R.; LOPEZ-DE-SILIANES, F.; SHLEIFER, A. The Regulation of Entry. **The Quarterly Journal of Economics**, Vol. 119, pp.1339-1382. 2004.
- [23]. EASTERLY, W.; LEVINE, R. It's Not Factor Accumulation: Stylized Facts and Growth Models. **The World Bank Economic Review**, Vol. 15, pp.177-219. 2001.
- [24]. FEENSTRA, R.C.; INKLAAR, ROBERT; TIMMER, MARCEL. The Next Generation of The Penn World Table. April 2013. Disponível em: www.ggd.net/pwt. acesso em: 30/07/2013.
- [25]. FERNÁNDEZ-MACHO, FRANCISCO JAVIER. A dynamic factor model for economic time series. **Kybernetika**,v. 33, no. 6, p. 583-606, 1997.
- [26]. FERREIRA, PEDRO CAVALCANTI; ELLERY JR., R. Crescimento econômico, retornos crescentes e concorrência monopolista. **Revista de Economia Política**, v. 16, no. 2, p. 86-104, 1996.

- [27]. FERREIRA, PEDRO CAVALCANTI; PESSÔA, S. de A; VELOSO, FERNANDO A. On The Evolution Of Total Factor Productivity In Latin America. **Economic Inquiry**, v. 51, p. 16-30, 2013.
- [28]. GEWEKE, J. The Dynamic Factor Analysis of Economic Time Series. In: AIGNER, D.J.; GOLDBERG, A.S. **Latent Variables in Socio-Economic Models**. Amsterdam: North-Holland, 1977.
- [29]. GIBBONS, S.; OVERMAN, H.; PELKONEN, P. The decomposition of variance into individual and group components with an application to area disparities. **Mimeo**. London, LSE, 2012.
- [30]. GLAESER, L. E; LA PORTA, R.; LOPEZ-DE-SILANES, F.; SHLEIFER, A. Do Institutions Cause Growth? **The Journal of Economic Growth**, vol. 9, p. 271-303.
- [31]. GONÇALVES, C. E. S. Desenvolvimento Econômico: Uma breve incursão teórica. In: FERREIRA, PEDRO CAVALCANTI (Org.); VELOSO, Fernando A (Org.); PESSÔA, Samuel de Abreu (Org.); GIAMBIAGI, F. (Org.). **Desenvolvimento Econômico: Uma Perspectiva Brasileira**. 1. ed. Rio de Janeiro: Campus-Elsevier, 2013.
- [32]. GROSSMAN, G. & E. HELPMAN. Quality ladders in the theory of growth. **Review of Economics Studies**, v. 58, pp.43-61, 1991.
- [33]. HALL, Robert E.; JONES, Charles I. Why do some countries produce so much more output per worker than others? **No. w6564**. National bureau of economic research, 1999.
- [34]. HANSEN, B. E. **Econometrics**. 2016. Disponível em: <http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>. Acesso em 10/06/2016.
- [35]. HANSEN, L. Large sample properties of generalized method of moments estimators. **Econometrica**, v.50(3), p. 1029-54, 1982.
- [36]. HAMILTON, J. D. **Time series analysis**. Princeton: Princeton University Press, 1994.
- [37]. HARVEY, ANDREW C. **Forecasting, structural series models and the Kalman Filter**. Cambridge University Press, 1989.
- [38]. HSIEH, CHANG-TAI; KLENOW, PETER J. Development Accounting. **American Economic Journal: Macroeconomics**, v. 2, p. 1 207-223, 2010.
- [39]. HOLTZ-EAKIN, D; NEWEY, W; ROSEN, H.S. Estimating vector autoregressions with panel data. **Econometrica**, v. 56, p. 1371-95, 1988.

- [40]. INKLAAR, R.; TIMMER, M. P. Capital, labor and tfp in pwt 8.0. 2013. Disponível em: www.ggd.net/pwt. Acesso em 10/06/2015.
- [41]. ISLAM, N. Growth empirics: A Panel Data Approach. **Quarterly Journal of Economics**, v. 110, p. 1127-1170, 1995.
- [42]. JONES, Charles I. ROMR, Paul M. The new Kaldor facts: ideas, institutions, population, and human capital. **No. w15094**. National Bureau of Economic Research, 2009.
- [43]. LOKO, B.; DIOUF, MA. Revisiting the Determinants of Productivity Growth: What's New? **IMF Working Papers**. 2009.
- [44]. KALDOR, N. Capital Accumulation and Economic Growth. In: LUTZ, F.A.; HAGUE, D.C., eds., **The Theory of Capital**, St. Martins Press, pp. 177–222, 1961.
- [45]. KARABARBOUNIS, L; NEIMAN, B. The Global Decline of the Labor Share. **NBER. Working Paper, no.1913**. 2013.
- [46]. KAUFMANN, D.; AART, K.; MASTRUZZI, M. The Worldwide Governance Indicators: Methodology and Analytical Issues. **World Bank Policy Research Working Paper No. 5430**. 2010.
- [47]. KOSE, M. A.; Otrok, C.; Prasad, E. GLOBAL BUSINESS CYCLES: CONVERGENCE OR DECOUPLING? **IZA DP No. 3442. Discussion Paper No. 3442**. April, 2008.
- [48]. KLENOW, P.; RODRIGUEZ-CLARE, A. The neoclassical revival in growth economics: Has it gone too far? In: **NBER Macroeconomics Annual 1997**, Volume 12, pp. 73-114. MIT Press, 1997.
- [49]. KRUEGER, D. Quantitative Macroeconomics: an introduction. 2007. **Notas de Aula. Mimeo**. Disponível em: <http://down.cenet.org.cn/upfile/8/200851785232169.pdf>. Acesso em 10/06/2015.
- [50]. MANKIW, N.G; ROMER, D; WEIL, D.; A Contribution to the Empirics of Economic Growth; **Quarterly Journal of Economics**, v. 107, p. 407-437, 1992.
- [51]. MCGUINNESS, A. Institutions and total factor productivity convergence. **No. 9/RT/07. Central Bank of Ireland**, 2007.
- [52]. MINGOTI, SUELI APARECIDA. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. 1º reimpressão. Editora UFMG, 2007.

- [53]. MORALES-ARIAS, Leonardo; MOURA, Guilherme V. A conditionally Heteroskedastic Global Inflation Model. **Journal of Economic Studies**, v. 40, no. 4, p. 572-596, 2013.
- [54]. MORALES-ARIAS, Leonardo; MOURA, Guilherme V. A Conditionally Heteroskedastic Global Inflation Model. **Kiel Working Papers 1666**, Kiel Institute for the World Economy. 2010.
- [55]. MORIER, B.; TELES, V. K. A time-varying markov-switching model for economic growth. **Textos para discussão 305**, Escola de Economia de São Paulo, Getúlio Vargas Foundation (Brazil). 2011.
- [56]. NELSON, RICHARD R.; PHELPS, EDMUND S. Investment in humans, technological diffusion, and economic growth. **The American Economic Review**, p. 69-75, 1966.
- [57]. NORTH, D. **Institutions, Institutional Change and Economic Performance**. Cambridge: Cambridge University Press, 1990.
- [58]. PARENTE, S.; PRESCOTT, E. **Barriers to Riches**. Cambridge: The MIT Press, 2000.
- [59]. PIRES, JORGE OLIVEIRA; GARCIA, FERNANDO. Productivity of nations: a stochastic frontier approach to TFP decomposition. **Economics Research International**, 2012.
- [60]. PRESCOTT, EDWARD C. Lawrence R. Klein lecture 1997: Needed: A theory of total factor productivity. **International economic review**, p. 525-551, 1998.
- [61]. ROMER, P. Increasing Returns and Long-Run Growth. **Journal of Political Economy**, v. 94, p. 1002-1037, 1986.
- [62]. ROMER, P. Endogenous Technological Change. **Journal of Political Economy**, v. 98, p. 71-102, 1990.
- [63]. ROODMAN, D. How to Do xtabond2: An Introduction to “Difference” and “System” GMM in Stata. **Working Paper Number 103**. December 2006.
- [64]. SARGENT, T.J.; SIMS, C.A. Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory. In: SIMS, C. (org.). **New Methods in Business Cycle Research**. Minneapolis: Federal Reserve Bank of Minneapolis, 1977.
- [65]. SOLOW, R. A Contribution to the Theory of Economic Growth. **Quarterly Journal of Economics**, v. 70, p. 65-94, 1956.

- [66]. SOLOW, R. Technical Change and Aggregate Production Function. **The Review of Economics and Statistics**, Vol. 39, p. 312-320, 1957.
- [67]. SWAN, TREVOR W. Economic growth and capital accumulation. **Economic record** 32, no. 2, p. 334-361, 1956.
- [68]. STOCK, JAMES H., AND MARK W. WATSON. **Dynamic factor models. Oxford Handbook of Economic Forecasting**. Oxford University Press, USA. 2011: 35-59.
- [69]. VELOSO, F.; FERREIRA, P. C.; PESSÔA, S. Experiências Comparadas de Crescimento Econômico no Pós-guerra. In: FERREIRA, PEDRO CAVALCANTI (Org.); VELOSO, Fernando A (Org.); PESSÔA, Samuel de Abreu (Org.); GIAMBIAGI, F. (Org.). **Desenvolvimento Econômico: Uma Perspectiva Brasileira**. 1. ed. Rio de Janeiro: Campus-Elsevier, 2013.
- [70]. VILLELA, A. O desenvolvimento econômico em perspectiva histórica. In: FERREIRA, PEDRO CAVALCANTI (Org.); VELOSO, Fernando A (Org.); PESSÔA, Samuel de Abreu (Org.); GIAMBIAGI, F. (Org.). **Desenvolvimento Econômico: Uma Perspectiva Brasileira**. 1. ed. Rio de Janeiro: Campus-Elsevier, 2013.
- [71]. WINDMEIJER, F. A finite sample correction for variance of linear efficient two-step GMM estimators. **Journal of Econometrics**, v. 126, p. 25-51, 2005.
- [72]. WOLCOTT, S. The Perils of Lifetime Employment Systems: Productivity Advance in the Indian and Japanese Textile Industries, 1920-1938. **Journal of Economic History**, v. 54, p. 307-324, 1994.
- [73]. WOOLDRIDGE, J. M. **Econometric Analysis of Cross Section and Panel Data, second edition**. Cambridge, MA: MIT Press, 2010.

7. APÊNDICE A

A1. Análise Fatorial

O objetivo da análise fatorial é descrever a variabilidade de um vetor aleatório X em termos de um número menor de fatores comuns, ou seja, cada variável pertencente ao vetor X seria explicada por um conjunto de fatores comuns e seu respectivo termo de erro aleatório. Desse modo, a análise fatorial permite que se construa um número menor de variáveis não correlacionadas entre si e que contém boa parte da informação contida no vetor original.

A análise fatorial se subdivide em duas vertentes no que concerne a construção dos fatores, que são a análise fatorial exploratória e confirmatória. Na análise fatorial exploratória, parte-se do pressuposto que o número de fatores é desconhecido, assim seu objetivo é também descobrir o número de fatores a partir da amostra. Já a análise fatorial confirmatória supõe que o número de fatores é conhecido e verifica se esse modelo é condizente com os dados disponíveis. Nessa seção é feita uma apresentação da análise fatorial exploratória, tomando como referência a exposição de Mingoti (2007).

Em equações, o modelo de fatores pode ser representado em sua forma normalizada:

$$Z_{px1} = D_{pxp}(X_{px1} - \mu_{px1}) = L_{pxm}F_{mx1} + \varepsilon_{px1} \quad (\text{A.1})$$

Onde X_{px1} é um vetor aleatório, $\mu_{px1} = E(X)$, Z_{px1} é o vetor das variáveis padronizadas de X , $D_{pxp} = \text{diag}(\sigma_1^{-1}, \dots, \sigma_p^{-1})$, onde σ_i é o desvio padrão do i -ésimo elemento de X , F_{mx1} é um vetor com m fatores comuns e L_{pxm} é a matriz com os coeficientes, $l_{i,j}$, de cada fator em cada uma das p equações. Por fim, ε_{px1} é um vetor de erros aleatórios, que correspondem a parte não explicada pelos fatores comuns. Adicionalmente, considere que $\Sigma_{pxp} = \text{Var}(X)$, P_{pxp} a matriz de correlação de X e matriz de covariâncias de Z .

Desse modelo, vale notar que os fatores comuns da matriz F são também chamados de variáveis latentes, pois não são observáveis. Os coeficientes $l_{i,j}$ por sua vez são comumente chamados de *loadings*, e também precisam ser estimados de alguma forma.

Os pressupostos para que se obtenham estimativas consistentes dos parâmetros da equação (A.1) são:

- i) $E[F_{mx1}] = 0, \forall j = 1, \dots, m;$
- ii) $\text{Var}[F_{mx1}] = I_m;$
- iii) $E[\varepsilon_{px1}] = 0;$

- iv) $Var[\varepsilon_{px1}] = \Psi_{p \times p} = \text{diag}(\psi_1, \dots, \psi_p)$;
 v) ε e F são independentes, portanto $Cov(\varepsilon_{px1}, F_{mx1}) = E(\varepsilon F') = 0$.

Esses pressupostos implicam que cada um dos fatores comuns não tem correlação com o outro e nem com os erros. Sob os pressupostos (i-v) temos o chamado modelo de fatores ortogonal. Uma implicação desses pressupostos é que

$$Var(Z) = Var(LF + \varepsilon) = Var(LF) + Var(\varepsilon) + 2Cov(LF, \varepsilon)$$

Como $Cov(LF, \varepsilon) = 0$:

$$Var(Z) = Var(LF) + Var(\varepsilon) = L L' + \Psi$$

$$P_{p \times p} = L L' + \Psi \quad (\text{A.2})$$

A partir dessa relação devemos encontrar L e Ψ tais que representem a matriz de correlação de X para um valor m tal que $m \leq p$. De (A.2) temos, também, para cada elemento de Z , $Var(Z_i) = \sum_{j=1}^m l_{i,j}^2 + \psi_i = h_i^2 + \psi_i$. Ou seja, a variância de Z é decomposta em duas partes: a parte devida aos m fatores (h) é denominada de comunalidade, pois os fatores estão presentes em cada uma das observações de Z ; a segunda parte é simplesmente a variabilidade devida ao termo de erro.

Outra implicação dessa estrutura para a matriz de correlação é que $L = Cov(F, Z)$ e, portanto, a proporção da variância que é explicada pelo fator F_j é dada por $\sum_{i=1}^p l_{i,j}^2 / P$. Quanto maior esse valor, mais representativo é aquele fator no modelo.

Para que o modelo de fatores ortogonal exposto acima seja aplicado, é necessário (i) estimar o número de fatores; (ii) estimar as matrizes L e Ψ ; e (iii) construir os escores de F para cada elemento amostral.

A1.1. Estimação do número de fatores

A forma usual de se escolher o número de fatores é feita com base nos autovalores da matriz de correlação amostral, $R_{p \times p}$. Os autovalores, $\hat{\lambda}_i$, são então postos em ordem decrescente para identificar quais desses autovalores são os mais importantes. A ideia subjacente a análise dos autovalores para determinação do número de fatores é devido ao fato de que a matriz de correlação pode ser representada em termos de seus autovetores e autovalores pelo Teorema da Decomposição Espectral, definido como:

Teorema A.1 (Teorema da Decomposição Espectral):

Seja $M \in \mathbb{R}$, onde M é uma matriz simétrica $d \times d$ com autovalores $\lambda_1, \dots, \lambda_d$ e respectivos autovetores ortonormais u_1, \dots, u_d , então:

$$M = Q\Lambda Q'$$

onde $Q = (u_1, \dots, u_d)$ é um vetor ortogonal contendo os autovetores de Λ , e $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$. Disso, temos também que $M = \sum_{i=1}^d \lambda_i u_i u_i'$.

Prova (DASGUPTA, 2008, Lecture): Como (u_1, \dots, u_d) são linearmente independentes, temos que $\forall N_{d \times d}$, $Nu_i = Mu_i \Leftrightarrow N = M$. Assim, $Q\Lambda Q' = M$ se $Q\Lambda Q' u_i = Mu_i$. Portanto, basta demonstrar que a última relação é verdadeira. Isso pode ser visto da seguinte forma: $Q\Lambda Q' u_i = Q\Lambda e_i$ (pois Q é ortonormal) $\Rightarrow Q\Lambda e_i = Q\lambda_i e_i = \lambda_i Q e_i = \lambda_i u_i = Mu_i$. De maneira similar, $(\sum_{i=1}^d \lambda_i u_i u_i') u_j = \lambda_j u_j = Mu_j$.

Q.E.D.

Pelo teorema,

$$R_{p \times p} = \sum_{i=1}^p \hat{\lambda}_i \hat{u}_i \hat{u}_i' \quad (\text{A.3})$$

Se esses autovalores estiverem ordenados em termos de grandeza, uma aproximação de (A.3) pode ser dada para um $m \leq p$:

$$R_{p \times p} \simeq \sum_{i=1}^m \hat{\lambda}_i \hat{u}_i \hat{u}_i' \quad (\text{A.4})$$

Essa aproximação será tanto melhor quanto menores forem os autovalores excluídos da equação (A.3). Sendo assim, podemos escolher o número m de fatores comuns tomando como base algum critério de escolha para os autovalores mais importantes em termos de grandeza. Mingoti (2007) sugere pelo menos três critérios de escolha: (1) analisar a proporção da variância total relacionada com cada autovalor ($\hat{\lambda}_i/p$, $i = 1, \dots, p$) e escolhendo um número de fatores igual ao número de m autovalores com as maiores proporções; (2) escolher m fatores de acordo com o número de autovalores maiores ou iguais a 1; (3) observar o gráfico *scree-plot*, que consiste num gráfico que dispõe os valores de cada autovalor ordenados em ordem decrescente. O objetivo do gráfico é encontrar o “ponto de salto”, isto é, o ponto em que há algum decréscimo significativo na grandeza dos autovalores, e, portanto, de sua importância em relação à variância total, nesse critério m é igual ao número de autovalores anteriores a esse ponto.

Dessas informações é evidente que a melhor aproximação de R consideraria todos os autovalores, porém um dos objetivos da análise fatorial é especificamente explicar um grande número de variáveis com um número menor de fatores, e por isso é importante prezar pela parcimônia também na escolha do número m . Vale lembrar que também que a ideia subjacente a isso é que as variáveis originais são correlacionadas de alguma forma. Conforme Mingoti (2007), se esse não fosse o caso, cada fator estaria relacionado com apenas uma variável original, implicando que $m=p$.

Mingoti (2007) também apresenta alguns critérios que podem auxiliar na decisão do valor ótimo de m . É possível formular um teste de hipóteses na qual H_0 : m fatores é o suficiente; e H_a : é necessário mais fatores se Z e F tiverem distribuições normais multivariadas²⁸. Além disso, os critérios de informação de Akaike e critério Bayesiano de Schwarz podem ser aplicados na escolha de m . Esses critérios contrapõem o número m de fatores com o valor máximo da função de verossimilhança. Assim como no teste de hipóteses, esses critérios exigem distribuição normal.

A autora, entretanto, ressalta que para um n muito pequeno em relação à p , o teste de hipóteses tenderá a rejeitar a hipótese nula. Enquanto isso, ambos os critérios de informação tendem a superdimensionar m . Apesar disso, segundo ela, o método de Schwarz geralmente produz resultados melhores do que o critério de Akaike.

AI.2. Estimação das matrizes de loading e covariância do erro

Há três formas de estimar L e Ψ . Uma consiste em aplicar o método dos componentes principais (PCA) para construir L e Ψ . A segunda forma é chamada de método dos fatores principais, e consiste num refinamento do método anterior. A última forma é via máxima verossimilhança. O último método é mais preciso que os demais, porém sua aplicação depende do conhecimento a priori da distribuição de Z .

AI.2.1 Método dos componentes principais

A estimação de L e Ψ via componentes principais consiste em encontrar os respectivos autovetores, \hat{u}_i , para cada autovalor, $\hat{\lambda}_i$, onde $i = 1, \dots, m$. Sendo m o número de fatores definidos anteriormente. Pelo Teorema de Decomposição Espectral visto anteriormente, podemos reescrever R como:

$$R_{p \times p} = \sum_{i=1}^m \hat{\lambda}_i \hat{u}_i \hat{u}_i' + \sum_{j=m+1}^p \hat{\lambda}_j \hat{u}_j \hat{u}_j' \quad (\text{A.5})$$

Assim, $\hat{L}\hat{L}'$ pode ser aproximado por $\sum_{i=1}^m \hat{\lambda}_i \hat{u}_i \hat{u}_i'$ e $\hat{\Psi}$ por $\text{diag}(R - \hat{L}\hat{L}')$. Vale notar que Ψ deve ser diagonal, mas a matriz $R - \hat{L}\hat{L}'$ não necessariamente é diagonal. Sendo assim, há a necessidade de tomar apenas a diagonal dessa matriz para a estimação de $\hat{\Psi}$. Portanto, na posse dos autovetores e autovalores de Z , as matrizes L e Ψ serão dadas por:

$$\hat{L}_{p \times m} = [\hat{\lambda}_1^{1/2} \hat{e}_1 \quad \dots \quad \hat{\lambda}_m^{1/2} \hat{e}_m] \quad (\text{A.6})$$

$$\hat{\Psi}_{p \times p} = \text{diag}(R - \hat{L}\hat{L}') \quad (\text{A.7})$$

²⁸ A estatística do teste é dada como $T = \left[n - 1 - \frac{2p+4m+5}{6} \right] \ln \left(\frac{\det(\hat{L}\hat{L}' + \Psi)}{\det(R)} \right)$.

Como é usada apenas a diagonal de $R - \hat{L}\hat{L}'$ na estimação de $\hat{\Psi}$, o resultado é que a soma $\hat{L}\hat{L}' + \hat{\Psi}$ não será exatamente igual a R . Evidentemente, quanto menor a diferença entre $\hat{L}\hat{L}' + \hat{\Psi}$ e R , melhor é o ajuste do modelo fatorial aos dados. Isso nos permite construir uma medida de ajuste do modelo, dada por

$$MRES = R - (\hat{L}\hat{L}' + \hat{\Psi}) \quad (\text{A.8})$$

Onde MRES é a matriz residual. Idealmente se espera que essa matriz residual tenha seus valores próximos de zero. Essa forma de aproximação é aceita porque MRES só será igual a uma matriz nula se $m = p$. A informação na MRES pode ser sintetizada considerando a média dos valores acima da diagonal principal (EM), ou então pela raiz quadrada da média desses mesmos valores ao quadrado (REQM)²⁹. Além dessas medidas de ajuste, há também o critério de Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett para a matriz de correlação, que podem ser vistos em mais detalhes em Mingoti (2007).

A1.2.2. Método dos fatores principais

Esse método é mais um refinamento do método anterior do que um método per si. Isso porque ele parte de uma estimativa inicial, e então iterativamente melhora essas estimativas. Seu procedimento consiste do fato de que LL' é constituída a partir das comunalidades, h^2 , e as correlações de cada uma das variáveis, ρ . Assim, seria possível estimar LL' através de uma matriz R^* dada por:

$$R^* = \begin{bmatrix} h_1^{*2} & r_{1,2} & \dots & r_{1,p} \\ r_{2,1} & h_2^{*2} & \dots & r_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p,1} & r_{p,2} & \dots & h_p^{*2} \end{bmatrix} = L^* L^{*'} \quad (\text{A.9})$$

Onde r_{ij} são as correlações estimadas e h_i^{*2} as comunalidades iniciais. A partir dos valores iniciais, é então usado o método dos componentes principais visto anteriormente para construir $L^* = [\hat{\lambda}_1^{*1/2}\hat{u}_1 \quad \dots \quad \hat{\lambda}_m^{*1/2}\hat{u}_m]$, onde $\hat{\lambda}_i^*$ e \hat{u}_i são os autovalores e autovetores de R^* . Essas estimativas, porém, irão produzir um novo conjunto de comunalidades. Estas, por sua vez, podem ser usadas para construir outra matriz R e obter novas estimativas de L , que darão novas comunalidades. Esse processo pode ser repetido até que a diferença entre as comunalidades de cada nova iteração sucessiva seja desprezível. Entretanto, dependendo dos valores iniciais, essa estimação pode resultar

²⁹ Ou seja, $EM = N^{-1} \sum_{a=2}^{p-1} \sum_{i=a}^p \sum_{j=1}^m l_{i,j} l_{j,i+1}$ e $REQM = \sqrt{N^{-1} \sum_{a=2}^{p-1} \sum_{i=a}^p \sum_{j=1}^m (l_{i,j} l_{j,i+1})^2}$, onde N é o número de elementos acima da diagonal principal.

em inconsistências, tal como autovalores negativos ou mesmo uma comunalidade maior do que 1.

A1.2.3. Máxima Verossimilhança

Se for possível assumir que X segue uma distribuição normal p -variada, $X \sim N_p(\mu, \Sigma)$, então também temos que $Z \sim N_p(0, P)$. Sob os pressupostos do modelo de fatores ortogonal também temos que $(F_1, \dots, F_m) \sim N_m(0, I_m)$ e $(\varepsilon_1, \dots, \varepsilon_p) \sim N_p(0, \Psi)$.

Sendo assim, a função de verossimilhança será dada por:

$$\mathcal{L}(0, P) = (2\pi)^{-\frac{np}{2}} \det(LL' + \Psi)^{-\frac{n}{2}} \exp\{-0,5 \sum_{j=1}^n z_j'(LL' + \Psi)^{-1} z_j\} \quad (\text{A.10})$$

Onde n é o número de observações de $Z = (z_1, \dots, z_n)$. Essa função de máxima verossimilhança irá gerar os valores de \hat{L} e $\hat{\Psi}$ que maximizam (A.10). Assim como os demais métodos, esse também exige que se tenha um número m de fatores determinados a priori. Conforme Mingoti (2007), esse método dá estimativas mais precisas para L e Ψ do que o método dos componentes principais. Entretanto, além da pressuposição de normalidade ser necessária, as estimativas via (A.9) são sensíveis ao número de fatores escolhidos. Para diferentes valores de m , a maximização via verossimilhança irá produzir diferentes valores para dos *loadings*. Isso não ocorre para a estimação via PCA.

A1.2.4. Estimação dos escores dos fatores comuns

Uma vez encontrado as matrizes L e Ψ , é então possível obter as estimativas de cada um dos fatores comuns para cada elemento amostral. Esses elementos são definidos como uma média ponderada das p variáveis Z , assim, para cada elemento amostral k , tal que $k=1, \dots, n$:

$$\hat{F}_{j,k} = \sum_{i=1}^p w_{j,i} Z_{i,k} \quad (\text{A.11})$$

Onde $w_{i,j}$ são os pesos dados a cada uma das variáveis Z_i . Mingoti (2007) sugere três formas de se estimar esses pesos: via método dos mínimos quadrados ponderados, outros métodos de regressão e métodos *ad hoc* de estimação.

A primeira forma de estimação via mínimos quadrados ponderados (MQP) consiste em tratar o vetor F como sendo um vetor de parâmetros a serem estimados. Ao sugerir isso, temos que a equação (A.1) pode ser tratada como um modelo de regressão linear múltipla, onde Z é a variável resposta, L é o regressor e ε é o termo de erro. Tendo posse dos valores de L , é possível estimar aplicar as técnicas de regressão sobre essa

equação e obter estimativas de F . Vale notar que, apesar de ser possível estimar via mínimos quadrados ordinários (MQO), a estimação via MQP é mais recomendada, pois raramente Ψ será homocedástica. Sendo assim, se for possível satisfazer os pressupostos do método dos mínimos quadrados generalizados³⁰, então é possível obter estimativas consistes de \hat{F}_j a partir do seguinte estimador:

$$\hat{F}_{jk} = (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}Z_k = WZ_k \quad (\text{A.12})$$

Os pressupostos para o modelo são:

- i. $E(\hat{L}_i \otimes \varepsilon_i) = 0$;
- ii. $\Psi = Var(\varepsilon)$ é positiva definida e $E(\hat{L}_i'\Psi^{-1}\hat{L}_i)$ é não singular.

Além disso, vale notar que não temos a variância populacional dos resíduos, o que há disponível é uma estimativa do mesmo. Sendo assim, é necessário que

$$\text{plim}_{N \rightarrow \infty} \hat{\Psi} = \Psi \quad (\text{A.13})$$

Se os pressupostos do MQG e (A.13) for satisfeito, então a estimação é consistente e factível (e nesse caso, temos a estimação via mínimos quadrados generalizados factíveis (MQGF)). Adicionalmente, se tivermos que $E(\hat{L}_i'\hat{\Psi}^{-1}\varepsilon_i\varepsilon_i'\hat{\Psi}^{-1}\hat{L}_i) = E(\hat{L}_i'\hat{\Psi}^{-1}\hat{L}_i)$, então o estimador será também eficiente (WOOLDRIDGE, 2010). Um adendo que deve ser feito em relação a esse método é que a metodologia utilizada trata o vetor F como um vetor de parâmetros a serem estimados, quando na realidade os elementos de F são variáveis aleatórias, não parâmetros. Outra desvantagem do método é que as observações amostrais de Z são correlacionadas entre si. Apesar dessas desvantagens, o método tem como vantagem o fato de que não é necessário assumir nada sobre a distribuição de Z .

Outra coisa a se notar é que o estimador (A.12) irá atribuir um peso a cada variável Z no modelo. Mingoti (2007) cita que há uma alternativa para a construção dos escores é manter apenas as variáveis com os maiores pesos, descartando aquelas que estão pouco correlacionadas com o fator em questão. Outra forma de estimação via métodos de regressão é, no modelo de fator ortogonal, assumir que Z , F e ε tenham distribuição normal multivariada. Nesse caso, $(Z, F) \sim N_{m+p}(0, \Sigma^*)$, onde

$$\Sigma_{(p+m) \times (p+m)}^* = \begin{bmatrix} P & \vdots & L \\ \cdots & \cdots & \cdots \\ L' & \vdots & I \end{bmatrix}$$

³⁰ O MQP é simplesmente um caso especial do MQG para quando a matriz de ponderação é diagonal (isto é, há pode haver heterocedasticidade, mas as variâncias não estão correlacionadas entre si), sendo assim os pressupostos do MQG devem ser satisfeitos.

Sob esse pressuposto, $(F|Z = z) \sim N_m(L'P^{-1}z, I - L'P^{-1}L)$, e então, dado L e Ψ estimados via Máxima Verossimilhança, é possível estimar os fatores comuns para cada elemento amostral $k, k=1, \dots, n$, da seguinte maneira:

$$\hat{F}_k = \hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1}Z_k \quad (\text{A.14})$$

Onde $F_k = (F_{1k}, \dots, F_{mk})$.

Por fim, existem também alguns métodos *ad hoc* para estimação dos escores a partir dos *loadings* de cada elemento do vetor Z . Por exemplo, um desses métodos consiste em construir o escore k do fator F_j de tal forma que este seja o valor correspondente da variável Z_i com maior *loading*. Outro método consiste em tomar a média a média das variáveis Z mais correlacionadas com o fator em questão, porém considerando o sinal da correlação. Uma lista mais completa de outros métodos *ad hoc* e suas respectivas referências pode ser visto em Mingoti (2007).

A1.3. Tópicos adicionais

Uma coisa a se notar no modelo de fatores ortogonal é que a solução do modelo para a escolha de \hat{L} não é única. Isto porque o que é realmente estimado é a matriz $\hat{L}\hat{L}'$. Como tal, para qualquer matriz ortogonal é possível encontrar outra matriz de *loading*, \tilde{L} , tal que $\tilde{L}\tilde{L}' = \hat{L}\hat{L}'$ ³¹. Sendo assim, é possível realizar transformações no vetor \hat{L} sem que isso altere a orientação entre os fatores ou mude a qualidade do ajuste do mesmo. Portanto, caso a matriz de *loading* original seja de difícil interpretação, é possível aplicar uma transformação nessa matriz de maneira a produzir uma nova matriz de *loading* com interpretação mais fácil. Essa técnica se chama rotação ortogonal, que consiste na aplicação de uma matriz ortogonal T para realizar tal transformação. Alguns critérios de escolha para a matriz T , tal como o critério varimax, quartimax e orthomax podem ser vistos em Mingoti (2007). Além da rotação ortogonal, também é possível fazer rotações não-ortogonais. Estas consistem em transformações que não preservam a estrutura original do modelo, alterando o ajuste original.

³¹ Seja T uma matriz ortogonal, então $TT' = I$. Assim, $\hat{L}TT'\hat{L}' = \hat{L}\hat{L}'$, ou seja, tanto $\tilde{L} = \hat{L}T$ quanto \hat{L} são soluções.

A2. Modelos de Estado Espaço e Filtro de Kalman

Diversos modelos dinâmicos podem ser representados sob a forma de estado espaço (*state space* ou SS). Essa representação reescreve o modelo com dois conjuntos de equações: um conjunto de equações para descrever as variáveis de estado do modelo (equações a diferenças) e outro conjunto de equações com as variáveis endógenas observadas. Tal forma é bastante conveniente por permitir a aplicação do Filtro de Kalman para obter as estimativas dos parâmetros. Esse filtro consiste em um algoritmo que usa as séries observadas ao longo do tempo para produzir recursivamente estimativas de variáveis desconhecidas. A partir dele também é possível obter a função de verossimilhança (na forma de previsão de erro), que permite a estimação dos parâmetros do modelo via máxima verossimilhança. O Filtro de Kalman e suas variações são aplicáveis sob o pressuposto de normalidade dos erros, porém é possível também realizar a estimação via quasi-máxima verossimilhança (QML) quando não é possível supor isso. Essa forma de estimação será consistente e assintoticamente normal quando o modelo for estacionário.

Esta seção está baseada em Harvey (1989) e apresenta as propriedades da forma de estado espaço e a estimação utilizando o filtro de Kalman.

A2.1 Modelo de Estado Espaço

A representação geral do modelo de SS considera o seguinte sistema de equações:

$$y_t = Z_t \alpha_t + d_t + \epsilon_t, \quad t = 1, \dots, T \quad (\text{A.15})$$

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \quad t = 1, \dots, T \quad (\text{A.16})$$

A equação (A.15) considera uma série de tempo multivariada y_t com N elementos. Essa série está relacionada com um vetor $mx1$, α_t , conhecido como vetor estado. Essa equação é conhecida como equação de mensuração. Z_t é uma matriz Nxm , d_t é um vetor $Nx1$ e ϵ_t é um vetor $Nx1$ de erros serialmente não correlacionados. A equação (A.16), por sua vez, é chamada de equação de transição, e ela apresenta a dinâmica do vetor de estado através de um processo de Markov de primeira ordem. T_t é uma matriz mxm , c_t é um vetor $mx1$, R_t é uma matriz mxg e η_t é um vetor $gx1$ de erros serialmente não correlacionados³².

³² A equação de estado também pode ser representada como $\alpha_{t+1} = T_t \alpha_t + c_t + R_t \eta_t$. Na prática não há grandes diferenças em considerar essa estrutura ou a originalmente apresentada, porém isso pode ter implicações se o modelo em questão permitir correlação entre os resíduos da equação de transição e mensuração.

Nesse modelo há como pressuposto que y_t é observado, porém α_t é desconhecido. Além disso, os seguintes pressupostos são feitos sobre ϵ_t e η_t :

$$\begin{aligned} E(\epsilon_t) &= 0 \text{ e } Var(\epsilon_t) = H_t \\ E(\eta_t) &= 0 \text{ e } Var(\eta_t) = Q_t \end{aligned} \quad (\text{A.17})$$

Onde H_t e Q_t são matrizes de covariâncias e seus respectivos erros são serialmente não correlacionados. Além desse pressuposto sobre os erros, a especificação do modelo depende adicionalmente dos seguintes pressupostos:

$$E(\alpha_0) = a_0 \text{ e } Var(\alpha_0) = P_0 \quad (\text{A.18})$$

$$E(\epsilon_t \eta_t') = 0, \quad \forall s, t = 1, \dots, T \quad (\text{A.19})$$

$$E(\epsilon_t \alpha_0') = 0, E(\eta_t \alpha_0') = 0, \quad \forall t = 1, \dots, T \quad (\text{A.20})$$

Sob esses pressupostos temos que os erros das duas equações não são correlacionados entre si em qualquer período de tempo. Esses erros também não se correlacionam com o termo o valor inicial do vetor de estado. Harvey (1989) também define um conjunto chamado Sistema de Matrizes, que inclui os seguintes elementos:

$$SM = \{Z_t, d_t, H_t, T_t, c_t, R_t, Q_t\}$$

Por fim, um pressuposto comum feito sobre SM é que seus elementos são não-estocásticos, sendo assim o sistema dado pelo modelo de estado espaço é linear para qualquer período. A partir de SM também é possível definir uma subclasse de modelos chamados de invariantes no tempo, que consiste no modelo de estado espaço quando os elementos de SM são invariantes no tempo.

Uma coisa a se notar na representação de estado espaço é que para muitos modelos ela não é única. Isto é, vários modelos de séries de tempo podem ser representados no formato descrito acima de diferentes maneiras. A forma como o modelo original deve ser reescrito para caber no formato dado pelas equações de transição e mensuração é bastante flexível, a exigência é apenas que os pressupostos dessa representação não sejam violados. Apesar disso, Harvey (1989) recomenda que α seja construído de maneira a conter toda a informação relevante do sistema no tempo t e com o menor número de elementos possível.

Por fim, Harvey (1989) também define um vetor φ chamado vetor de hiperparâmetros. Esse vetor consiste no conjunto de parâmetros de SM exceto os parâmetros d_t e c_t . Geralmente os hiperparâmetros serão desconhecidos e precisam ser estimados de alguma forma.

A2.2. Filtro de Kalman

Como dito, a estimação dos parâmetros do modelo de estado espaço geralmente é via máxima verossimilhança, e o filtro de Kalman é um processo recursivo para obter o estimador ótimo do vetor de estado no tempo t (com base no conjunto de informações observáveis até t). Com isso, é então possível calcular a função de verossimilhança através da decomposição da previsão dos erros. Através desse procedimento é possível estimar e testar qualquer parâmetro desconhecido no modelo. O conjunto de informações do processo considera todas as observações de y até o tempo t . O sistema de matrizes, bem como a_0 e P_0 são considerados conhecidos em todos os períodos, e, portanto não é necessário que sejam explicitados no conjunto de informações.

A forma geral do Filtro de Kalman toma como base o modelo de espaço estado com base nas equações (A.15-16). Seja $E(\alpha_{t-1}|Y_{t-1} = y_0, \dots, y_{t-1}) = a_{t-1}$, a matriz de covariâncias do erro estimado em $t-1$ é dada por:

$$P_{t-1} = E[(\alpha_{t-1} - a_{t-1})(\alpha_{t-1} - a_{t-1})']$$

Com base apenas no conjunto de informações até $t-1$, teremos que o estimador ótimo de α_t e a matriz de covariâncias dos erros estimados em t serão respectivamente:

$$E(\alpha_t|Y_{t-1}) = a_{t|t-1} = T_t a_{t-1} + c_t \quad (\text{A.21})$$

$$P_{t|t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t' \quad (\text{A.22})$$

As equações (A.21) e (A.22) são as equações de previsão (*prediction equations*) da equação de estado. Além disso, quando a informação em t se torna disponível, podemos atualizar essas estimativas a partir das chamadas equações de atualização (*updating equations*), dadas por:

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t a_{t|t-1} - d_t) \quad (\text{A.23})$$

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \quad (\text{A.24})$$

$$F_t = Z_t P_{t|t-1} Z_t' + H_t \quad (\text{A.25})$$

O sistema (A.23-25) pode ser reescrito de forma a consistir em uma única recursão de $a_{t|t-1}$ até a_t , basta reescrever o sistema como:

$$a_{t+1|t} = (T_{t+1} - K_t Z_t) a_{t|t-1} + K_t y_t + c_{t+1} - K_t d_t \quad (\text{A.26})$$

Onde K é chamada de matriz de ganho, e é dada por $K_t = T_{t+1} P_{t|t-1} Z_t' F_t^{-1}$. A recursão para a matriz de covariância do erro é:

$$P_{t+1|t} = T_{t+1} (P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}) T_{t+1}' + R_{t+1} Q_{t+1} R_{t+1}' \quad (\text{A.27})$$

Essa recursão 'direta' é conhecida como equação de Riccati.

Juntas, as equações de previsão e atualização formam o filtro de Kalman. Uma vez especificados os valores iniciais de a_0 e P_0 , é então possível obter os estimadores ótimos do vetor estado em cada período no tempo, realizar previsões sobre o mesmo e atualizar o modelo conforme novas informações são adicionadas ao conjunto de informação, Y . Vale notar que o filtro permite realizar previsões ótimas sobre os valores do modelo como um todo, isso inclui tanto a equação de estado quanto a das observações. Apesar do filtro de Kalman ser factível, Harvey (1989) argumenta que nem sempre a recursão dada por ele é a mais adequada. Quando o número de séries no modelo é maior do que a dimensão do vetor de estado, o autor argumenta que o filtro de informação pode levar aos mesmos resultados, porém é mais facilmente computável. Para mais informações sobre, ver Harvey (1989).

A derivação das equações de previsão é bastante direta. Já a derivação das equações de atualização se baseia na suposição de normalidade. Sob o pressuposto de normalidade, $\alpha_0 \sim N_m(a_0, P_0)$. Sob os pressupostos usuais do modelo de estado espaço, também temos que os resíduos ϵ e η têm distribuição normal multivariada e são independentes entre si e de α_0 . Sendo assim, podemos fazer uso do seguinte lema:

Lema A.1: Considere que os vetores x e y seguem uma distribuição multivariada conjunta tal que

$$(x', y') \sim NM(\mu, \Sigma)$$

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \text{ e } \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

Onde $NM(\cdot)$ designa uma distribuição normal multivariada, μ é o vetor de médias e Σ é a matriz covariâncias. Se isso vale, então a distribuição de x condicionada em y , $f(x|y)$, também segue uma distribuição normal multivariada tal que:

$$f(x|y) \sim NM(\mu_{x|y}, \Sigma_{xx|y})$$

$$\mu_{x|y} = \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y)$$

$$\Sigma_{xx|y} = \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}$$

Note que $\Sigma_{xx|y}$ não depende de y e Σ e Σ_{yy} devem ser não-singulares. Ainda assim, o lema continua válido caso se possa obter uma pseudo-inversa de Σ_{yy}^{-1} .

A prova do Lema acima pode ser vista em Harvey (1989).

Se for possível assumir que o par $[\alpha_1' \ y_1']'$ segue uma distribuição normal multivariada, então

$$[\alpha_1' \ y_1']' \sim NM \left(\left[\alpha_{1|0}' \ (Z_1 \alpha_{1|0} + d_1)' \right], \begin{bmatrix} P_{1|0} & P_{1|0} Z_1' \\ Z_1 P_{1|0} & Z_1 P_{1|0} Z_1' + H_1 \end{bmatrix} \right)$$

Assim, podemos obter α_1 e P_1 simplesmente aplicando diretamente o Lema A1. Nesse caso,

$$\begin{aligned} \underbrace{\alpha_1}_{E(\alpha_1|y_1)} &= \underbrace{\alpha_{1|0}}_{E(\alpha_1)} + \underbrace{P_{1|0} Z_1'}_{Cov(\alpha_1, y_1)} \cdot \underbrace{F_1^{-1}}_{Var(y_1)^{-1}} \cdot \left(y_1 - \frac{(Z_1 \alpha_{1|0} - d_1)}{E(y_1)} \right) \\ \underbrace{P_1}_{Var(\alpha_1|y_1)} &= \underbrace{P_{1|0}}_{Var(\alpha_1)} - \underbrace{P_{1|0} Z_1'}_{Cov(\alpha_1, y_1)} \cdot \underbrace{F_1^{-1}}_{Var(y_1)^{-1}} \cdot \underbrace{Z_1 P_{1|0}}_{Cov(y_1, \alpha_1)} \end{aligned}$$

Esse resultado é exatamente a fórmula dada pelas equações (A.23) e (A.24), porém para o caso em que $t=1$. A partir dessas equações acima é possível obter recursivamente todos demais períodos³³.

Uma coisa a se notar no modelo apresentado acima é que ele trabalha com a suposição de que ϵ e η não são correlacionados em nenhum período. Essa hipótese pode ser relaxada para permitir a correlação entre ϵ e η contemporaneamente, porém neste caso as equações de atualização sofrem pequenas modificações.

Seja

$$E(\eta_t \epsilon_s') = \begin{cases} G_t, & t = s \\ 0, & t \neq s \end{cases}$$

Onde G_t é uma matriz $g \times N$. Nesse caso, teremos que:

$$Cov(\alpha_1, y_1) = P_{t|t-1} Z_t' + R_t G_t$$

$$Var(y_t) = Z_t P_{t|t-1} Z_t' + Z_t R_t G_t + G_t' R_t' Z_t' + H_t$$

Portanto essas modificações devem ser levadas em conta ao construir as equações de atualização. Isso, evidentemente, também irá alterar a equação de Riccati. Além disso, é necessário tomar cuidado com a especificação da equação de estado, pois os resultados podem não serão mais os mesmos quando ele é definido em termos de α_{t+1} . Harvey (1989) menciona, porém, que quando este for o caso, uma alternativa é transformar o sistema de maneira que os erros sejam não correlacionados.

³³ A média não condicional de α_1 e y_1 é obtida rearranjando a equação de mensuração. Para verificar isso, basta somar e subtrair $\alpha_{1|0}$ de α_1 , assim:

$$\begin{aligned} \alpha_1 &= \alpha_{1|0} + (\alpha_1 - \alpha_{1|0}) \\ y_1 &= Z_1 \alpha_{1|0} + d_1 + (Z_1 (\alpha_1 - \alpha_{1|0}) + \epsilon_1) \end{aligned}$$

Note que $(\alpha_1 - \alpha_{1|0})$ tem média zero, portanto $E(\alpha_1) = \alpha_{1|0}$ e $E(y_1) = Z_1 \alpha_{1|0} + d_1$.

A2.3. Modelo invariante no tempo

Como dito anteriormente, uma subclasse de modelos de estado espaço é quando os elementos de SM são constantes no tempo. Quando isso ocorre, é possível derivar uma série de propriedades: controlabilidade, observabilidade, estabilizabilidade, e detectabilidade. Essas propriedades podem eventualmente serem úteis em determinadas aplicações, como quando se quer descobrir se um modelo específico converge para um estado estacionário. Apesar de essa classe ser aplicada para o caso em que todos os elementos são constantes, tais propriedades se mantêm mesmo quando c e d variam no tempo. Portanto, o pressuposto básico dessa classe de modelos é que seus hiperparâmetros são invariantes no tempo.

Para a definição dessas propriedades, Harvey (1989) parte de um sistema dinâmico determinístico dado pelas equações:

$$\begin{aligned} y_t &= Z\alpha_t \\ \alpha_t &= T\alpha_{t-1} + Gu_t \end{aligned} \tag{A.28}$$

Ou seja, é o mesmo modelo de estado espaço definido anteriormente, porém sem seus respectivos termos estocásticos e sem as variáveis c e d . Sendo assim, temos as seguintes definições:

Definição A.1: O sistema (A.28) é controlável se $\text{posto}[G, TG, \dots, T^{m-1}G] = m$.

Se o sistema é controlável, isso significa que os vetores u_{t+1}, \dots, u_{t+m} podem ser escolhidos de maneira que um α_{t+m}^* possa ser obtido. Se G for invertível, essa condição é claramente satisfeita e a equação de estado pode ser rearranjada de tal modo que u possa ser posto em termos de α e T .

Definição A.2: O sistema (A.28) é observável se $\text{posto}[Z', T'Z', \dots, (T')^{m-1}Z'] = m$.

A implicação do vetor α_t ser observável é que ele pode ser determinado de maneira exata se as informações de y_t, \dots, y_{t+m-1} estiverem disponíveis. Também se supõe que os valores de $u_{t+1}, \dots, u_{t+m-1}$ são conhecidos.

Definição A.3: O sistema (A.28) é estabilizável se $\exists S_{m \times m}: |\lambda_i(T + GS')| < 1, i = 1, \dots, m$.

Definição A.4: O sistema (A.28) é detectável se $\exists D_{m \times N}: |\lambda_i(T - DZ)| < 1, i = 1, \dots, m$.

Esses dois últimos casos são versões mais fracas do que as primeiras duas propriedades. Nesse sentido, a controlabilidade implica em estabilizabilidade e observabilidade implica em detectabilidade.

Apesar disso, tais propriedades formuladas para esse modelo são aplicáveis para o modelo de estado espaço estocástico invariante no tempo, basta ter como condição que a matriz de covariâncias do termo de erro seja uma identidade. Isso pode ser feito com uma simples transformação.

Seja modelo:

$$\begin{aligned} y_t &= Z\alpha_t + d_t + \epsilon_t, & \text{Var}(\epsilon_t) &= H \\ \alpha_t &= T\alpha_{t-1} + c_t + R\eta_t, & \text{Var}(\eta_t) &= Q \end{aligned} \quad (\text{A.29})$$

Defina R^* como uma matriz $n \times n$ tal que $R^*R^{*'} = Q$, então a equação de transição pode ser reescrita como $\alpha_t = T\alpha_{t-1} + c_t + G\eta_t^*$, onde $G = RR^*$ e η_t^* é redefinido como o termo de erro com $E(\eta_t^*) = 0$ e $\text{Var}(\eta_t^*) = I$. Com essa transformação é possível verificar sem grandes dificuldades as propriedades a serem formuladas para o modelo determinístico.

Podemos também aplicar o filtro de Kalman para o modelo (A.29). Desse modelo, podemos verificar se o modelo está no estado estacionário observando sua matriz de covariâncias. Para isso, basta que a matriz de covariâncias seja invariante no tempo, \bar{P} . Nesse caso, a recursão para a covariância é trivial, resumindo o modelo em termos de:

$$\begin{aligned} a_{t+1|t} &= \bar{T}a_{t|t-1} + \bar{K}y_t + (c_{t+1} - \bar{K}d_t), & \bar{T} &= \bar{T} - \bar{K}Z \\ \bar{K} &= T\bar{P}Z'(Z\bar{P}Z' + H)^{-1} \end{aligned} \quad (\text{A.30})$$

Para que o filtro seja estável, as raízes de T devem ser menor que um em valor absoluto. Além disso, é possível verificar se há alguma solução de estado estacionário observando se a equação de Riccati é satisfeita quando P é constante. Assim, temos a seguinte condição chamada equação de Riccati algébrica (ARE):

$$\bar{P} - T\bar{P}T' + T\bar{P}Z'(Z\bar{P}Z' + H)^{-1}Z\bar{P}T' - RQR' = 0 \quad (\text{A.31})$$

É nessa verificação que os conceitos anteriormente definidos podem ser úteis. Isto porque em geral a solução de ARE pode ser difícil ou mesmo impossível obter uma

solução explícita. Sendo assim, Harvey (1989) enuncia três resultados que podem jogar luz sobre a possibilidade de existência de convergência para um estado estacionário:

- i) Se $|\lambda_i(T)| < 1, i = 1, \dots, m$ e $P_{1|0}$ é positiva semidefinida, então $\lim_{t \rightarrow \infty} P_{t+1|t} = \bar{P}$ com \bar{P} independente de $P_{1|0}$. Nesse caso a convergência para o estado estacionário é exponencialmente rápida se \bar{P} é a única solução positiva semidefinida de ARE;
- ii) Se o sistema é detectável, estabilizável e $P_{1|0}$ é positiva semidefinida, então o resultado acima se mantém. Mesmo que o sistema não seja estável;
- iii) Se o sistema é observável e $P_{1|0} = \bar{P}$, então o resultado (i) se mantém.

O resultado (ii) é um caso mais geral de (i), e o resultado (iii) não garante que a convergência será exponencialmente rápida.

Para os valores iniciais do Filtro de Kalman, a primeira alternativa é utilizar como valores iniciais da média e variância dados pela distribuição incondicional do vetor estado. Se T e c forem constantes, então $E(\alpha) = (I - T)^{-1}c$. A matriz de covariância incondicional é aquela que resolve a equação $P = TPT' + RQR'$. Para modelos mais complexos, Harvey (1989) também sugere que P pode ser avaliada a partir do fato que $vec(P) = [I - T \otimes T]^{-1}vec(RQR')$. Uma coisa que o autor chama atenção é que, como a distribuição incondicional de α_0 e α_1 é igual, então uma opção é iniciar o filtro com $a_0 = a_{1|0} = 0$ e $P_0 = P_{1|0} = P$. Porém essas opções só valem quando o sistema é estacionário. Quando ele é não-estacionário, não há uma distribuição incondicional definida. Nesse caso os valores iniciais devem ser definidos em termos de um *prior* difuso. Nesse caso, supõe-se que $P_0 = \kappa I$, onde κ é um escalar positivo e a *prior* difusa é obtida quando $\kappa \rightarrow \infty$. Quando este é o caso, $P_0^{-1} = 0$. Na prática o que pode ser feito para se aproximar o *prior* é definir um κ suficientemente grande. Uma exposição para a definição do *prior* e condições para convergência pode ser encontrada em Harvey (1989).

A2.4 Estimação por máxima verossimilhança e decomposição da previsão dos erros

Como se trata de um modelo para séries de tempo, as observações não são i.i.d. e, portanto, a função de verossimilhança deve ser definida em termos da função de probabilidade condicional:

$$\mathcal{L}(y; \varphi) = \prod_{t=1}^T p(y_t | Y_{t-1} = \{y_{t-1}, \dots, y_1\}) \quad (\text{A.32})$$

Onde Y_{t-1} é o conjunto de informação em $t - 1$. Se o *prior* tiver uma distribuição normal multivariada, então $y_t|Y_{t-1}$ terá também distribuição normal. Além disso, o próprio Filtro de Kalman dá os valores da média e covariância. Com isso, a função de verossimilhança construída pelo sistema é dada por:

$$\ln \mathcal{L} = -0,5(NT \ln 2\pi + \sum_{t=1}^T \ln |F_t| + \sum_{t=1}^T v_t' F_t^{-1} v_t) \quad (\text{A.33})$$

Onde $v_t = y_t - \tilde{y}_{t|t-1}$, $t = 1, \dots, T$. E \tilde{y} é o estimador de mínimos quadrados de y . Assim, v é o vetor de predição dos erros. Harvey comenta que por isso é comum chamar a forma (A.33) de forma de decomposição da previsão dos erros da verossimilhança. Se existir um *prior* não impróprio, então o filtro de Kalman resultará numa função de verossimilhança exata com base nas observações via (A.33). Na ausência dessa informação prévia sobre o *prior*, é ainda possível encontrar técnicas alternativas para a estimação da mesma. Harvey (1989) cita duas alternativas: i) a utilização de uma abordagem bayesiana na definição do *prior*; ou ii) a utilização de uma amostra grande o suficiente, tornando desimportante a especificação da condição inicial. Como visto, com o Filtro de Kalman há formas de se encontrar o *prior* próprio no caso de um modelo invariante no tempo estacionário. Para o caso não-estacionário também há como definir um *prior* difuso. A computação nesse caso pode ser mais complexa. Harvey (1989) apresenta alguns algoritmos para simplificar a computação da condição inicial no caso não-estacionário. As propriedades assintóticas para essa função de verossimilhança podem ser vistas em Harvey (1989) e Hamilton (1994)³⁴.

A2.5 Estimação por Quase-máxima Verossimilhança

O filtro de Kalman ainda é aplicável para o cálculo da projeção linear de y_{t+s} nas variáveis passada mesmo que ϵ e η sejam não-gaussianos. Conforme Hamilton (1994), a função de verossimilhança pode ser maximizada em relação ao vetor de hiperparâmetros, ϕ , mesmo dentro de um sistema não gaussiano. Os resultados das estimativas ainda serão consistentes e assintoticamente normais for satisfeita a seguinte propriedade:

³⁴ Seguindo Harvey (1989), temos que $\hat{\beta}A(\psi) = \lim T^{-1}\hat{\beta}(\psi)$ e $Avar(\hat{\psi}) = T^{-1}\hat{\beta}A(\Psi)$, onde $\hat{\psi}$ é o estimador de ψ e $\hat{\beta}(\psi)$ é a matriz de informação, se três condições forem satisfeitas. A saber: (i) ψ deve ser um ponto interior do espaço de parâmetros; (ii) $\ln \mathcal{L}$ deve ser triplamente diferenciável em relação a ψ e contínuo na vizinhança do verdadeiro parâmetro; e (iii) ψ é identificável. Em Hamilton (1994), se, além disso, (iv) todos os autovalores de dos parâmetros das variáveis exógenas estiverem dentro do círculo unitário; (v) as variáveis em Z , exceto a constante, se comportarem assintoticamente como um processo estacionário-covariante não-determinístico linear e tiver posto cheio, então o vetor de hiperparâmetros estimado é assintoticamente normal e consistente.

$$T^{-\frac{1}{2}}(\tilde{\varphi} - \varphi_0) \xrightarrow{L} N(0, [\mathcal{I}_{2D} \mathcal{I}_{OP}^{-1} \mathcal{I}_{2D}]^{-1}) \quad (\text{A.34})$$

Onde \mathcal{I}_{OP} é o produto externo do estimador da matriz de informação, \mathcal{I}_{2D} é uma matriz de informação de tamanho T , φ_0 é o valor verdadeiro de φ e $\tilde{\varphi}$ é o estimador de φ . Adicionalmente, \mathcal{I}_{2D} e \mathcal{I}_{OP} são dados por:

$$\mathcal{I}_{2D} = \text{plim } T^{-1} \sum_{t=1}^T \left. \frac{\partial^2 \ln p(y_t | Y_{t-1}; \varphi)}{\partial \varphi \partial \varphi} \right|_{\varphi=\varphi_0}$$

$$\mathcal{I}_{OP} = \lim T^{-1} \sum_{t=1}^T \left[\left. \frac{\partial \ln p(y_t | Y_{t-1}; \varphi)}{\partial \varphi} \right|_{\varphi=\varphi_0} \right] \left[\left. \frac{\partial \ln p(y_t | Y_{t-1}; \varphi)}{\partial \varphi} \right|_{\varphi=\varphi_0} \right]'$$

A3. Modelo de Fatores Dinâmicos

A ideia do modelo de fatores dinâmicos é bastante similar à análise de fatores, com a diferença que os fatores latentes são postos dentro de um contexto dinâmico. Dentro desse contexto podemos também chamar tais fatores de tendências comuns. Sendo assim, o pressuposto básico do modelo é que q fatores latentes, f_t , dirigem os movimentos de um vetor de séries de tempo X_t . Assim como no caso estático, o que dá sentido ao modelo é supor que o número de séries de tempo, N , é maior do que o número de fatores latentes. Ou seja, $N > q$.

Quando aplicado a um contexto de séries de tempo, o modelo básico pode ser expresso no seguinte sistema:

$$X_t = l(L)f_t + \epsilon_t \quad (\text{A.35})$$

$$f_t = \gamma(L)f_{t-1} + \eta_t \quad (\text{A.36})$$

Onde X_t é um vetor $N \times 1$, f_t é $q \times 1$, $l(L)$ é um vetor de parâmetros $N \times q$, $\gamma(L)$ é um vetor de parâmetros $q \times q$ e L é o operador de defasagens³⁵. Dentro desse arcabouço básico, parte-se dos seguintes pressupostos:

- i) (A.34) e (A.35) são estacionários;
- ii) $E(\epsilon_t \eta'_{t-k}) = 0, \forall k$.

Para o chamado Modelo de fatores dinâmicos *exato*, adicionalmente é pressuposto:

- iii) $E(\epsilon_{it} \epsilon_{js}) = 0, \forall s \text{ se } i \neq j$.

Além disso, se for possível assumir que (ϵ_t, η_t) são gaussianos, então é possível obter previsões eficientes para a variável com uma regressão populacional dessa

³⁵ Vale lembrar que $a(L) = a_0 + a_1L + a_2L^2 + \dots + a_pL^p$.

variável sendo explicada pelos fatores e suas defasagens. A vantagem nesse tipo de aplicação é reduzir a dimensão do modelo.

Outra coisa que se deve levar em consideração é a identificação do modelo. Por simplicidade, considere o caso em que os fatores seguem um processo de passeio aleatório com *drift*. Nesse caso,

$$X_t = lf_t + \mu_0 + \epsilon_t, \quad Var(\epsilon_t) = \Sigma_\epsilon \quad (\text{A.37})$$

$$f_t = f_{t-1} + \mu_1 + \eta_t, \quad Var(\eta_t) = \Sigma_\eta \quad (\text{A.38})$$

Nessa forma o modelo não é diretamente identificável. Assim, é necessário impor restrições na matriz de loadings, l_0 , e na matriz de covariâncias do erro dos fatores, Σ_η . Na análise de fatores usual é usual impor algum tipo de padronização de maneira que $\Sigma_\eta = I_k$, porém ainda assim essa transformação é insuficiente para identificação. Isto porque ainda não se mantém possível a premultiplicação por uma matriz H ortogonal sem alterar o formato do modelo, ou seja, mantendo possível a existência de múltiplas soluções. Duas recomendações para identificação podem ser encontradas em Harvey (1989) e Fernandez-Macho (1997), que são expressas definindo que:

$$l = \left\{ \frac{l_{ij}}{l_{ij}} = 0, \text{ se } i < j; a_{ii} = 1 \right\} \quad (\text{A.39})$$

$$\Sigma_\eta = \left\{ \frac{\sigma_{\eta,ij}}{\sigma_{\eta,ij}} = 0, i \neq j; \sigma_{\eta,ii} = \sigma_{\eta,i}^2 \right\} \quad (\text{A.40})$$

Basicamente o que as equações (A.39) e (A.40) dizem é que os elementos $l_{ij} \in l$ devem ser zero sempre que $j > i$. Adicionalmente, Σ_η deve ser uma matriz diagonal. Essas restrições implicam que os fatores comuns serão não correlacionados entre si (devido à Σ_η ser diagonal) e também que o vetor de variáveis observáveis, X_t , é tal que X_{1t} depende da primeira tendência comum, mas não das demais; já X_{2t} depende da primeira e segunda tendência comum, mas não das demais; e assim sucessivamente.

Por fim, dentro do *framework* geral do modelo de fatores dinâmicos, restam duas questões. A primeira questão concerne o critério de decisão para o número de fatores que devem ser incluídos no modelo, isto é, encontrar o valor correto de q . Em segundo lugar, cabe saber como estimar os coeficientes desses q fatores e, eventualmente, construir os escores desses fatores. As próximas subseções tratarão dessas questões. Após isso, também é feita uma subseção para discutir o tratamento adequado para do modelo não estacionário.

A3.1. Número de fatores

Os procedimentos para a obtenção do número r de fatores estáticos é igual ao visto anteriormente na análise de fatores visto na seção A1. Para a escolha do número dos q fatores dinâmicos, Stock e Watson (2010) citam alguns procedimentos. Em um deles, proposto por Bai e Ng (2007), que consiste em estimar um VAR sobre \hat{F}_t e seus respectivos lags. Desse modelo é possível comparar os autovalores da matriz de covariâncias do modelo a partir de um limite que depende em N e T . Stock e Watson (2010) também citam alguns estimadores. Stock e Watson (2010) também notam que o critério de informação de Bai-Ng (2002) pode de ser aplicado à matriz de covariâncias amostral dos resíduos também resulta numa estimativa consistente do número de fatores dinâmicos.

A3.2. Estimação dos fatores

Para a estimação dos fatores no contexto de DFMs, Stock e Watson (2010) fazem uma breve exposição da forma como eles vêm sendo historicamente tratados. A utilização de fatores dinâmicos parte do trabalho seminal de Geweke (1977), com uma aplicação em Sargent e Sims (1977). Esses primeiros trabalhos buscavam por evidências de fatores dinâmicos usando métodos de domínio da frequência. Essa forma, entretanto, não é a melhor possível, pois não permitem a estimação direta de f_t . Isso foi resolvido ao focar a estimação para métodos de domínio do tempo. Dentro desse contexto, os autores citam três gerações de DFMs, que diferem entre si na forma de estimação.

A primeira geração o problema consiste em colocar o sistema (A.34-35) na forma de estado espaço e estimar via Máxima Verossimilhança utilizando Filtro de Kalman. Para tal, uma forma possível de colocar o sistema (A.34-35) na forma de estado espaço é reescrever o sistema da seguinte forma:

$$X_t = \Lambda F_t + \epsilon_t \quad (\text{A.37})$$

$$\Phi(L)F_t = G\eta_t \quad (\text{A.38})$$

Onde há p defasagens, $F_t = (f'_t, \dots, f'_{t-p})$, $\Lambda = (l_0, \dots, l_p)$ e $\Phi(L)$ é uma matriz formada em termos de 1s, 0s e os valores de $\gamma(L)$ tais que representem a equação (A.36). Por sua vez, G é uma matriz formada de 1s e 0s escolhidos também de maneira a representar (A.36)³⁶. Vale lembrar que cada l_i é uma matriz $N \times q$. Esse modelo

³⁶ Por exemplo, suponha:

também permite que o erro da equação de mensuração, ϵ_t , seja um processo autoregressivo, ou seja:

$$d_i(L)e_t = \epsilon_{it}, \quad i = 1, \dots, N \quad (\text{A.39})$$

Onde $\epsilon_{it} \sim NID(0, \sigma_{\epsilon_i}^2)$ e $\eta_{jt} \sim NID(0, \sigma_{\eta_j}^2)$, com $\{\epsilon_t\}$ e $\{\eta_t\}$ sendo independentes entre si. Stock e Watson (2010) chama essa forma dada por (A.37-38) de “forma estática”, por conta dos fatores aparecerem apenas contemporaneamente na forma de estado espaço. Uma vez na forma de estado espaço, é então possível aplicar o Filtro de Kalman descrito na seção anterior e obter a função de verossimilhança. Isso permite a obtenção dos parâmetros da equação, incluindo o valor dos *loadings* e também permite construir os valores estimados para F_t .

Uma dificuldade inerente nessa forma é devido aos custos computacionais em se otimizar funções não lineares. Para tanto, tal abordagem se via restrita a modelos com N pequeno. Para tentar superar essa dificuldade, uma segunda geração de estimadores se baseou na construção de estimadores não paramétricos quando N é grande a partir de métodos de ponderação, sendo a Análise de Componentes Principais (PCA) e suas variações a principal forma de se estimar os *loadings* do modelo de fatores. Esse método é consistente e permite a obtenção dos escores de f_t .

Como em uma ponderação os efeitos idiossincráticos irão convergir para zero pela lei dos grandes números, isso resulta que a média ponderada dos elementos de X_t gera uma combinação linear em que apenas os fatores comuns permanecem, e é por conta disso que a ponderação é uma forma de realizar a estimação. Esse tipo de estratégia é consistente com a forma de estado espaço apresentada anteriormente. Partindo desse modelo, temos que o fator, F_t , pode ser tratado como um vetor de parâmetros a ser estimado a partir dos dados. Para que isso ocorra, Stock e Watson (2010) listam as seguintes condições:

- i) $N \rightarrow \infty \Rightarrow N^{-1}\Lambda'\Lambda \rightarrow D_\Lambda$, onde D_Λ é rxr e tem posto cheio;
- ii) $\max_{\epsilon} \text{eval}(\Sigma_\epsilon) \leq c < \infty, \forall N$.

$$\begin{aligned}
 X_t &= l_0 f_t + \epsilon_t \\
 f_t &= \gamma_0 f_{t-1} + \eta_t
 \end{aligned}$$

Se $\Lambda = (l_0, 0)$, $F_t = (f_t', f_{t-1}')$, $\Phi = (1, -\gamma_0)$ e $G = (1, 0)$, então:

$$X_t = \Lambda F_t + \epsilon_t = [l_0 \quad 0] \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} + \epsilon_t = l_0 f_t + \epsilon_t$$

$$\Phi F_t = G \eta_t \Rightarrow [1 \quad -\gamma_0] \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} = [1 \quad 0] \eta_t \Rightarrow f_t = \gamma_0 f_{t-1} + \eta_t.$$

Onde $\max_{eval}(\Sigma_\epsilon)$ busca o maior autovalor da matriz de covariâncias de Σ_ϵ . Essas condições garantem que os fatores são heterogêneos, afetam todos ou quase todas as séries e os efeitos idiossincráticos têm correlação limitada entre as séries.

Seja um vetor de pesos W com dimensão $N \times r$ e normalizado de forma que $\frac{W'W}{N} = I$, então:

$$\hat{F}_t(N^{-1}W) = N^{-1}W'X_t \quad (\text{A.40})$$

Se $N^{-1}W'\Lambda \rightarrow H$ quando $N \rightarrow \infty$, onde H é uma matriz $r \times r$ com posto cheio, então isso atende as condições (i) e (ii), e então o estimador $\hat{F}_t(N^{-1}W)$ é consistente para F_t ³⁷. Para a estimação via PCA, a estratégia consiste em fazer $W = \hat{\Lambda}$, onde $\hat{\Lambda}$ é a matriz dos maiores autovetores amostrais. Stock e Watson (2010) sugerem um procedimento de estimação dos PCAs via mínimos quadrados:

$$\min_{F_1, \dots, F_T, \Lambda} V_r(\Lambda, F) = (NT)^{-1} \sum_{t=1}^T (X_t - \Lambda F_t)'(X_t - \Lambda F_t) \quad (\text{A.41})$$

$$s. a. N^{-1}\Lambda'\Lambda = I_r \quad (\text{A.42})$$

A solução desse problema chega um $\hat{\Lambda}$ igual aos autovetores escalonados de $\hat{\Sigma}_{XX}$ correspondendo aos maiores r . É também possível obter uma estimação via componentes principais generalizados (GPC). Esse é aplicado quando a matriz de covariância dos erros não é proporcional à matriz identidade. Para corrigir isso, basta corrigir o procedimento de minimização acima com o inverso da matriz de covariância dos erros amostrais:

$$\min_{F_1, \dots, F_T, \Lambda} V_r(\Lambda, F) = T^{-1} \sum_{t=1}^T (X_t - \Lambda F_t)' \Sigma_\epsilon^{-1} (X_t - \Lambda F_t) \quad (\text{A.43})$$

Nesse caso $\hat{\Lambda}$ é formado pelos autovetores escalonados dos r maiores autovetores de $\Sigma_\epsilon^{-0,5} \hat{\Sigma}_X \Sigma_\epsilon^{-0,5'}$. Na prática a matriz Σ_ϵ não é observada, porém o estimador construído com base nos resíduos tende a se comportar de maneira pobre (Stock e Watson, 2010). Stock e Watson (2010) sugerem pelo menos três procedimentos para obter $\hat{\Sigma}_\epsilon$:

- 1) Dado que $\Sigma_X = \Sigma_{\Lambda F} + \Sigma_\epsilon$, então $\hat{\Sigma}_\epsilon = \hat{\Sigma}_X - \hat{\Sigma}_{\Lambda F}$;
- 2) $\hat{\Sigma}_\epsilon = \text{diag}(s_{\epsilon_1}^2, \dots, s_{\epsilon_N}^2)$, onde s é o estimador da variância do erro de cada uma das séries;
- 3) Obter F_t via componentes principais e então separar N autoregressões para acomodar os resíduos da regressão de X_{it} em \hat{F}_t . Após isso, X_{it} é quase-

³⁷ Note que $\hat{F}_t(N^{-1}W) = N^{-1}W'(\Lambda F_t + \epsilon_t) = N^{-1}W'\Lambda F_t + N^{-1}W'\epsilon_t$. Por construção, se $N \rightarrow \infty$, $N^{-1}W'\Lambda \rightarrow H$ e, pela lei dos grandes números, $N^{-1}W'\epsilon_t \xrightarrow{p} 0$. Assim, $\hat{F}_t(N^{-1}W) \xrightarrow{p} HF_t$.

diferenciado utilizando os coeficientes da i -ésima autoregressão. Então é aplicado o método (2).

Desses, apenas o terceiro método consegue lidar com a possibilidade de autocorrelação serial nos termos de erro. O terceiro método é exposto em Stock e Watson (2005 *apud* STOCK E WATSON, 2010) em maiores detalhes.

Apesar da consistência dos modelos da segunda geração, eles não são necessariamente os mais eficientes. É possível melhorar sua precisão, e é nesse sentido que surge uma terceira geração de estimadores. Esses estimadores combinam as duas abordagens anteriores. Assim, os estimadores não-paramétricos são usados como uma primeira aproximação, e então essas estimativas são utilizadas para estimar os parâmetros do modelo de estado e espaço, resolvendo o problema de dimensionalidade presente na primeira geração. Além disso, a junção dos componentes principais com a estimação via Filtro de Kalman permite um tratamento mais adequado para casos em que há *missing data*.

A aplicação desse híbrido consiste em dois estágios. Primeiro os fatores são estimados via componentes principais ou algum outro método similar. Após isso, os fatores estimados, \hat{F}_t , são usados para estimar os parâmetros desconhecidos da representação em estado espaço via Filtro de Kalman. A estimação nesse caso irá depender de como a representação em estado espaço é construída. Se ela for construída como em (A.37-39), formando o modelo de estado espaço “estático”, então Λ é estimado a partir de uma regressão das séries nos fatores estimados, e os resíduos são usados para estimar as regressões autoregressões univariadas que o modelo permite. Os parâmetros de \hat{F}_t em seus lags podem ser estimados a partir de um VAR, bem como a variância de seus erros.

Outra possibilidade é a estimação do modelo em estado espaço com fatores dinâmicos aplicando diretamente o sistema (A.35-36) e (A.39). Nesse caso o vetor estado é formado por f_t . Quando formulado dessa forma, há menos parâmetros a serem estimados, e o VAR para f_t resulta em restrições sobre o vetor $\Phi(L)$. Dadas as estimativas \hat{f}_t , é possível computar os coeficientes de (A.35) via regressão das séries nos fatores estimados e seus lags; os parâmetros de (A.36) são obtidos a partir de um VAR nos fatores estimados; e, finalmente, os resíduos (A.39) podem ser estimados a partir da regressão usada em (A.35). Uma vez obtidos os parâmetros, é possível melhorar as estimativas computando o alisamento do Filtro de Kalman (*Kalman*

Smoother). Também é possível utilizar as estimativas via componentes principais como *prior* na estimação via máxima verossimilhança.

Há também uma forma alternativa de prosseguir, com a utilização de métodos bayesianos para a estimação, que pode ser vista em mais detalhes em Stock e Watson (2010).

A3.3 Não-estacionariedade

Em geral se assume o vetor de variáveis observáveis são estacionários. Isso não é um problema quando é possível tomar a primeira diferença dessas séries. Ainda assim, certos cuidados devem ser tomados para lidar com a cointegração entre as séries. Conforme Stock e Watson (2010), a teoria econométrica dos componentes principais dos fatores necessita de modificações para lidar com variáveis cointegradas. No caso, sem modificações em $\hat{\Sigma}_x$, essa matriz terá muitos autovalores grandes, mesmo que os elementos de X_t sejam independentes. Como a PCA e a escolha do número de fatores dependem da informação dos autovalores, é provável que ignorar resulte num modelo mal especificado.

No caso em que os fatores são não-estacionários e $I(1)$, é possível obter estimadores consistentes se os erros forem $I(0)$ ou $I(1)$ (BAI, 2004 *apud* STOCK E WATSON, 2010; e BAI E NG, 2004 *apud* STOCK E WATSON, 2010). Para casos mais gerais, é possível também trabalhar com modelos de correção de erros aumentados por fatores (FECM). Nesse caso Banerjee, Marcellino e Masten (2008) apresentam um modelo em que X_t é um vetor de séries não-estacionárias e que pode ser modelado como um $\text{Var}(p)$. O modelo base apresentado pelos autores supõe a existência de uma tendência comum, f_t , porém essa tendência não é autorregressiva. Sob essa representação os autores introduzem o chamado modelo de correção de erros fatorial (FECM). De maneira geral Fernández-Macho (1997) também apresenta uma metodologia de estimação no domínio da frequência para o modelo de fatores dinâmicos no caso não-estacionário, quando os fatores não observados apresentam raiz unitária.

A4. Decomposição da Variância

Seja

$$Y = \sum_{k=0}^K \eta_k + \varepsilon \quad (\text{A.44})$$

Onde $\eta_k = x_k \beta_k$ são as variáveis explicativas e seus respectivos coeficientes. Note que omitimos o subscrito de observações e tempo para não carregar a notação.

A variância de (A.44) é dada por:

$$\text{Var}(Y) = \sum_{k=0}^K \text{Var}(\eta_k) + \sum_{k \neq w} \sum_{w \neq k} \text{Cov}(\eta_k, \eta_w) + \text{Var}(\varepsilon) \quad (\text{A.45})$$

Dividindo (A.45) por $\text{Var}(Y_{i,t})$, resulta em:

$$\sum_{k=0}^K \frac{\text{Var}(\eta_k)}{\text{Var}(Y)} + 2 \sum_{k \neq w} \frac{\text{Cov}(\eta_k, \eta_w)}{\text{Var}(Y)} + \frac{\text{Var}(\varepsilon)}{\text{Var}(Y)} = 1 = R^2(Y_{i,t}; \eta_1, \dots, \eta_K) \quad (\text{A.46})$$

Onde $R^2(Y_{i,t}; \eta_1, \dots, \eta_K)$ é simplesmente o R^2 de uma regressão de cada efeito sobre a variável dependente.

Gibbons et. al. (2012) apresentam, além da Parcela da Variância Bruta (PVB) ou *Raw Variance Share* (RVS), também outras maneiras de avaliar a contribuição de uma variável explicativa sobre a variância da variável dependente: a Parcela da Variância Correlacionada (PVC) ou *Correlated Variance Share* (CVS), a Parcela da Variância Não-correlacionada (PVN) ou *Uncorrelated Variance Share* (UVS), e Parcela da Variância Balanceada (PVBal) ou *Balanced Variance Share* (BVS).

A primeira, a PVB é simplesmente o R^2 de uma regressão simples entre Y e uma das variáveis explicativas, assim,

$$\frac{\text{Var}(\tilde{\eta}_k)}{\text{Var}(Y)} = R^2(Y; \tilde{\eta}_k) \quad (\text{A.47})$$

Onde $\tilde{\eta}_k \equiv x_k \tilde{\beta}_k$. $\tilde{\beta}_k$ é o coeficiente da regressão de Y em x_k sem controlar pelas demais variáveis. Como tal, por não ser correlacionada com o termo de erro, indiretamente ela também assume que não há qualquer correlação entre η_k qualquer outro η_w , $w \neq k$.

Já a PVC considera a proporção da variância diretamente de (A.46). Primeiro é feita uma regressão de Y em todas as variáveis explicativas e então se obtém a estimativa de cada beta, $\hat{\beta}_k$. A partir disso é construído o PVC como:

$$\frac{\text{Var}(\hat{\eta}_k)}{\text{Var}(Y)} = \frac{\hat{\beta}_k^2 \text{Var}(x_k)}{\text{Var}(Y)} \quad (\text{A.48})$$

Por sua vez, a PVN é uma forma de cálculo que mostra a parcela da variância explicada apenas pela porção da variável explicativa não-correlacionada com nenhuma das outras. Apesar de o significado ser similar a PVC, os resultados não são necessariamente iguais. Neste caso é feita uma regressão de Y em cada um dos x_k para obter

$R^2(Y; \eta_1, \dots, \eta_K)$. Em seguida é feita uma regressão apenas de Y em um dos x_k para obter $R^2(Y; \eta_k)$. Resultando em

$$PVN = R^2(Y; \eta_1, \dots, \eta_K) - R^2(Y; \eta_k) \quad (\text{A.49})$$

Isso dará a parcela da variância explicada apenas pelo x_k selecionado. Por fim, os autores apresentam a PVBal, dada por:

$$PVBal = \frac{s(\hat{\eta}_k)}{s(Y)} r(Y; \hat{\eta}_k) = \frac{Cov(Y; \hat{\eta}_k)}{Var(Y)} = \frac{(Var(\hat{\eta}_k) + \sum_{1 < j} Cov(\eta_k, \eta_j))}{Var(Y)} \quad (\text{A.50})$$

Onde $s(\cdot)$ é o desvio-padrão da variável em questão e $r(\cdot)$ é o coeficiente de correlação. O resultado do lado direito vem diretamente de (A.46), onde as covariâncias são divididas de maneira que seja considerada a covariância do elemento k com cada uma das demais variáveis. Como ressalta Gibbons et. al. (2012), não há uma justificativa óbvia para essa divisão na variância. Tal divisão é feita assim apenas para que a soma das parcelas individuais somem $R^2(Y; \eta_1, \dots, \eta_K)$.

A5. Método dos Momentos Generalizados

Esta seção apresenta o Método dos Momentos Generalizados e algumas extensões (first-differenced GMM e GMM sistêmico, tal que o primeiro foi desenvolvido por Holtz-Eakin, Newey e Rosen (1988) e Arellano e Bond (1991), e o segundo foi delineado por Bundell e Bond (1998) e Arellano e Bover (1995)). Esta exposição toma como referência Hansen (2016), Bond, Hoeffler e Temple (2001), Roodman (2009) e Wooldridge (2010).

O Método dos Momentos consiste em utilizar as *condições de momento* de um modelo específico para encontrar uma estimativa dos parâmetros a partir do conjunto de dados. A ideia aqui é que, se conhecermos os momentos de um dado modelo paramétrico, então podemos organizar esses momentos em um sistema de equações e resolvê-lo, encontrando assim o valor dos parâmetros. De modo geral, seja uma variável X com função de densidade de probabilidade $f(X; \theta)$, tal que $\theta = (\theta_1, \dots, \theta_L)$ são os L parâmetros dessa distribuição. Se conhecermos a forma funcional dos K primeiros momentos de $f(X; \theta)$, então podemos estabelecer o seguinte sistema de equações:

$$\begin{aligned} E(W) &= h_1(\theta) \\ E(W^2) &= h_2(\theta) \\ &\vdots \\ E(W^k) &= h_k(\theta) \end{aligned}$$

Que possui os parâmetros como incógnitas. O método dos momentos consiste em resolver a contraparte amostral deste sistema. Pelo Teorema do Limite Central e Teorema do Mapeamento Contínuo, sabemos que um estimador consistente de um momento pode ser dado por sua respectiva média amostra, isto é, $\left(\frac{1}{n}\sum_{i=1}^n W_i^j = h_j(\hat{\theta})\right)$ para cada j momento. O θ que resolve o sistema acima, $\hat{\theta}$, é chamado de estimador do Método dos Momentos. Isso implica que teremos um sistema com L incógnitas e K equações. Uma dificuldade com este método é que ele nem sempre terá solução. Quando $K > L$ o sistema estará sobreidentificado e, portanto, será impossível resolver o sistema com exatidão.

O Método dos Momentos Generalizados (GMM) surgem como uma generalização do Método dos Momentos, que busca contornar este problema minimizando uma determinada normal das médias amostrais das condições de momento.

Desse modo, podemos definir o GMM da seguinte forma:

Seja w_i um vetor de variáveis $L \times 1$, θ_0 é um vetor de parâmetros populacionais $K \times 1$ e $g(w_i; \theta_0): \mathbb{R}^L \rightarrow \mathbb{R}^M$ como sendo um vetor de funções contínuas, que podem ser ou não ser lineares. Por hipótese, suponhamos que $E[g(w_i; \theta_0)] = 0$, então teremos M condições de momentos e podemos aplicar o GMM para estimar θ . Neste caso, o estimador de GMM para θ_0 é:

$$\hat{\theta} = \operatorname{argmin}_{\theta} (\sum_{i=1}^n g(w_i; \theta))' W (\sum_{i=1}^n g(w_i; \theta)) \quad (\text{A.51})$$

Tal que W é uma matriz positiva semidefinida qualquer. A condição de primeira ordem pode ser escrita como:

$$[\sum_{i=1}^n \partial_{\theta} g(w_i; \hat{\theta})]' W [\sum_{i=1}^n \partial_{\theta} g(w_i; \hat{\theta})] = 0 \quad (\text{A.52})$$

É possível mostrar que o estimador do GMM é consistente, com variância assintótica dada por:

$$AVar(\hat{\theta}) = \{H' \tilde{W} H\}^{-1} H' \tilde{W} G \tilde{W} H \{H' \tilde{W} H\}^{-1} \quad (\text{A.53})$$

Tal que $H = E[\partial_{\theta} g(w_i; \theta_0)]$ e $G = E[(g(w_i; \theta_0))(g(w_i; \theta_0))']$.

Se $\tilde{W} = G^{-1}$, então teremos o estimador eficiente do GMM, e a variância assintótica se reduz para:

$$AVar(\hat{\theta}) = \{H' G^{-1} H\}^{-1} \quad (\text{A.54})$$

Em geral não conhecemos G , então uma solução é substituir G por sua contraparte amostral, \hat{G} , neste caso a matriz de ponderação ótima será dada por:

$$W_n = \left(\frac{1}{n} \sum_{i=1}^n \left[g(w_i; \tilde{\theta}) g(w_i; \tilde{\theta})' \right] \right)^{-1} \quad (\text{A.55})$$

Tal que $\tilde{\theta}$ é um estimador consistente de θ_0 . Observe que a escolha de W ótima afeta apenas a eficiência, o estimador do GMM é consistente para qualquer W positiva semidefinida e simétrica. Desse modo, uma estratégia comum é estimar o GMM em dois estágios: primeiro estimar $\tilde{\theta}$ com $W_n = I$ e então usar essa estimativa para construir a matriz de ponderação ótima. O GMM estimado utilizando $W_n = \hat{G}^{-1}$ é conhecido como estimador GMM factível.

Apesar de consistente e assintoticamente eficiente, os desvios-padrão computados pelo GMM em dois estágios possui viés negativo. Windmeijer (2005) propõe uma correção para diminuir esse viés em amostras finitas.

A5.1. GMM Linear

Suponha que o processo gerador de dados seja dado por

$$y = x_i' \beta + \varepsilon_i \quad (\text{A.56})$$

Para $i = 1, \dots, N$. Tal que y é um vetor $N \times 1$, x_i é um vetor $K \times N$, β é um vetor de parâmetros $K \times 1$ e ε_i é um vetor $N \times 1$ tal que $E[x_i | \varepsilon_i] = 0$ (o que implica que $E[x_i \varepsilon_i] = 0$). Como temos K parâmetros e K momentos, as condições de momento serão dadas por:

$$E[x_i (y_i - x_i' \beta)] = 0 \quad (\text{A.57})$$

Neste caso o estimador do método dos momentos será dado por:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \{ (X' \varepsilon)' I (X' \varepsilon) \}$$

$$\hat{\beta} = \left(\sum_i x_i x_i' \right)^{-1} \sum_i x_i y_i \quad (\text{A.58})$$

Se houver algum x_i endógeno, então não é possível assumir que $E[\varepsilon_i | x_i] = 0$ para qualquer x_i e o método dos momentos não pode ser aplicado. Entretanto, se for possível obter um vetor de instrumentos, z_i , de dimensão $M \times 1$, então teremos as seguintes M condições de momento:

$$E[z_i \varepsilon_i] = E[z_i (y_i - x_i' \beta)] = 0 \quad (\text{A.59})$$

Se o número de parâmetros, K , for igual ao número de momentos, M , então o método dos momentos é aplicável. Se o sistema for sobreidentificado ($M > K$), então o procedimento é o do GMM. Neste caso,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\{(Z'\varepsilon)'W(Z'\varepsilon)\} \quad (\text{A.60})$$

Cuja condição de primeira ordem é:

$$\begin{aligned} -2Y'ZWZ'X + 2X'ZWZ'X\hat{\beta} &= 0 \\ \hat{\beta}_{GMM} &= (X'ZWZ'X)^{-1}X'ZWZ'Y \end{aligned} \quad (\text{A.61})$$

Se $W = (Z'Z)^{-1}$, então teremos o estimador de mínimos quadrados em dois estágios (2SLS), que é um estimador de GMM eficiente sob homocedasticidade. Sob heterocedasticidade, $W = (Z'DZ)^{-1}$ é a matriz que minimiza a variância condicional, e neste caso $\hat{\beta}$ é também o estimador de mínimos quadrados de três estágios (3SLS). A variância assintótica do GMM linear é dada, no caso homocedástico, por:

$$AVar(\hat{\beta}_{GMM}) = \sigma^2(X'ZWZ'X)^{-1}X'ZWZ'ZWZ'X(X'ZWZ'X)^{-1} \quad (\text{A.62})$$

Se $W = (Z'Z)^{-1}$, então

$$AVar(\hat{\beta}_{GMM}) = \sigma^2(X'Z(Z'Z)^{-1}Z'X)^{-1} \quad (\text{A.63})$$

No caso Heterocedástico e assumindo que a amostra é iid:

$$E[\varepsilon\varepsilon'|X, Z] = \operatorname{diag}(\sigma_i^2) = D$$

Neste caso,

$$AVar(\hat{\beta}_{GMM}) = (X'ZWZ'X)^{-1}X'ZWZ'DZWZ'X(X'ZWZ'X)^{-1}$$

E a matriz de ponderação ótima é dada por $\tilde{W} = (Z'DZ)^{-1}$, que reduz a variância assintótica para:

$$AVar(\hat{\beta}_{GMM}) = (X'Z(Z'DZ)^{-1}Z'X)^{-1} \quad (\text{A.64})$$

A5.2. First-differenced GMM

Em um painel, quando as realizações correntes da variável dependente são influenciadas por suas próprias realizações passadas, teremos um problema de endogeneidade, em que a variável defasada estará necessariamente correlacionada com o efeito fixo. Quando se suspeita que um ou mais regressores esteja relacionados com o efeito fixo, há duas formas usuais de se trabalhar com isso: (i) transformar os dados de modo a remover os efeitos fixos; ou (ii) instrumentalizar a variável endógena com variáveis que o pesquisador acredite que não estão correlacionadas com os efeitos fixos.

Um problema, porém, é que em um painel dinâmico, as formas usuais de transformação dos dados para remoção do efeito fixo não são capazes de remover a endogeneidade. Na transformação do estimador *within-group*, a variável dependente defasada continua correlacionada com o efeito fixo mesmo após a transformação. O

first-differenced GMM (FD-GMM) e GMM sistêmico (SGMM) surgem para lidar com esse tipo de caso. E agem tanto no sentido de transformar os dados quando utilizar instrumentos para corrigir o problema de endogeneidade em painéis dinâmicos.

A ideia básica do FD-GMM é escrever a equação de regressão como um modelo de dados de um painel dinâmico, para então tomar a primeira diferença com o intuito de remover a heterogeneidade não-observada e invariante no tempo, e então instrumentalizar as variáveis em primeiras diferenças do lado direito da equação usando os níveis das séries defasadas em dois ou mais períodos, sob o pressuposto de que os erros na equação original, em nível, não são serialmente correlacionados. Para explicitar esta técnica de estimação, considere um modelo AR (1) com um efeito fixo:

$$\begin{aligned} y_{i,t} &= ay_{i,t-1} + u_{i,t} & |a| < 1 & \quad (A.65) \\ u_{i,t} &= v_{i,t} + \eta_i \end{aligned}$$

Pressupostos:

$$\begin{aligned} E[\eta_i] &= 0 \\ E[v_{i,t}] &= 0 \\ E[v_{i,t}\eta_i] &= 0, \text{ para } i = 1, \dots, N \text{ e } t = 2, \dots, T \\ E[v_{i,t}v_{i,s}] &= 0, \text{ para } i = 1, \dots, N \text{ e } s \neq t \\ E[y_{i,t}v_{i,t}] &= 0, \text{ para } i = 1, \dots, N \text{ e } t = 2, \dots, T \end{aligned} \quad (A.66)$$

Sob esses pressupostos o GMM terá $m = 0.5(T-1)(T-2)$ restrições de momentos do tipo $E[y_{i,t-s}\Delta v_{i,t}] = 0$ para $t = 3, \dots, T$ e $s \geq 2$. Ou, escrito de maneira sustenta, $E[Z_i'\Delta v_{i,t}] = 0$ tal que

$$Z_i = \begin{bmatrix} y_{i,1} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & y_{i,1} & y_{i,2} & \dots & 0 & 0 & 0 \\ & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & y_{i,1} & \dots & y_{i,T-2} \end{bmatrix}$$

Apesar de o estimador ser, em teoria, consistente, este método é pouco eficaz caso os lags da variável, em nível, possuírem baixa correlação com suas subsequentes primeiras-diferenças, neste caso os instrumentos serão fracos. Esse tipo de problema também pode ocorrer quando o coeficiente autorregressivo, a , se aproxima de um ou quando $Var(\eta_i)$ aumenta relativamente à $Var(v_{i,t})$. Este viés irá subestimar o coeficiente, principalmente quando há poucos períodos de tempo. Uma forma de detectar esse tipo de viés consiste em comparar o FD-GMM com outras estimativas para a . Sabe-se que o Método dos Mínimos Quadrados Ordinários (MQO) irá superestimar a na presença de efeitos fixos, por sua vez, o estimador de efeitos fixos subestima o

mesmo. Sendo assim, temos um limite inferior e superior para comparar a razoabilidade da estimativa.

A5.3. GMM Sistemico

O GMM Sistemico (SGMM) é tido como um estimador com propriedades de amostras finitas superiores ao FD-GMM, sendo, portanto, um estimador mais apropriado para painéis dinâmicos. Basicamente a ideia aqui consiste adicionar uma nova hipótese e, ao invés de transformar os regressores para expurgar os efeitos fixos, a técnica consiste em transformar os instrumentos (via primeira diferença) de modo que estes sejam exógenos ao efeito fixo. Sendo assim, é possível combinar o modelo com variáveis em níveis (instrumentalizada com defasagens em primeiras diferenças) e o modelo com variáveis em primeiras diferenças (instrumentalizada com defasagens em nível) em um único sistema. Essa forma de estimação foi delineada por Bundell e Bond (1998).

Além das hipóteses do FD-GMM, a nova hipótese adicionada é:

$$E[\eta_i \Delta y_{i,2}] = 0, \text{ para } i = 1, \dots, N \quad (\text{A.67})$$

Para que ela seja válida, é necessário que o processo gerador de dados seja tal que os efeitos fixos e o processo autorregressivo governado por a se compensem ao longo de todo o painel. Combinando com as hipóteses do FD-GMM, ficamos com mais $T - 2$ restrições de momentos:

$$E[u_{i,t} \Delta y_{i,t-1}] = 0, \text{ para } i = 1, \dots, N \text{ e } t = 3, \dots, T \quad (\text{A.68})$$

Essas novas restrições permitem a utilização das defasagens em primeiras diferenças a série como instrumentos na equação em nível, como sugerido em Arellano e Bover (1995).

Como dito anteriormente, o SGMM usa um sistema empilhado de $T - 2$ equações em primeira diferença e mais $T - 2$ equações em nível, correspondendo aos períodos $t \geq 3$ disponíveis para os quais os instrumentos são observados. Assim, a matriz de instrumentos será escrita como:

$$Z_i^+ = \begin{bmatrix} Z_i & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i,2} & 0 & \dots & 0 \\ 0 & 0 & \Delta y_{i,3} & & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Delta y_{i,T-1} \end{bmatrix}$$

O conjunto inteiro de condições de momentos de segunda ordem disponíveis pela nova hipótese (X) é expresso por:

$$E \left[Z_i^+{}' u_i^+ \right] = 0$$

Tal que $u_i^+ = (\Delta v_{i,3}, \dots, \Delta v_{i,T}, u_{i,3}, \dots, u_{i,T})'$. Uma vez dentro deste arcabouço, é possível testar a validade dos instrumentos utilizando os testes de Sargan/Hansen e “*difference-in-Sargan/Hansen*”. Estes últimos são mais recomendados, visto que os testes usuais de Sargan e Hansen são fracos na presença de muitos instrumentos.

Podemos a forma de estimação para o caso em que há regressores endógenos. Considere o seguinte processo gerador de dados:

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \eta_i + v_{i,t} \quad (\text{A.69})$$

$$|\alpha| < 1$$

Para $i = 1, \dots, N$ e $t = 2, \dots, T$, tal que $x_{i,t}$ é correlacionado com η_i e também endógeno, $E[x_{i,t}v_{i,s}] \neq 0$, para $i = 1, \dots, N$ e $s \leq t$ (ou seja, os regressores são afetados tanto por choques contemporâneos quanto por *feedbacks* dos choques passados). Tomando a primeira diferença é possível eliminar o efeito fixo e obter as seguintes condições de momento:

$$E[x_{i,t-s}\Delta v_{i,t}] = 0 \quad (\text{A.70})$$

Para $t = 3, \dots, T$ e $s \geq 2$. Que se tornam disponíveis além daquelas já citadas no FD-GMM, com um AR(1). As defasagens das variáveis endógenas $x_{i,t}$ iguais ou anteriores à $t - 2$ podem ser usadas como instrumentos para a equação em primeira diferença. De modo similar, se $E[\eta_i\Delta x_{i,t}] = 0$ para $i = 1, \dots, N$ e $t = 2, \dots, T$, então os seguintes momentos estão disponíveis:

$$E[\Delta x_{i,t-1}u_{i,t}] = 0 \quad (\text{A.71})$$

Para $i = 1, \dots, N$ e $t = 3, \dots, T$. Além, é claro, dos momentos já citados no SGMM. Variáveis defasadas em primeira-diferença do regressor endógeno podem ser usadas como instrumentos para as equações em nível. É também válido ressaltar que a validade das hipóteses de ambos os modelos é necessário que a tendência temporal esteja devidamente controlada. Isto é, a regressão deve conter as *dummies* de tempo, do contrário as condições de momento não serão válidas caso haja variáveis não-estacionárias.

8. APÊNDICE B

B1. Modelo de Fatores Dinâmicos com os países do G7

B1.1. Com dados da PWT 8.0

Estimamos o modelo de fatores dinâmicos, porém dessa vez com dados de PTF dos países do G7. A ideia é verificar se os resultados se mantêm similares ao observado no modelo com os 10 países com maior índice de eficiência técnica. Assim como no outro, neste modelo conseguimos notar que os *loadings* estão inversamente relacionados com variáveis relacionadas com desenvolvimento técnico. Neste ponto, duas coisas chamam atenção: a ausência de correlação entre a PTF e os *loadings* e a correlação negativa fraca entre os *loadings* e as taxas de crescimento das PTFs. Esta última chama mais atenção por ser contra a intuição dada pelo modelo, visto que sinaliza que países com maior sensibilidade à tecnologia mundial teriam taxas de crescimento menores. Ainda assim, é sempre bom ressaltar que esse exercício é tão somente um esboço, visto o reduzido tamanho da amostra. Em todo caso, as demais correlações vão ao sentido esperado, mostrando que os resultados, em grande parte, se mantêm. Outra coisa que chama atenção, mas sem óbvias implicações, é que o fator comum deste modelo é estacionário.

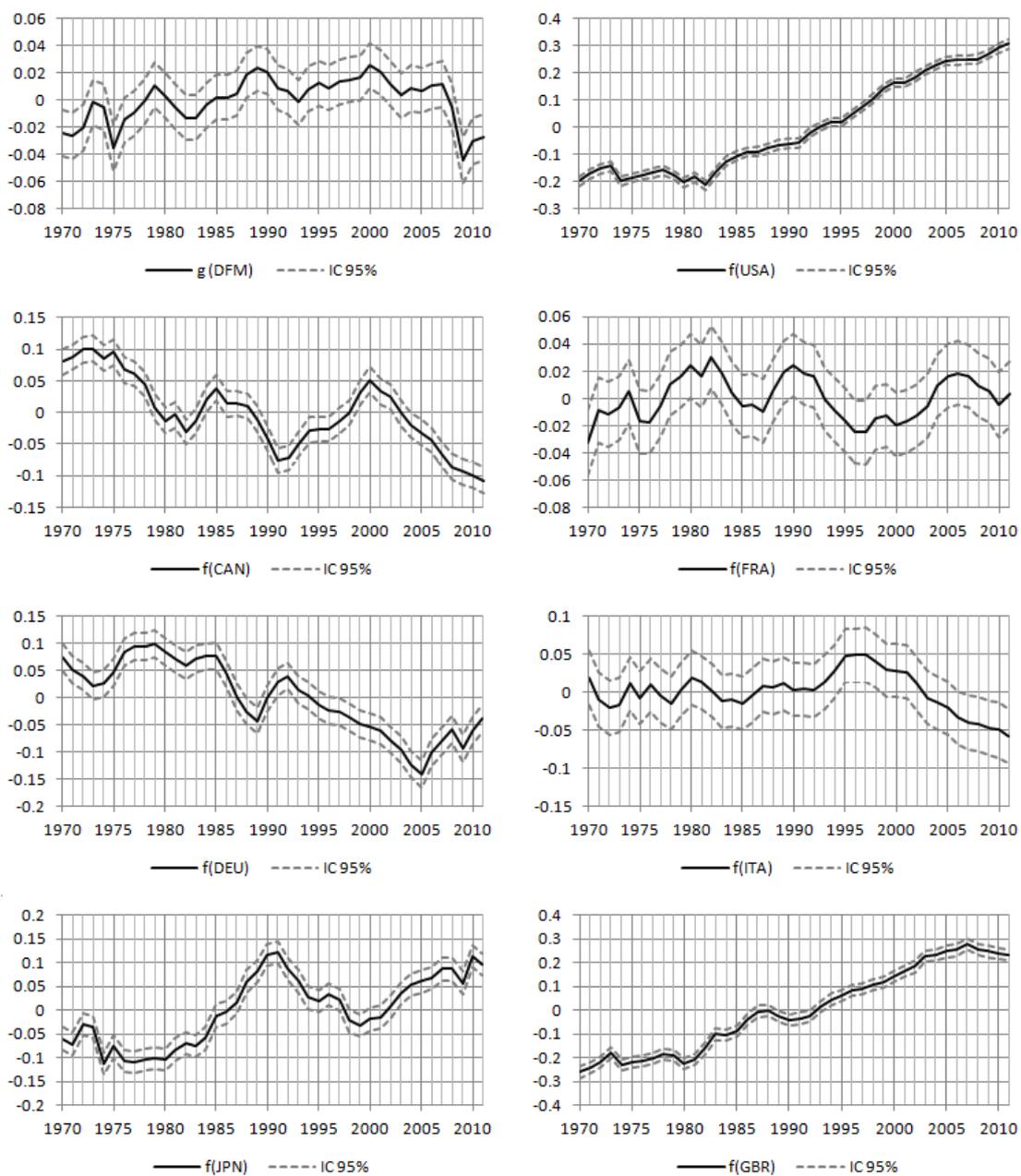
Tabela 16 - Resultados para o modelo de fatores dinâmicos (G7, PWT 8.0)

Modelo de Fatores Dinâmicos					
	ρ	λ_i^*	ϕ_i	σ_v^2	$\sigma_{u_i}^2$
g_t	0.8074419*** (0.1189949)	-	-	0.0001326* (0.0000919)	-
USA	-	1 (constrained)	0.995636*** (0.0060206)	-	0.0005344*** (0.0001242)
CAN	-	1.198553** (0.4657303)	0.9646631*** (0.0387766)	-	0.0003842*** (0.0000933)
FRA	-	1.400748*** (0.5009493)	0.833819*** (0.1105007)	-	0.0001257*** (0.0000406)
DEU	-	1.490341** (0.5875384)	0.9416524*** (0.0455918)	-	0.0004864*** (0.0001174)
ITA	-	2.115675*** (0.7659078)	0.9074424*** (0.0816714)	-	0.0002154*** (0.0000849)
JPN	-	1.412456** (0.5841262)	0.9313432*** (0.0522689)	-	0.0007397*** (0.0001743)
GBR	-	1.420897** (0.5643134)	0.9938579*** (0.008281)	-	0.0006793*** (0.0001605)
Amostra	1970-2011				

Erros-padrão em parêntesis; *** p<0.01, ** p<0.05, * p<0.1.

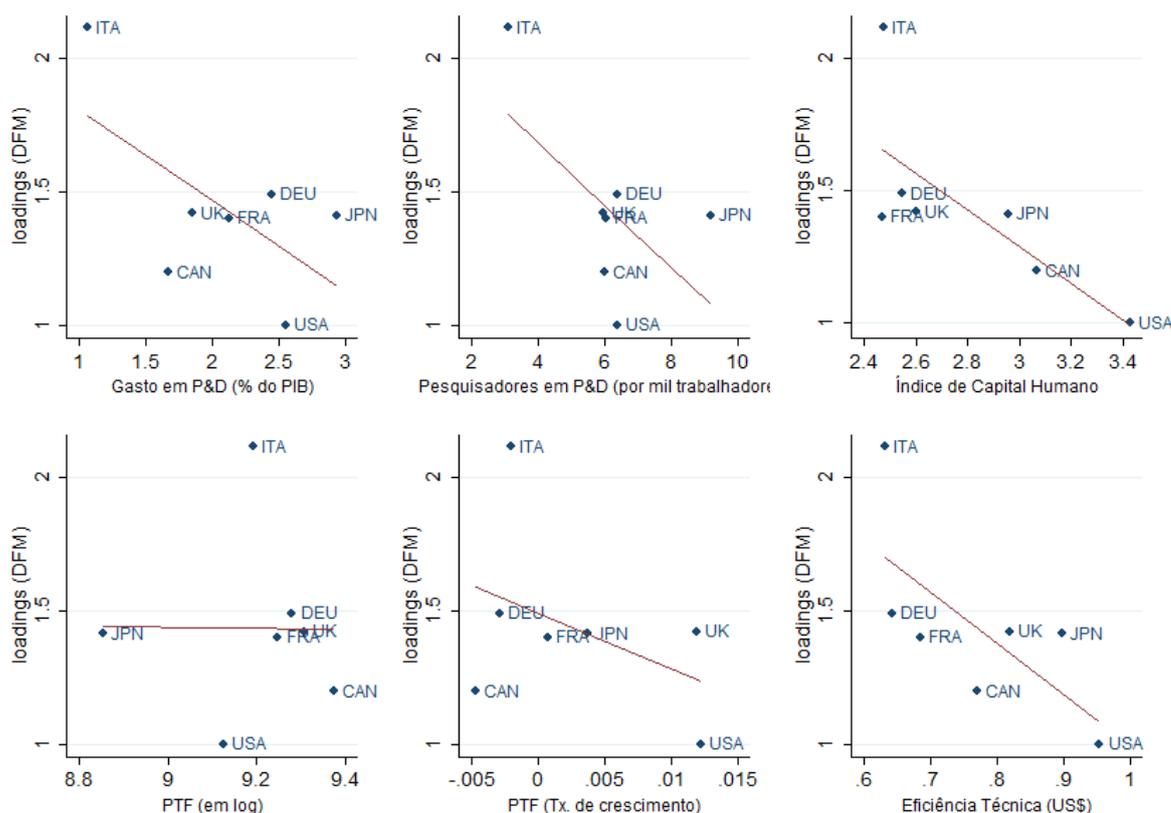
Fonte: cálculo dos autores.

Figura 9 – Fator comum e Fatores específicos (G7, PWT 8.0)



Fonte: cálculo dos autores.

Figura 10 - Correlação entre variáveis selecionadas e loadings (G7, PWT 8.0)



Fonte: cálculo dos autores e dados da OCDE, PWT 8.0 e Pires e Garcia (2011).

Tabela 17 - Decomposição da Variância (G7, PWT 8.0)

$$PVB = \hat{\beta}_i^2 Var(g_t) / Var(Y_i)$$

	USA	CAN	FRA	DEU	ITA	JPN	GBR
$\ln A_{i,t}^*$	5.75%	7.82%	71.87%	1.20%	83.31%	24.71%	14.27%

Fonte: cálculo dos autores.

B1.2. Com dados da PWT 9.0

É possível notar algumas mudanças significativas na mesma estimação para os dados da PWT 9.0. Primeiro, o fator comum passa a ser não-estacionário, e o coeficiente da PTF japonesa, assim como no modelo com os 10 países, não é significativo a 10%. Os *loadings* mudam, com magnitudes menores e com o Japão e Canadá tendo coeficientes menores do que os Estados Unidos. E a correlação entre esses *loadings* e os R^2 s passa a ser praticamente nula (0.0002). Apesar disso tudo, a correlação das variáveis selecionadas com os *loadings* permanece e fica bem mais fiel ao encontrado no modelo com 10 países. Agora a correlação com as taxas de crescimento é positiva e a correlação com o logaritmo das PTFs é negativo. Desse

modo, a principal intuição em relação aos *loadings* se mantém mesmo com um diferente grupo de países selecionado.

Tabela 18 - Resultados para o modelo de fatores dinâmicos (G7, PWT 9.0)

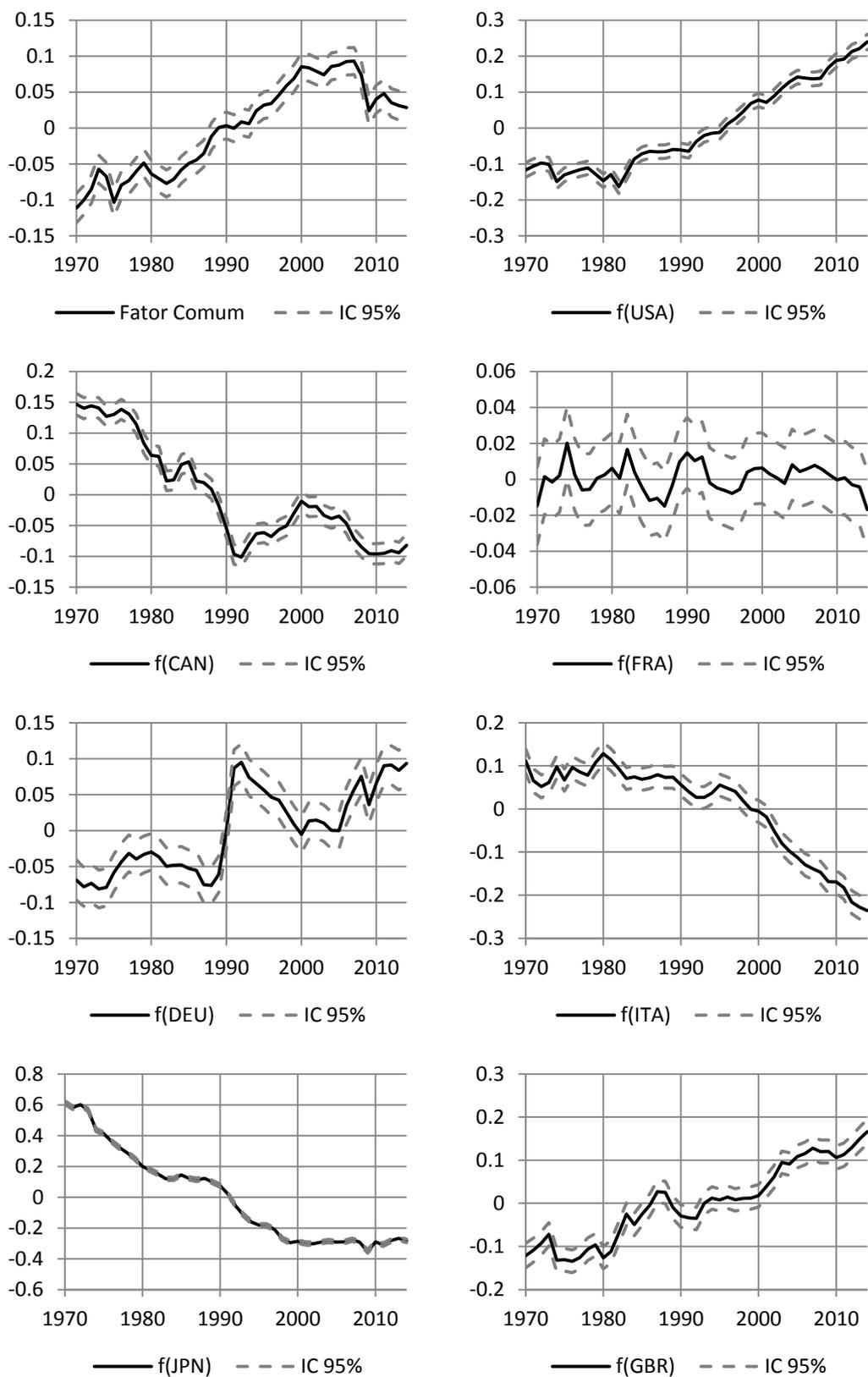
Modelo de Fatores Dinâmicos					
	ρ	λ_i^*	ϕ_i	σ_v^2	$\sigma_{u_i}^2$
g_t	0.9766731*** (0.0258495)	-	-	0.0002471** (0.000118)	-
USA	-	1 (constrained)	0.99413*** (0.008026)	-	0.0003652*** (0.0000946)
CAN	-	0.8554692*** (0.2401567)	0.9865174*** (0.0177029)	-	0.0003049** (0.0000739)
FRA	-	1.04779*** (0.2783765)	0.7360072*** (0.2398984)	-	0.0000819*** (0.0000382)
DEU	-	1.355369*** (0.3854276)	0.9407222*** (0.0599764)	-	0.0005062*** (0.000122)
ITA	-	1.359486*** (0.3854829)	1.027581*** (0.0323273)	-	0.0003975*** (0.0001057)
JPN	-	0.5753764 (0.3970744)	0.9965295*** (0.0048527)	-	0.0014523*** (0.0003116)
GBR	-	1.386884*** (0.3506306)	0.9839828*** (0.0207031)	-	0.0005001*** (0.0001307)
Amostra	1970-2014				
Erros-padrão em parêntesis; *** p<0.01, ** p<0.05, * p<0.1.					
Fonte: cálculo dos autores.					

Tabela 19 - Decomposição da Variância (G7, PWT 9.0)

$PVB = \hat{\beta}_i^2 Var(g_t) / Var(Y_i)$							
	USA	CAN	FRA	DEU	ITA	JPN	GBR
$\ln A_{i,t}^*$	85.8%	5.6%	98.6%	90%	3%	81.4%	91.5%

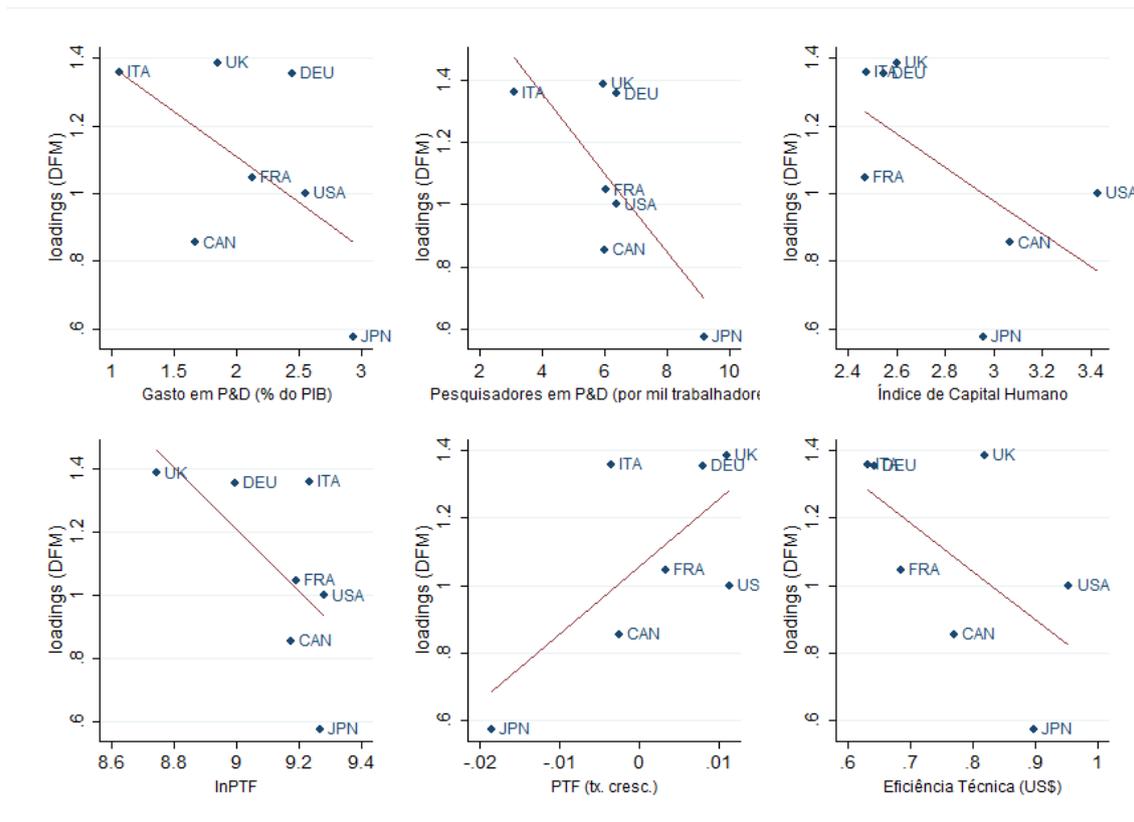
Fonte: cálculo dos autores.

Figura 11 – Fator comum e Fatores específicos (G7, PWT 9.0)



Fonte: cálculo dos autores.

Figura 12 - Correlação entre variáveis selecionadas e loadings (G7, PWT 9.0)



Fonte: cálculo dos autores e dados da OCDE, PWT 9.0 e Pires e Garcia (2011).