

Bruno Freitas Boynad de Vasconcelos

**Poder preditivo de métodos de Machine Learning com
processos de seleção de variáveis: uma aplicação às projeções
de produto de países**

Universidade de Brasília - UNB

2017

Bruno Freitas Boynad de Vasconcelos

Poder preditivo de métodos de Machine Learning com processos de seleção de variáveis: uma aplicação às projeções de produto de países

Tese apresentado ao curso de doutorado em Economia, como requisito para obtenção do título de Doutor em Economia, Faculdade de Economia, Administração e Contabilidade

Orientador: Victor Gomes

Universidade de Brasília - UNB
2017

Sumário

	Sumário	3
1	INTRODUÇÃO	5
1.1	Aprendizado de Máquina e Big Data	6
1.2	Especificação Econométrica	7
1.3	Capítulo 3: relação de causalidade entre produto e investimento	8
1.4	Capítulo 4: política fiscal, mercado financeiro e produtividade	8
1.5	Dados	9
1.6	Modelos de previsão de PIB como balizadores	9
2	APRENDIZADO DE MÁQUINA	11
2.1	Modelos lineares	12
2.1.1	Regressão linear	13
2.1.2	Métodos de seleção de sub-conjunto de preditores	13
2.1.3	Métodos de Shrinkage	14
2.1.3.1	Custo-benefício do viés e variância	15
2.1.3.2	Regressão ridge	16
2.1.3.3	Regressão Lasso	16
2.1.3.4	Regressão ridge & lasso	17
2.1.3.5	Elastic net e parâmetro de sintonia (λ)	18
2.1.4	Métodos de redução de dimensão dos dados	18
2.2	Modelos Não lineares	19
2.2.1	Redes neurais	20
2.2.2	Modelos Aditivos	21
2.2.3	Regressão em árvore	22
2.2.3.1	CHAID	24
2.2.4	Support vector machines	24
2.3	Seleção de modelos e Model Averaging	26
2.3.1	Critérios de informação	27
2.3.2	Amostra de validação	27
2.3.3	Model Averaging e Ensemble	29
3	A RELAÇÃO DE CAUSA EFEITO DO INVESTIMENTO E PRODUTO DENTRO DOS CANAIS DE TRANSMISSÃO DOS JUROS, POUPANÇA, INVESTIMENTO E PRODUTO	31
3.1	Introdução	31
3.2	Metodologia e Estimação	36
3.2.1	Dados, variáveis e amostra	36

3.2.2	Distribuição conjunta e simulações	38
3.2.3	Modelo estrutural e identificação	39
3.2.4	Fatoração, seleção de variáveis e clusterização dos betas	40
3.3	Resultados	43
3.3.1	Relação entre Investimento e Produto	43
3.3.2	Similaridades nas dinâmicas totais da economia	45
3.3.2.1	Análise de fatores	46
3.3.2.2	Análise de clusters	48
3.4	Conclusão	50
4	POLÍTICA FISCAL, MERCADO INTERNACIONAL E ANTECIPAÇÃO DAS FLUTUAÇÕES DO PRODUTO: UMA ABORDAGEM USANDO APRENDIZADO DE MÁQUINA	57
4.1	Introdução	57
4.2	Ciclos de negócios, crise de 2008 e política fiscal	60
4.3	Metodologia e estimação	62
4.3.1	Especificação econométrica	63
4.3.2	Dados	64
4.3.3	Variáveis Signals e Noise	67
4.3.4	Montagem de base, variáveis signals e validação cruzada	67
4.3.4.1	Montagem de base	68
4.3.4.1.1	Primeira abordagem: etapa exploratória	68
4.3.4.1.2	Segunda abordagem: determinantes do produto	68
4.3.4.2	Seleção de variáveis signals	70
4.3.4.3	Classificação por validação cruzada	71
4.4	Resultados	73
4.4.1	Eficiência preditiva com dados novos	73
4.4.2	Política fiscal e mercado financeiro	74
4.5	Conclusão	78
Referências		80
	Referências	81
5	APÊNDICE - DADOS	87

1 Introdução

O trabalho tem como objetivo testar hipóteses relacionadas ao comportamento das economias mundiais, nos últimos 30 anos, através de métodos de aprendizado de máquina / machine learning (ML) em evidência na literatura atual¹.

Em específico, a idéia central da tese é estender os artigos de Bajari et al. [2015a] e Bajari et al. [2015b], aplicados a micro dados em Organização Industrial, com metodologias de escolha de modelo e seleção de variáveis utilizando base de dados de painel de países. Essa aplicação foi explorada por Varian [2014] com o objetivo de estudar essas técnicas de ML. O estudo das forças motoras dos movimentos do produto sempre é um tema de interesse dos economistas e estas bases de dados possibilitam exercícios e análises atrativas. Nesse sentido, também podemos mencionar que a tese se assemelha a trabalhos que usam painel de países para estudar métodos de fronteiras (Mello and Perrelli [2003] e Sala-I-Martin et al. [2004]).

O trabalho está estruturado de forma a abordar as hipóteses, sobre o comportamento das economias mundiais, em duas análises de painel de países, cada qual em um capítulo específico. Para tanto, essas hipóteses estão distribuídas sob duas perguntas principais que serão tratadas e desenvolvidas cada qual em um capítulo específico. No capítulo 2 vamos fazer uma revisão teórica dos métodos relevantes para a tese. No capítulo 3 fazemos nossa primeira análise do comportamento das economias mundiais baseada nas idiosincrasias dos países. No capítulo 4 apresentamos um trabalho baseado na proposta de um procedimento de estimação que resume um volume grande de informação. Apesar de diferentes, os trabalhos dos capítulos 3 e 4 chegam a resultados similares de que a característica (ou grupo de informação) mais relevante para os ciclos de negócios do produto é como o país interage com o mercado internacional.

No que se refere à análise econômica, a primeira pergunta, tratada no capítulo 3, é a relação de causalidade entre investimento e produto. Para esta, vamos usar uma abordagem mais simples para tratar do problema. Neste trabalho destacamos idiosincrasias dos países com sistemas de equações individuais e nossa metodologia está focada na estrutura de dados.

A segunda pergunta é o papel da política fiscal, do mercado financeiro e da produtividade da mão de obra para antecipar as flutuações do produto no período anterior, durante e na recuperação mundial da crise de 2008. Para abordar essa questão vamos usar técnicas de ML relacionadas diretamente com a fronteira dessa literatura. Nessa análise do capítulo 4, vamos desenvolver uma abordagem baseada no estimador de dois estágios de Meinshausen [2007], onde estendemos o processo de seleção de variáveis signals (primeiro estágio) e a escolha do modelo final (segundo estágio). Para avaliar os resultados da nossa metodologia utilizamos estimadores em evidência na literatura de ML destacados por Fernández-Delgado et al. [2014] (SVM, deep learning e random forest).

As principais contribuições dessa tese são divididas em dois grupos. O primeiro reúne as contribuições em termos de métodos para identificação de variáveis signals dentro de um universo significativo de

¹ Dessa forma, primeiramente, é importante destacar que por métodos de ML não nos referimos a estimadores. Na verdade, tratamos do assunto de forma mais abrangente e por ML nos referimos ao raciocínio orientado aos dados (*data-driven*), voltado principalmente à previsão, na tomada de decisão.

preditores². O segundo grupo de contribuições são relacionadas (i) a relação entre produto, investimento, poupança e preço do investimento e da poupança e (ii) as análises das forças motoras do produto no período antes, durante e após a crise de 2008.

No que se refere ao primeiro grupo de contribuições, nossa abordagem desenvolvida no capítulo 4, apresentou resultados melhores dentre outras consideradas e destacadas na literatura ao mesmo tempo em que possui vantagens interpretativas (derivação de efeito marginal) e de aplicação a grandes bases de dados.

Por sua vez, sobre o segundo grupo de contribuições, tanto o trabalho do capítulo 3, como o do capítulo 4, usando análises, base de dados e estruturas bem diferentes, chegam a conclusões similares. Quando procuramos entender como as economias internacionais se diferem ou se assemelham, nossos resultados indicam que a principal característica (ou grupo de características) que tem efeito sobre o produto é como a economia interage com o mercado internacional. O que chamamos de interação com o mercado internacional inclui questões como por exemplo, como a economia acomoda ou reage a choques do mercado internacional, como esses choques se propagam pela economia, como a economia compete no mercado internacional e como absorve know-how e tecnologia para alterar sua produtividade interna.

1.1 Aprendizado de Máquina e Big Data

As motivações do trabalho estão relacionados à forma diferente de tratar estimações, aos benefícios relacionados a previsão com dados novos e à soluções aplicadas à situação de grandes amostras / “big data” que são frequentes atualmente no ambiente dos economistas e em específico do Banco Central do Brasil (BCB).

O aprendizado de máquina tem ganhado atenção especial dos economistas nos últimos anos no meio acadêmico em economia (Varian [2014], Bajari et al. [2015b] e Bajari et al. [2015a])³. Destacamos nessa tese três motivos principais para esse aumento de interesse. O principal é que o ML tem uma abordagem orientada aos dados com soluções eficientes quando o objetivo é previsão *out of sample*⁴. Em segundo, essa abordagem permite um elevado grau de automatização do processo de modelagem, estimação, teste e previsão. Por fim, o campo de estudo da ML trata diretamente de problemas de “big data”.

A discussão de ML e desempenho preditivo *out of sample*, passa necessariamente pelo tópico do *bias-variance trade-off* (Hastie et al. [2009]). De uma forma geral, o custo-benefício do viés e variância ocorre quando é possível reduzir as incertezas das previsões e projeções ao custo de um viés nos estimadores. O tema de desempenho preditivo tem um apelo empírico muito significativo para instituições que tratam de projeções e previsão como bancos centrais.

O segundo motivo também merece destaque porque a possibilidade de automatizar uma parcela relevante do processo traz muitos benefícios em áreas que tratam de previsão e projeção. Além das questões relacionadas às estimações e testes, existem também os desafios de grandes bases de dados (big data). Esse ponto conecta o segundo motivo ao terceiro motivo.

² Definiremos variáveis “signals” a seguir nessa introdução.

³ Em outras áreas relacionadas a economia como finanças, negócios e marketing, essas técnicas orientadas a dados já são usadas com maior frequência a mais tempo (Khandani et al. [2010], Peña et al. [2011], Plakandaras et al. [2013] e Bontempi et al. [2013]).

⁴ Por abordagem orientada aos dados queremos dizer que a modelagem (especificação, adequação, avaliação, etc.) é direcionada e definida pelo feedback dos dados e não por teoria por exemplo.

Por sua vez, essas situações de "big data" estão presentes no cotidiano dos economistas com uma frequência cada vez maior nas últimas décadas. O "big data" pode ser caracterizado por grandes bases de dados com um número elevado de preditores relativos às observações disponíveis. Nesses casos dois problemas relacionados a modelagem se destacam. O primeiro é a insuficiência de graus de liberdade para estimação de métodos frequentistas. O segundo é que em bases com muitos preditores existe uma probabilidade significativa de haver multicolinearidade elevada ou perfeita entre esses preditores. E isso pode ocorrer independente da definição conceitual das variáveis, pois pode simplesmente haver coincidência numérica.

Assim, diante dos três motivos apresentados para o ganho de popularidade do ML junto aos economistas no meio acadêmico, podemos destacar três problemas que motivaram a tese.

Primeiro que os métodos e abordagens usados em econometria tradicionalmente para previsão e projeção podem não estar entre os mais eficientes disponíveis. Esse problema é o alvo principal da ML que busca melhores formas de fazer projeções através de abordagem orientada aos dados.

Segundo que diante de bases de dados cada vez maiores e mais complexas, a tarefa de testar e escolher modelos de forma otimizada é custosa. A questão de modelar de forma otimizada é importante porque devido ao volume de informação (dados) disponível é possível encontrar, dentre as opções factíveis, um conjunto de modelos que podem ser classificados como "melhores" relativos aos demais. Dessa forma, por tratar de abordagens orientadas aos dados, processos baseados nas ferramentas e lógicas de ML como a modelagem otimizada podem ser automatizados, o que reflete em mais eficiência para processar informações disponíveis, analisar o espaço de possibilidades e em conseqüentemente melhores resultados preditivos.

Terceiro que esse grande volume de dados oferece desafios para definição de critérios de seleção de modelos, variáveis e manipulação dessas informações. A ML, apesar de não tratar desse tipo de problema diretamente nos seus objetivos, fornece ferramentas que podem ser usadas nessas situações.

1.2 Especificação Econométrica

Para explicar as perguntas propostas vamos utilizar bases de dados e métodos distintos que apresentam problemas metodológicos específicos que por sua vez levam a soluções particulares em termos metodológicos.

No capítulo 3 vamos apresentar um ambiente onde nosso objetivo é estudar sistemas de equações para cada país individualmente ao mesmo tempo em que temos restrições de graus de liberdade. Ao contrário de uma abordagem de painel onde estimamos o efeito médio, aqui procuramos evidenciar as particularidades de cada país. Essa é uma abordagem mais simples e ingênua. Os algoritmos utilizados são classificados na literatura de ML e computação como algoritmos "burros" e/ou "greed". É um trabalho que usa uma visão orientada a dados, mas que não está relacionada com as melhores soluções de ML.

No capítulo 4 vamos utilizar abordagens que procuram corrigir fragilidades da metodologia da análise 1 principalmente no que se refere à abordagem ingênua e otimização dos códigos. Vamos utilizar estimações em painel e abordagens de ML propriamente dita.

1.3 Capítulo 3: relação de causalidade entre produto e investimento

Para destacar as particularidades de cada país e estudar os canais de transmissão da economia envolvidos no objetivo do capítulo 3, vamos estimar sistemas de equações com transmissão contemporânea por país (equação (1.1)).

$$A_c Y_{c,t} = B_c X_{c,t} + \varepsilon_{c,t} \quad (1.1)$$

A matriz A_c (7×7) é triangular superior e faz com que o sistema seja exatamente identificado. $Y_{c,t}$ é o vetor de variáveis endógenas do país c no tempo t , $X_{c,t}$ é o vetor de variáveis exógenas desse mesmo país no tempo t e B_c tem dimensão 4×7 (4 variáveis exógenas em 7 equações de variáveis endógenas).

Assim, ao contrário da análise do capítulo 4, que estima uma equação em painel, aqui vamos estimar modelos por país (A_c e B_c variam de país para país).

Na verdade a matriz A_c vai ser estimada para diversas especificações mantendo o sistema da equação (1.1) identificado. A idéia é escolher a melhor especificação para cada país. Com base nisso vamos tentar identificar comportamento e estrutura mais usuais, ou agrupamento de estruturas, que sustentem teorias que relacionam produto, investimento, poupança e preço do investimento e poupança. Em específico, a pergunta principal feita aqui é: são mais frequentes as evidências de causalidade de produto para investimento ou vice versa?

1.4 Capítulo 4: política fiscal, mercado financeiro e produtividade

Para abordar o assunto, dado um painel de países com muitos preditores em relação ao número de observações, o trabalho começa com uma equação em painel que explica a variação do produto dos países dada pela equação (1.2).

$$\Delta y_{c,t} = \alpha \Delta y_{c,t-1} + \beta X_{c,t-1} + \delta Z_t + \varepsilon_{c,t} \quad (1.2)$$

Onde $y_{c,t}$ é a variação do produto do país c no tempo t e $\Delta y_{c,t}$ é a primeira diferença de $y_{c,t}$. $X_{c,t}$ é uma matriz de variáveis específicas do país c que incluem observações tanto no tempo t como em defasagens. Essas variáveis incluem dados de contas nacionais, preços, mercado financeiro, risco, etc. Z_t é uma matriz que incluem variáveis constantes no cross section dos países tanto no período t como defasadas. Essas variáveis incluem variáveis mundiais, como medidas de aversão ao risco internacional e observações específicas de países que afetam o painel todo (como PIB da China, ou crédito nos EUA, etc), ou PCA em todas as variáveis.

Na verdade, a equação (1.2) é uma especificação genérica e não necessariamente será utilizada com essa estrutura. O capítulo 4 vai desenvolver essa discussão.

A idéia principal aqui é, para um número grande de testes, verificar se a política fiscal, o mercado financeiro e a produtividade da mão de obra (medidas de diferentes formas) foram relevantes (seja por si próprias ou por modificar o comportamento de outra variável) para explicar as variações do produto em um período onde essas questões estiveram em evidência na economia internacional e no Brasil. Assim, as perguntas principais a serem feitas são: podemos argumentar que a política fiscal, o mercado financeiro

ou a produtividade da mão de obra foram importantes forças motoras do produto no período da Grande Recessão? Podemos usar as forças motoras desse período para antecipar flutuações do produto?

1.5 Dados

Para a aplicação prática vamos usar uma base de dados econômicos e financeiros de países com periodicidade anual ou trimestral dependendo da análise.

Essas bases combinam dados de domínio público (Banco Mundial, OCDE, BIS, dentre outros) e dados de acesso mais restrito (Bloomberg). As bases utilizadas são diferenciadas para o estudo dos capítulos 3 e 4.

Para o estudo da relação de causalidade entre produto e investimento o período de treinamento se inicia em 1990 e vai até 2012. É uma base anual mais longa possível que podemos montar para incluir países interessantes a essa análise (países do bloco comunista não tinham dados antes de 1990).

Para o estudo do papel da política fiscal, mercado financeiro e produtividade, o período de treinamento básico começa em 2001 e vai até 2015 (dependendo do caso a série pode ser mais extensa e começar antes). É uma base trimestral mais curta que a anterior, pois possui como objetivo estudar características do período e está sujeita a limitação de dados para observações anteriores a 2003.

Esse é um período muito interessante para análise pois contempla informação da economia mundial antes da crise de 2008, durante e da recuperação desta crise.

Em termos teóricos o período é interessante para a proposta do capítulo 4 por 3 motivos principais. Primeiro por causa da disponibilidade de dados (quantidade e qualidade) que permite avaliar muitas dimensões dos eventos. Segundo que foi um período com muitos eventos de stress, em termos de situações ou comportamentos não usuais, limites e políticas. Terceiro que, além de situações de stress, também foi um período rico em políticas e eventos.

1.6 Modelos de previsão de PIB como balizadores

As discussões desse tópico estão mais relacionadas ao capítulo 4, no entanto, optamos por trazê-la para o início para melhor contextualizar o leitor e/ou esclarecer algumas questões que este pode levantar.

O trabalho tem como principal variável de aprendizado supervisionado o produto dos países⁵. Com isso, queremos dizer que o produto é uma variável a ser usada como medida de desempenho dos procedimentos desenvolvidos e conseqüentemente testar as hipóteses e argumentos do trabalho. Essa visão é diferente de “prever PIB”.

No entanto, como os conceitos não são mutuamente excludentes, uma preocupação relevante é balizar as estimações e procedimentos com os melhores modelos disponíveis em termos de previsão de produto. Por sua vez, apesar de reconhecer a conexão entre assuntos, é importante destacar que nesse aspecto, o trabalho não tem como objetivo competir com os melhores modelos de previsão de produto do mercado.

Tendo isso em mente, para identificar alguns dos melhores modelos do mercado, em um primeiro passo, devemos identificar os interessados no assunto. Fora do meio acadêmico, o assunto de previsão de PIB é especialmente interessante para instituições como FMI e Bancos Centrais. Dentre os métodos

⁵ Produto, PIB ou GDP se referem a mesma variável nesse trabalho.

utilizados por estes órgãos vale destacar o Multmod 3 (FMI - Laxton et al. [1998]), NiGEM (OCDE, BoE e BCE - NiGEM [2016]) e equações bridge (Bayoumi et al. [2004]). No entanto, estes modelos estão fora do escopo desse trabalho seja por serem fechados, ou muito complexos para figurar um papel secundário no trabalho, ou demandarem bases de dados específicas.

Outra linha de trabalho que está relacionada ao tema é a literatura de ciclo de negócios que trata das forças motoras do produto de curto prazo (Neumeyer and Perri [2005], Aguiar and Gopinath [2007], Smets and Wouters [2007] e Chang and Fernandez [2013]). Mais uma vez, estes trabalhos não serão explorados nessa tese por dois motivos. Primeiro que não possuem como objetivo principal o desempenho preditivo. Segundo que estendem essas discussões também para tópicos também fora da proposta e objetivos desse trabalho.

Assim, como esses modelos, em geral, fogem do escopo dessa tese, vamos utilizar 2 grupos de abordagens como referência para o estimador proposto no capítulo 4⁶. A primeira é mais tradicional e baseada em Bayoumi et al. [2004]. O segundo grupo usa estimadores de ML em evidência na literatura e destacados em Fernández-Delgado et al. [2014]. Essas abordagens serão discutidas no capítulo 4 que trata especificamente desse assunto.

⁶ Na verdade, no capítulo 4 falamos em 3 grupos de balizadores porque 2 desses grupos estão relacionados ao trabalho de Bayoumi et al. [2004].

2 Aprendizado de Máquina

Nesse capítulo, vamos fazer uma revisão teórica das metodologias relacionadas com a tese. Lembramos que o foco é o ML e por ML queremos destacar o raciocínio de metodologias orientadas aos dados.

Dessa forma, eventualmente vamos usar algum método que não é ligado a essa literatura. Isso é importante ressaltar porque quando falamos em ML e métodos orientados a dados não estamos restringindo a um conjunto limitado de metodologias estatísticas e computacionais. Na verdade estamos nos referindo a uma forma de raciocínio que pode usar diferentes métodos para atingir seus objetivos.

Podemos pensar a ML como um ramo da estatística e da computação que reúne uma série de métodos que possuem dois objetivos principais. O primeiro é o desempenho preditivo de modelos e o segundo é automatizar o processo de modelagem das bases de dados observados ou aprendizado com os dados observados. Ademais, boa parte da ML envolve métodos de calcular o erro de validação (erro de previsão *out of sample*), e selecionar ou ponderar modelos com base nesses erros de previsão.

No que se refere ao desempenho preditivo dos modelos, a acurácia das estimativas ou previsões das estimativas ou previsões é um aspecto importante. Nesse sentido, uma parte importante da literatura de ML se dedica a reduzir a variância das estimativas ou a parcela redutível da variância das estimativas. A parcela não redutível da variância não pode ser trabalhada, por exemplo, devido à omissão de variáveis do modelo. Para visualizar essa decomposição, seja a variância dos erros de previsão dadas por $E(Y - \hat{Y})$ de acordo com a equação (2.1).

$$E(Y - \hat{Y})^2 = E[f(X) + \epsilon - \hat{f}(X)]^2$$
$$E(Y - \hat{Y})^2 = [f(X) - \hat{f}(X)]^2 + var(\epsilon) \quad (2.1)$$

Onde \hat{Y} é a estimativa de Y dado pelo modelo estimado $\hat{f}(X)$ que é função dos dados observados X . O termo $(Y - \hat{Y}) * \epsilon = 0$ por hipótese. O termo $[f(X) - \hat{f}(X)]^2$ é a parcela redutível do erro e $var(\epsilon)$ é a não redutível.

Os métodos de ML, ou métodos estatísticos, podem ser divididos em aprendizado supervisionado e aprendizado não supervisionado. Learning supervisionado, que é foco desse trabalho, reúne métodos de estimação onde se tem pelo menos uma variável dependente explicada, ou variável de resposta, que é explicada. Isso quer dizer que o que se está querendo modelar é objetivo e mensurável e pode ser previsto ou projetável. O aprendizado supervisionado envolve problemas de classificação, estimação da distribuição a posterior ($P(y|X)$), interpolação (métodos não lineares) e regressão linear. Por outro lado o aprendizado não supervisionado estuda métodos onde não se tem uma variável objetivo ou variável de resposta que é usada para supervisionar o modelo, como análise de cluster por exemplo.

Assim, dividimos os métodos de estimação em 3 grupos, métodos lineares, não lineares e bayesianos. Os dois primeiros estão relacionados à literatura de ML, o último não. Por fim, vamos tratar da avaliação de modelos e de como combinar modelos menores para reduzir as incertezas das previsões usando dados novos.

2.1 Modelos lineares

Os modelos lineares tem esse nome porque assumem que a função que define $E(Y|X)$ é linear nos preditores X_1, X_2, \dots, X_p . Três aspectos importantes com relação a esses modelos merecem destaque. Primeiro que, apesar do nome, esses modelos permitem que sejam aplicadas transformações não lineares nos preditores preservando a relação linear na função. Com isso, os métodos lineares possuem uma flexibilidade significativa.

Segundo que mesmo que essa aproximação linear seja uma visão simplista da maioria dos problemas reais, os métodos lineares, como os mínimos quadrados ordinários (MQO), possuem, na prática, um desempenho muito competitivo quando comparado a modelos não lineares mais sofisticados.

Terceiro que a estrutura linear da(s) equação(ões) permite que a relação de causalidade entre as variáveis seja vista de forma direta. Isso simplifica a interpretação e as inferências no modelos (James et al. [2013]).

De uma maneira geral os métodos lineares definem, ou estimam, uma função $E(Y|X)$ como uma combinação linear de preditores¹.

$$E(Y|X) = \beta_0 + \sum_{j=1}^P X_j \beta_j$$

Dentre os métodos lineares, a abordagem mais comum é o MQO. É um método simples e em muitos casos com eficiência satisfatória.

No entanto, apesar de suas vantagens, a literatura de ML destaca dois motivos principais para se buscar modelos mais elaborados que as estimativas de MQO. O primeiro se refere a *acurácia das previsões* e o segundo a *melhor interpretabilidade do modelo*.²

Assim, o MQO é uma foma referência para comparar abordagens e métodos de estimação.

O problema da acurácia das previsões está relacionado com o fato de que, no método de MQO, para um elevado número de parâmetros relativo ao tamanho da amostra, a variância das estimativas do modelo é elevada. Se o número de parâmetros for superior ao tamanho da amostra, então existem infinitas soluções e o MQO não pode ser usado. Assim, para reduzir a variância das estimativas de mínimos quadrados a literatura sugere o uso de métodos que aplicam penalidades nos coeficientes dos estimadores (métodos de *shrinkage*). O *trade off* desses modelos é que se diminui a variância das previsões ao custo de um viés nos parâmetros estimados.

O segundo motivo para se buscar métodos diferentes do MQO se refere às limitações desse método em selecionar as variáveis predictoras realmente importantes no modelo. De fato, o MQO não tem um critério direto para seleção de um subconjunto de preditores dentro de um conjunto maior. Nesse sentido, para tratar desse problema, a literatura de aprendizado de máquina sugere métodos que automatizam esse processo de seleção.

¹ Além da questão das transformações não lineares nos preditores que confere flexibilidade aos métodos lineares, tal como mencionado, outra forma de aumentar a flexibilidade desses modelos é relaxar essa estrutura aditiva e trabalhar com interações de preditores.

² Note que esse raciocínio se aplica principalmente quando temos como objetivos mais importantes a capacidade preditiva do modelo *out of sample* e a automatização do processo de aprendizado, que tal como comentamos, são os objetivos principais do aprendizado de máquina.

2.1.1 Regressão linear

O método de regressão linear envolve os métodos de mínimos quadrados generalizados (GLS ou MQG) e os métodos de variáveis instrumentais³. Esses estimadores são obtidos minimizando uma função perda que é igual à soma do quadrado dos resíduos.

$$\beta_{linear} = \underset{\beta}{\operatorname{argmin}} \left\{ (Y - X\beta)^T (Y - X\beta) \right\},$$

Na qual X é uma matriz de preditores com a primeira coluna formado por 1s.

O MQO é um método de regressão linear abrangente que envolve vários casos específicos. A solução do problema de MQO na forma matricial é mostrada na equação (2.2).

$$\beta_{glz} = \left(X' \Omega^{-1} X \right)^{-1} X' \Omega y, \quad (2.2)$$

Na qual Ω no caso do MQO é uma matriz identidade.

O estimador de variáveis instrumentais é um estimador de dois estágios que pode ser escrito diretamente como na equação (2.3).

$$\beta_{iv} = \left(X' Z (Z' Z)^{-1} Z' X \right)^{-1} X' Z (Z' Z)^{-1} Z' y, \quad (2.3)$$

Na qual Z é uma matriz de instrumentos.

Dentre esses, o mais usado é o MQO, que é um caso específico do MQG. O MQO tem benefícios bem conhecidos na literatura pois, segundo o teorema de Gauss-Markov, dentre os estimadores não viesados lineares, o MQO é o que apresenta a menor variância.

Além dessa característica, o MQO tem vantagens em termos de simplicidade e facilidade para se fazer inferências. Por fim vale destacar que em testes empíricos, apesar da simplicidade, o MQO tem um desempenho preditivo muito bom quando comparado a modelos mais sofisticados.

No entanto, o MQO, em sua estrutura simples, apresenta limitações. Kuhn and Johnson [2013] aponta quatro limitações principais dos modelos de mínimos quadrados. Essas limitações são situações em que os estimadores de mínimos quadrados são instáveis ou não são os mais adequados. A primeira é situações em que se tem a presença de elevada multicolinearidade nos dados (preditores). A segunda é que a solução desses modelos é uma aproximação linear que em muitos casos não representa adequadamente o espaço estudado. Terceiro é que os estimadores são muito sensíveis a *outliers*. Quarto que ele não pode ser usado em situações onde se tem mais preditores que observações.

2.1.2 Métodos de seleção de sub-conjunto de preditores

Para a primeira limitação do MQO importante para o aprendizado de máquina, vamos assumir que o conjunto de preditores de uma função $E(Y|X)$, dado pela matriz X de dimensão $n \times p$ (onde n é o

³ Na verdade, a literatura de aprendizado de máquina geralmente se refere a apenas aos mínimos quadrados (MQO) nos métodos de regressão lineares. No entanto, aqui vamos usar um conceito mais abrangente e incluir os métodos de regressão lineares resolvidos por solução linear de matrizes. Isso porque assim cobrimos os principais problemas do MQO, como erros não esféricos e endogeneidade dos preditores. O problema de colinearidade significativa dos preditores é tratada aqui quando debatemos processos de seleção de variáveis.

número de observações e p o número de preditores), seja grande. Isso que dizer que p é um número elevado quando comparado a n mas não necessariamente $p > n$).

Um primeiro método para tratar do problema é o **best-subset selection** que consiste em procurar o subconjunto de preditores que maximiza um critério de ajustamento (*fit*) do modelo. Esse critério de ajustamento do modelo pode ser tanto *in sample* como *out of sample* tal como discutido em detalhes no tópico sobre seleção de modelos e model averaging. No entanto o método de *best sample selection* é custoso computacionalmente para um número elevado de preditores. Um algoritmo clássico é o *leaps and bounds*.

Um método muito conhecido é o **stepwise selection** que escolhe as variáveis usando um processo iterativo baseado no incremento do ajustamento (*fit*) do modelo. O *forward stepwise selection* começa com o intercepto e a cada iteração adiciona a variável, dentro do conjunto total de preditores, que tem maior correlação com o resíduo, ou tem maior impacto positivo no ajustamento do modelo. O *backward selection* começa com o modelo com todos os preditores e apaga, de forma iterativa, a variável que tem menor impacto no ajustamento do modelo. Existem também métodos de *hybrid stepwise selection* que combinam os critérios de seleção de *backward* e *forward* em cada iteração.

Outro método relacionado é o **forward-stagewise regression**. Assim como o *forward stepwise selection*, o *forward stagewise selection* inicia com o intercepto e com os coeficientes de todas as variáveis iguais a 0. Em cada iteração selecionamos a variável com maior correlação com resíduo e regredimos o resíduo contra essa variável. O coeficiente dessa variável na equação principal é atualizado por um pequeno ϵ . O procedimento continua até que não existam variáveis correlacionadas com o resíduo.

2.1.3 Métodos de Shrinkage

Os métodos de seleção de subconjunto de preditores não tratam do problema relacionado à elevada variância relativa das previsões e estimação do MQO. Para abordar o assunto a literatura de ML indica os métodos de *shrinkage* que, segundo Hastie et al. [2009], são mais contínuos e não apresentam elevada variabilidade. Estes tipos de regressão fazem parte da classe de estimadores de regressão com penalidade.

A penalidade, ou restrição, existe para limitar a variância dos estimadores. Assim, o ponto central aqui é o *trade-off* entre variância das estimativas do modelo e viés dos coeficientes estimados não viesados. De fato, o MQO gera os estimadores não viesados com menor variância dentro do conjunto de estimadores lineares. No entanto, é possível obter modelos com variância menor ao custo de acrescentar viés aos estimadores. Segundo Kuhn and Johnson [2013], com um pequeno viés nos estimadores é possível diminuir substancialmente a variância do modelo.

Ao contrário do método de MQO, os parâmetros estimados pelos métodos de *shrinkage* não são invariantes ao reescalonamento das variáveis. Na verdade eles podem variar significativamente com esse reescalonamento. Isso porque, o reescalonamento das variáveis tem efeito direto na restrição do problema. Dessa forma, daqui em diante, vamos assumir que as variáveis de Y e X foram padronizadas ($Y, X \sim (0, 1)$) pois caso contrário os resultados ficam comprometidos.

O problema da regressão com penalidade para a *elastic net* (Zou and Hastie [2005]):

$$\beta_{en} = \underset{\beta}{\operatorname{argmin}} \left\{ (Y - X\beta)^T (Y - X\beta) + \lambda \left(\beta^T \beta (1 - \alpha) + \alpha |\beta| \right) \right\},$$

Onde X é uma matriz de preditores normalizados sem uma coluna de 1s para estimação de intercepto. A *elastic net* é um método de regressão que usa como restrição uma combinação linear entre a restrição L_2 da ridge e a L_1 da lasso. Dessa forma, a *elastic net* tem três outros métodos como casos especiais.

Se não houver restrição ($\lambda = 0$), então estamos em um problema de minimizar uma função perda dada pela soma dos quadrados dos resíduos $((Y - \beta_0 - X\beta)^T (Y - \beta_0 - X\beta))$ que é o problema do MQO.

Se a restrição for ativa ($\lambda \neq 0$) e $\alpha = 0$, então temos apenas a restrição L_2 (quadrática - $\beta^T \beta$) que define a regressão de ridge. Se por outro lado, $\alpha = 1$, então temos apenas a restrição L_1 e estamos no problema da regressão lasso (*least absolute shrinkage and selection operator*).

A restrição da regressão ridge é uma restrição de norma 2 (l_2) e aplica uma penalidade relativamente homogênea a todos os coeficientes do modelo. No caso da regressão ridge, a idéia da penalidade pela soma dos quadrados dos parâmetros é a mesma usada nas redes neurais onde essa é chamada de *weight decay*. A restrição da regressão lasso é uma restrição de norma 1 (l_1) e tende a destacar os coeficientes de algumas variáveis enquanto que as outras variáveis têm coeficientes iguais a 0. Com isso a lasso é usada também como um processo de seleção de variáveis.

Note que, com o objetivo de fazer com que a penalização não dependa da origem, o intercepto (β_0) foi deixado fora da restrição de penalidade.

Estas regressões com penalidade são exemplos clássicos de regularização. Todos esses métodos tendem a reduzir os estimadores de mínimos quadrados em direção ao 0 (*shrink toward 0*). Com isso os estimadores desses métodos são viesados. A vantagem desses métodos é abrir mão de um estimador não viesado para obter menor variância nas previsões.

O termo λ é chamado de parâmetro de sintonia (tuning parameter) e a restrição $\lambda (\beta^T \beta (1 - \alpha) + \alpha |\beta|)$ é chamada de *shrinkage penalty*. Assim, a escolha, ou calibragem, do parâmetro de sintonia (λ) é essencial para a estimação. Este é responsável pela magnitude da restrição imposta aos coeficientes (regularização).

As perguntas que têm que ser feitas aqui é porque existe esse *trade-off* entre variância do modelo e viés dos parâmetros e como que a sintonia de λ reduz a volatilidade das previsões do modelo.

Para responder a essas perguntas, primeiro vamos definir a chamada *bias-variance decomposition* e o *bias-variance trade-off*.

2.1.3.1 Custo-benefício do viés e variância

Segundo a decomposição do viés-variância (*bias-variance decomposition*), para uma amostra teste, o erro quadrático esperado em um ponto qualquer x_0 ($E(y_0 - \hat{f}(x_0))^2$) sempre pode ser decomposto em (i) variância do erro do modelo (σ_ε^2), (ii) quadrado do viés do modelo no ponto ($Bias^2(\hat{f}(x_0)) = [E\hat{f}(x_0) - f(x_0)]^2$) e (iii) variância de $Var(\hat{f}(x_0)) = E[\hat{f}(x_0) - E\hat{f}(x_0)]^2$.⁴

$$E(y_0 - \hat{f}(x_0))^2 = \sigma_\varepsilon^2 + Bias^2(\hat{f}(x_0)) + Var(\hat{f}(x_0)) \quad (2.4)$$

A equação (2.4) desenvolve a equação (2.1) para um ponto *out of sample*. O termo de variância do erro (σ_ε^2) não pode ser reduzido. Como $Bias^2(\hat{f}(x_0))$ e $Var(\hat{f}(x_0))$ são sempre positivos, o ponto aqui é encontrar um método que reduza a $Var(\hat{f}(x_0))$ em uma proporção maior que o aumento em

⁴ A prova matemática dessa decomposição não será desenvolvida. Para maiores detalhes ver Hastie et al. [2009]

$Bias^2(\hat{f}(x_0))$. A calibragem do parâmetro de sintonia (λ) é feita com o objetivo de ter um *trade-off* ótimo. James et al. [2013] desenvolve uma discussão detalhada sobre esse *trade-off* e a *bias-variance decomposition*.

2.1.3.2 Regressão ridge

No caso específico da regressão de ridge, podemos encontrar, para um determinado valor do parâmetro de sintonia (λ), uma solução fechada na forma da equação (2.5).

$$\beta_{\lambda}^{ridge} = (X^T X + \lambda I_p)^{-1} X^T y \quad (2.5)$$

A motivação inicial da regressão de ridge foi incluir uma perturbação na diagonal principal de $X^T X$ de forma que essa possa ser invertida mesmo no caso de ser singular.

Os benefícios da regressão ridge são especialmente sentidos na presença de multicolinearidade dos regressores. Isso porque, nesses casos, os coeficientes dos regressores tendem a ser “inflados”.

No entanto, uma das principais críticas à regressão de ridge é que ela não exclui nenhuma variável preditora e por isso os modelos estimados pela regressão de ridge tendem a ferir o princípio da parcimônia.

2.1.3.3 Regressão Lasso

Ao contrário da regressão ridge, a regressão lasso permite que alguns coeficientes sejam iguais a 0.

A solução do problema de otimização da lasso, para restrição l_1 é mais complexo que a regressão ridge porque não conseguimos uma solução fechada como a equação (2.5). Inicialmente o problema da estimação da regressão lasso era resolvido por otimização convexa. No entanto, esse procedimento foi substituído posteriormente por algoritmos mais eficientes como a **least angle regression** - LARS (Efron et al. [2004]) e a **pathwise coordinate descent** - PCD (Friedman et al. [2007]).

O LARS faz um link entre a lasso e a *forward stepwise*. Na verdade é um algoritmo de regressão que pode ser usado para diferentes métodos. A idéia é seguir uma lógica similar a *forward stepwise* em que, a cada iteração se identifica a variável mais correlacionada com o resíduo corrente. Uma vez identificada se aumenta em um ε o coeficiente dessa variável na direção da correlação. Quando um outro regressor apresentar uma correlação mais alta com o resíduo, se aumenta os dois coeficientes em um ε . O procedimento continua até incluir todas as variáveis no modelo. Esse procedimento desenvolve o caminho dos regressores de uma regressão de lasso para o parâmetro de sintonia (λ) variando de 0 a infinito.

O algoritmo da *least square angle* é apresentando no algoritmo (1). O processo inicia com o vetor β , de ordem $(p \times 1)$, igual a 0 para todos os p coeficientes a serem estimados. Dessa forma, Y é igual ao resíduo (r). No segundo passo se identifica o regressor mais correlacionado com o resíduo atual e se soma ou diminui um ε ao coeficiente da variável j (β_j), dependendo do sinal da correlação entre r e X_j ($sign(X_j, r)$). No penúltimo passo, se atualiza o resíduo e se reinicia o processo do passo 2..

O PCD é um algoritmo antigo e que ganhou recentemente muita atenção por ser eficiente na estimação da regressão de lasso. A idéia é otimizar uma variável (coordenada) de cada vez, mantendo as demais fixas.

Algorithm 1 Least angle regression - lasso

1. inicia $y = \beta X + r$, com $\beta = 0$ tal que $y = r$;
 2. identifica o preditor j em X que é mais correlacionado com o resíduo (r);
 3. atualiza o coeficiente do preditor j (β_j) em ε , tal que $\beta_j = \beta_j + \varepsilon_j$ e $\varepsilon_j = \varepsilon * \text{sign}(X_j, r)$;
 4. re-defina $r = r - \varepsilon_j X_j$;
 5. volte ao passo 2.;
-

2.1.3.4 Regressão ridge & lasso

Considerando apenas as duas situações extremas da *elastic net*, regressão ridge e lasso, podemos ressaltar algumas diferenças entre esses casos. Essas diferenças destacam vantagens de um método sobre o outro.

Primeiro que a ridge, devido a sua penalização quadrática, não exclui nenhuma variável. Isso quer dizer que o método aplica uma penalidade sobre os coeficientes no sentido de diminuir sua magnitude absoluta em direção do 0 (*shrink toward 0*), mas sem definir nenhum exatamente igual a 0. Isso pode comprometer a interpretação do modelo para um elevado número de preditores. A lasso por outro lado, ao aplicar a penalidade sobre os coeficientes do modelo (*shrink toward 0*) permite que alguns coeficientes sejam exatamente iguais a 0⁵. Com isso a lasso realiza um tipo de processo de seleção de variáveis⁶. Esse processo de seleção de variáveis, que faz como que os coeficientes de algumas variáveis sejam 0, é conhecido na literatura como *sparsity-based model regularization*. Esse processo de seleção de variáveis também permite que os modelos estimados pela lasso sejam mais fáceis de serem interpretados que os modelos gerados pela ridge.

Segundo que pode haver diferença em termos de desempenho em previsões *out of sample* dependendo da importância relativa dos preditores dentro do processo real gerador de dados. Como a lasso pode fazer com que alguns coeficientes sejam iguais a 0, então em funções onde os coeficientes verdadeiros dos preditores são mais homogêneos, a acurácia de suas previsões é menor que a ridge. Em funções onde alguns preditores têm coeficientes relativamente muito mais importantes que o resto, então a lasso terá um desempenho relativo melhor. No entanto, não podemos dizer que existe uma dominância de um método por outro (Tibshirani [1996] e Fu [1998]).

Terceiro que testes empíricos mostram que a lasso é mais eficiente em destacar os coeficientes das variáveis *signal* (variáveis realmente importantes para explicar a variável objetivo) e variáveis *noise* (variáveis não relacionados com o que se quer estudar)⁷. Tendo em mente essa vantagem da lasso em identificar as variáveis *signals* e a já mencionada característica de que esses estimadores são viesados, destacamos o método conhecido como **relaxed lasso** (Meinshausen [2007]). Esse procedimento visa estimar estimadores não viesados das variáveis *signals*. Ele consiste em primeiro utilizar a lasso para selecionar as variáveis

⁵ Ver James et al. [2013] para uma explanação didática de como essas restrições permitem que os coeficientes assumam valores iguais a 0 (solução de canto) ou não (solução interior).

⁶ Na verdade, tanto a *ridge regression* quanto a *lasso regression* podem ser encaradas como alternativas aos procedimento de *best subset selection*. A *lasso regression* especificamente é utilizada como uma alternativa computacionalmente eficiente em relação aos métodos de *best subset selection* que são computacionalmente custosos.

⁷ Na literatura de processamento de sinais a *lasso regression* é chamada de *Basis Pursuit Denoising*.

signals em um processo de seleção de variáveis e em segundo rodar um método linear que gere estimadores não viesados apenas para as variáveis selecionadas no primeiro passo.

Quarto que a *lasso* tende a ignorar o problema de multicolinearidade das variáveis enquanto que a *ridge* tende a aproximar os coeficientes das variáveis multicolineares. Esse na verdade é apontando como um dos maiores problemas da *lasso*.

2.1.3.5 Elastic net e parâmetro de sintonia (λ)

O objetivo inicial da *elastic net regression* foi abordar o problema da *lasso* em ignorar a multicolinearidade dos regressores (Zou and Hastie [2005]).

Ela pondera (faz uma combinação convexa entre) a restrição que destaca a individualidade (penalidade l_1) com a restrição que garante uma penalidade democrática nos regressores (penalidade l_2). Assim, é possível verificar que ela tem um desempenho melhor que a regressão *lasso* ao mesmo tempo em que mantém a característica de selecionar variáveis. Uma outra vantagem da *elastic net* sobre a *lasso* é que a *lasso* não funciona muito bem, como método de seleção de variável, em casos em que o número de preditores é muito maior que o de observações. Nesses casos, a *elastic net* tem um desempenho melhor que a *lasso* na seleção de variáveis.

A otimização do parâmetro de sintonia (λ) pode ser feita de diferentes formas. A abordagem mais simples é plotar os diferentes valores do critério de ajustamento (*fit*) (RMSE por exemplo) contra o parâmetro de sintonia. Outra forma é plotar usando validação cruzada para computar o modelo com menor erro.

Apesar de interessante e pertinente, não vamos entrar em detalhes na discussão sobre as diferenças e implicações das diferentes restrições (penalidade l_1 e l_2). Para uma discussão aprofundada sobre o assunto ver Hastie et al. [2009].

Outro método é o **reduced rank regression**.

2.1.4 Métodos de redução de dimensão dos dados

Esses métodos consistem em projetar a base de dados de dimensão p em um espaço de dimensão M tal que $p < M$. Com isso, se reduz o número de preditores dos modelos.

A idéia desses métodos é basicamente uma regressão em dois estágios. No primeiro estágio se cria F variáveis latentes (Z_t^j como $j \in \{1, \dots, F\}$) que são combinações lineares de todos os preditores do modelo (X_t) tal como descrito na equação (2.6). Essas variáveis latentes são ortogonais entre si ($E(Z_t^j, Z_t^i) = 0 \forall i \neq j$). No segundo estágio se faz individualmente uma regressão de cada variável latente contra a variável explicada do modelo (equação (2.7)). O modelo final é o somatório dos coeficientes encontrados em cada regressão (equação (2.8)). Isso pode ser feito porque as variáveis latentes são ortogonais entre si.

$$Z_t^j = \beta^{fst,j} X_t^j \tag{2.6}$$

$$y_t = \beta^{scd,j} Z_t^j + \varepsilon_t^j \tag{2.7}$$

Algorithm 2 Principal Component regression

1. Fazer uma análise de componentes principais (PCA) a base de dados X ;
 2. identificar F componentes principais (Z_t^j) ortogonais entre si ($E(Z_t^i, Z_t^j) = 0 \forall i \text{ e } j \leq F$);
 3. Calcular os F componentes principais (Z_t^1, \dots, Z_t^F) tal que $Z_t^j = \beta^{fst,j} X_t^j$;
 4. Calcular o coeficiente linear de cada componente principal:
for $j = 1$ to F
 . $b^j = \left[\begin{pmatrix} (Z_t^j)^T & Z_t^j \end{pmatrix}^{-1} \begin{pmatrix} (Z_t^j)^T \\ Z_t^j \end{pmatrix} y_t$
end
 5. Modelo final: $y_t = \sum_{j=1}^F \beta^{scd,j} Z_t^j$
-

$$y_t = \sum_{j=1}^F \beta^{scd,j} Z_t^j \quad (2.8)$$

Uma outra abordagem, caso se tenha graus de liberdade suficientes, seria pular a equação (2.7) e estimar diretamente a equação (2.8).

O método de redução da dimensão dos dados mais conhecido é o **componente principal** (PCA). Assim como a **análise de fatores**, o componente principal é usado em situações onde conjuntos de dados possuem variáveis com elevado nível de multicolinearidade ou quando se tem um número elevado de preditores em relação ao número de observações.

Outro método é o **partial least squares** (PLS) que, assim como o PCA, busca uma combinação linear das variáveis explicativas. A diferença dos dois, segundo Kuhn and Johnson [2013], é que o PCA tem como solução uma combinação linear dos preditores que maximiza a variabilidade conjunta destes no espaço, enquanto o PLS tem como solução uma combinação linear dos preditores que maximiza a covariância destes com a variável explicada. Segundo os autores, o método funciona bem com amostras menores, mas é ineficiente para base de dados maiores. Por ser um método menos usado vamos omitir o algoritmo. Para maiores detalhes ver Hastie et al. [2009].

2.2 Modelos Não lineares

Ao contrário dos modelos lineares, os modelos não lineares têm estruturas para $E(Y|X)$ que não são lineares nos parâmetros dos preditores (X). A flexibilização do pressuposto de linearidade nos parâmetros tem implicações sobre a interpretabilidade das estimações. Por exemplo, quando aplicável, o efeito marginal de cada preditor é mais complexo de ser obtido que nos modelos lineares.

A literatura sobre esses modelos é ampla. Como exemplos de métodos de estimação não-lineares supervisionados temos regressões polinomiais, splines, logística, métodos de kernel, smoothing, modelos aditivos generalizados (GAMs), modelos de regressões em árvore, redes neurais, etc.

Como o objetivo dessa tese é tratar métodos em destaque na literatura recente, vamos selecionar apenas alguns deles.

O primeiro que vamos abordar são as redes neurais em suas abordagens mais modernas. Esses modelos se tornaram muito populares nos anos 80 e 90, mas sofreram muitas críticas relacionadas às interpretações das estimações, a inexistência de bons testes para a qualidade do modelo e o fraco desempenho preditivo *out of sample*. Destacamos esses métodos porque, mais recentemente, têm ganhado atenção dentro do conceito de *deep learning* em função da eficiência preditiva com novas abordagens.

O segundo grupo de métodos são os derivados dos modelos de regressão em árvore. As abordagens mais simples de regressão em árvore estão sujeitas a problemas de *overfitting* e consequente desempenho preditivo *out of sample* insatisfatório. Por outro lado, técnicas mais sofisticadas de regressão em árvore que envolvem *prunning*, geração randômica de árvores e ponderação de modelos, estão entre os métodos reconhecidos na literatura mais eficientes em previsão *out of sample*.

Por fim, vamos tratar dos métodos de estimação do *vector support machine* que assim como os 2 outros mencionados anteriormente estão entre os métodos com melhor desempenho preditivo *out of sample* na literatura.

2.2.1 Redes neurais

As redes neurais podem ser entendidas como um método de regressão, ou de modelagem, em dois estágios, assim como a PCA ou a PLS. No entanto, ao contrário desses métodos, as redes neurais estimam relações não lineares nas suas etapas intermediárias.

No primeiro estágio se estima as chamadas *hidden variables* ou *hidden units*, que são variáveis não observadas equivalentes às variáveis latentes de uma PCA ou PLS. Elas são funções não lineares de combinações lineares de todos ou alguns preditores (ou variáveis observadas) do modelo.

Assim, para K *hidden variables*, uma rede neural estima k funções não lineares (equação (2.10)) de M combinações lineares dos preditores observados X (equação (2.9)). O *output* final da rede neural ($f(x)$) é uma combinação linear as K *hidden variables* (equação (2.11)).

$$Z_m = \beta_m X \quad (2.9)$$

$$h_k(x) = g(Z) \quad (2.10)$$

$$f(x) = \sum_{k=1}^K \alpha_k h_k \quad (2.11)$$

Seja o conjunto total de parâmetros de uma rede neural dado pelo vetor θ . Esses parâmetros são chamados de *weights* (pesos) da rede neural e o vetor θ é composto por:

$$\{\beta_1, \dots, \beta_J, \alpha_1, \dots, \alpha_M\}$$

Onde K é o número de *hidden variables*. β_j é o vetor de coeficientes da combinação linear dos preditores na *hidden variable* j (equação (2.9)) de dimensão $J \times 1$, onde J é o número de preditores (colunas) em

Algorithm 3 Neural Net - gradient descent (back-propagation)

1. inicializar os weights $\theta = 0$;
 2. definir os inputs X e outputs Y ;
 3. estimar os erros por hidden variables;
 4. atualizar os weights;
-

X . α_m é o coeficiente da *hidden variable* m no *output* final da rede neural (equação (2.11)) e β_m é o vetor ($1 \times J$) de J coeficientes dos inputs X na variável Z_m .

Assim, para uma rede neural sem intercepto nas equações (2.9) e (2.11) e com apenas um *output* na equação (2.11), temos $M * J + K$ *weights* em θ a serem estimados. Isso leva a um número elevado de parâmetros no caso de ausência de restrição para o número de *hidden units*. Dessa forma, as redes neurais tendem a ser sobreparametrizadas (*overfitting*), pois uma rede neural com mais *hidden units* é melhor para representar as não linearidades dos dados. Esse problema de *overfitting* geralmente é tratado com a utilização de algum método de regularização através de um termo de penalidade ou por algum mecanismo de *early stopping*.

Um método de regularização é o uso da *weight decay* que é uma penalização quadrática equivalente a restrição da regressão de ridge.

Outro problema das redes neurais está relacionado a não convexidade do problema. Assim, por haver vários pontos ótimos locais, é interessante rodar para vários valores iniciais diferentes.

No entanto, a maior crítica que se faz as redes neurais está relacionada à interpretabilidade dos coeficientes estimados. Como a otimização do problema da rede neural é irrestrita, os coeficientes estimados assumem valores que não podem ser interpretados.

Para estimar o modelo dado pelas equações (2.9), (2.10) e (2.11), vamos assumir como critério de ajustamento (*fit*) o erro quadrático médio.

$$R(\theta) = \sum_{i=1}^N (y_i - f(x_i))^2 \quad (2.12)$$

Um método comum de estimar redes neurais é por *backward propagation* (algoritmo (3)). A ideia é treinar a rede ajustando os *weight* de cada *hidden variable* ($h(\cdot)$) pela derivada da função custo (equação (2.12)) em relação ao *weight* (α_m ou β_m).

A abordagem de redes neurais, que foi muito popular nos anos 80 e 90, sofreu muitas críticas. Uma das principais se refere ao *overfitting* e fraco desempenho preditivo. No entanto, versões mais modernas usam o *output* da equação (2.11) com *inputs* em novas etapas equivalentes a equação (2.9). Isso pode ser feito várias vezes criando vários níveis dentro da rede. A isso se chama de *deep learning* e métodos que seguem este raciocínio estão entre os mais competitivos na ML.

2.2.2 Modelos Aditivos

Os modelos aditivos podem ser entendidos como médias ponderadas de outros modelos. Nesse sentido esses modelos podem ser gerados por procedimentos de *boosting*.

Seja a formulação genérica de um modelo aditivo dada por:

$$y_t = \beta_0 + \sum g_i(x_i) + \varepsilon_t$$

Onde $g_i(x_i)$ é uma função, linear ou não, de x_i . Assim, a regressão linear múltipla é um caso especial quando $g_i(x_i) = \beta_i x_i$.

Em um procedimento de ponderação de modelos do tipo *boosting*, as funções $g_i(x_i)$ são modelos estimados individualmente.

2.2.3 Regressão em árvore

Os métodos de regressão em árvore podem ser usados tanto para modelos de classificação ou categorização quanto para regressão e são chamados genericamente de modelos CART. No caso específico de modelos onde a variável explicada é binária, os métodos de regressão em árvore são alternativas para os modelos de regressão logística. Nessa tese vamos nos referir genericamente a esses modelos como *regressões em árvore* e uma aplicação a um modelo com variável dependente binária, ordenada ou contínua vai depender do contexto.

A idéia da regressão em árvore é particionar a amostra de *treinamento* em sub-espacos menores que destaquem características que melhor funcionem para prever a variável objetivo. É um tipo de *piecewise regression model*. Esse processo tem como objetivo maximizar o ajustamento (*fit*) do modelo total ao estimar modelos específicos para cada subgrupo da base de dados. Na verdade a modelagem por regressão de árvore não envolve processos usuais de estimação e sim algoritmos de otimização⁸.

Regressões em árvore são métodos mais indicados para situações onde as relações entre as variáveis são complexas e não são bem aproximadas por um modelo linear. Nesse sentido, esses métodos são usados em etapas exploratórias de base de dados. São úteis em termos de interpretação, mas que não são competitivos em termos de previsão *out of sample* quando comparados com outros métodos mais eficientes. Por isso esses métodos são usados geralmente em conjuntos com processos de aleatorização da amostra de treinamento como o *bagging*, o *boosting* e o *random forest*, o que melhora significativamente o seu desempenho. Este é provavelmente o método de *aprendizado de máquina* mais usado em *datamining* (Witten et al. [2011]).

Uma regressão em árvore pode ser vista como um modelo aditivo. James et al. [2013] compara um modelo de regressão linear (aditivo) com uma regressão em árvore. As similaridades podem ser visualizadas na equação (2.13), para um modelo de regressão linear, e na equação (2.14), para um modelo de regressão em árvore.

$$f(X) = \beta_0 + \sum_{j=1}^p X_j \beta_j \quad (2.13)$$

$$f(X) = \sum_{m=1}^M c_m \cdot I(X \in \mathbb{R}_m) \quad (2.14)$$

⁸ Nesse sentido, na abordagem tradicional da regressão em árvore não existe medida de significância ou distribuição dos coeficientes do modelo. Para contornar essa limitação Hothorn et al. [2006] desenvolveu a chamada **conditional inference tree**.

Onde, R_1, \dots, R_m são partições do espaço. $I(X \in \mathbb{R}_m)$ é uma função de indicação que assume o valor 1 quando $X \in \mathbb{R}_m$ (a observação pertence ao espaço em questão, ou partição do espaço total) e 0 caso contrário. c_m são os valores pela variável objetivo ($y = f(x) + \varepsilon$) em \mathbb{R}^m . No caso de uma variável binária $c_m \in \{0, 1\}$.

Para construir uma árvore, vamos definir o problema do modelo de regressão em árvore como minimizar a soma do quadrado dos erros de previsão dentro de cada subconjunto do espaço da base de dados. Seja R_j o subespaço j do espaço da base de dados, então a equação (2.15) descreve a função perda da regressão em árvore.

$$SQR_{tree} = \sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \quad (2.15)$$

Onde \hat{y}_{R_j} é a resposta média dentro do subconjunto R_j .

Podemos resumir a idéia dos algoritmos de regressão árvore como um procedimento no qual, para cada nó, identificamos qual a melhor variável j para dividir a base e qual o ponto de corte s que deve ser usado. Se o critério da regressão, para uma variável contínua, é minimizar a soma dos quadrados dos resíduos, então, em cada nó se resolve o problema da equação (2.16). Ou seja, em cada nó se escolhe a variável j e o ponto de corte que mais reduz a soma do quadrado dos erros naquele nó.

$$\min_{j,s} \left[\sum_{x_i \in R_1 \in (j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{x_i \in R_2 \in (j,s)} (y_i - \hat{y}_{R_2})^2 \right] \quad (2.16)$$

Onde, R_1 e R_2 são os espaços em que R é subdividido. Cada um desses espaços é escolhido ao se minimizar o quadrado dos desvios de cada variável (y_i) em relação a sua média nesse espaço (\hat{y}_{R_1} e \hat{y}_{R_2}).

Um aspecto importante dos modelos baseados em regressão em árvore é que estes fornecem uma representação lógica dessas regiões na forma de árvore. Cada *path* que tem origem na base da árvore e vai até uma folha corresponde a uma região. Cada nó interior de uma árvore corresponde a um teste sobre uma variável preditora. Isso quer dizer que para cada partição associada D_i nós temos um path P_i que consiste de uma conjunção lógica de testes sobre as variáveis predictoras.

Por outro lado, esse tipo de modelagem possui questões importantes que devem ser consideradas. A literatura destaca três principais. Primeiro que esses modelos de regressão em árvore geralmente possuem problemas de *overfitting*. Segundo que não possuem bom desempenho preditivo *out of sample* em função de sua metodologia de estimação, ou derivação e de possível *overfitting*. O terceiro problema que vale a pena mencionar é que os resultados da regressão em árvore são sensíveis a mudanças na amostra de treinamento.

Nesse sentido, devem-se ponderar formas de tratar essas questões. No que se refere ao *overfitting*, uma forma de controlar para esse problema é usar abordagens de *prunning* de árvores (limitar o seu crescimento). Para os problemas de deficiências no desempenho *out of sample* e da sensibilidade do modelo estimado à amostra de treinamento, é indicado procedimentos que aumentem a robustez dos estimadores. Dentre esses procedimentos, a *random forest* é conhecida na literatura como um método de modelagem com desempenho *out of sample* bastante satisfatório. Na verdade, o *random forest* está entre os métodos com melhor desempenho preditivo *out of sample*. Outros métodos também utilizados e que melhoram o desempenho *out of sample* das árvores são *bagging* e o *boosting*. Todos esses métodos

Algorithm 4 CHAID (chi-squared automatic interaction detection)

1. **preparar preditores:** preditores mensurados em categoria;
 2. **unir categorias:** unir pares de categorias dos preditores e calcular estatística do teste χ^2 ou F ;
 3. **selecionar a variável de divisão:** escolher o par de preditores com a divisão (split) mais significativa;
-

envolvem estimar múltiplas regressões de árvore a partir de múltiplas amostras de treinamento geradas da amostra de treinamento original. Esses procedimentos partem da idéia de que se reduz a variância de uma amostra de observações, cada uma com uma variância individual, ao se ponderar elas. No entanto, esses processos de ponderar árvores ao mesmo tempo em que reduzem a variância das projeções comprometem a facilidade de interpretação das estimações finais. Isso porque como o resultado final são várias árvores diferentes, então nem sempre é possível dizer, por exemplo, quais variáveis preditivas são mais importantes. Também não é possível fazer uma representação em árvore, que é um dos principais atrativos desse tipo de modelagem.

No caso do *bagging*, as múltiplas árvores geradas pelos múltiplos processos de *bootstrap* não são submetidas à *prunning*. Isso quer dizer que crescem sem limitações e por isso todas possuem *overfitting*. Dessa forma, cada árvore individualmente possui elevada variância nos estimadores, mas baixo viés (James et al. [2013]).

O *boosting* para regressões em árvore é um processo de estimação iterativo em que para uma regressão em árvore inicial de referência em cada iteração se estima uma árvore para explicar o resíduo da árvore estimada na iteração anterior. Com isso, em cada iteração se estima uma árvore menor.

O *random forest* proposto por Breiman [2001] guarda similaridades com o *bagging*. Assim como o *bagging*, o *random forest* estima uma árvore para cada amostra do *bootstrap*. A diferença é que, para determinada árvore, em cada nó, se seleciona um número aleatório de preditores para a próxima decisão da árvore.

Essa abordagem “diferente”, segundo Breiman [2001], tem um desempenho preditivo *out of sample* superior a outros métodos de classificação. De fato, a eficiência do *random forest* foi comprovada em diversos trabalhos posteriores como o de Bajari et al. [2015a], Bajari et al. [2015b] e Fernández-Delgado et al. [2014] para citar exemplos mais próximos em economia.

2.2.3.1 CHAID

O algoritmo de CHAID (chi-squared automatic interaction detection) é um dos métodos mais antigos e tradicionais de regressão em árvore. É um algoritmo rápido indicado para grandes bases de dados.

O processo (algoritmo (4)) requer inicialmente que os preditores sejam transformados em variáveis categóricas. A idéia central é que a divisão em cada nó com base em pares de categorias de preditores.

2.2.4 Support vector machines

O SVM é um método que generaliza a idéia do *maximal margin classifier*. Por sua vez, o *maximal margin classifier* é um método de classificação que separa as observações da amostra de treinamento através de um

hiperplano separador ótimo. Esse hiperplano separador ótimo é definido com base em alguns pontos da amostra de treinamento chamadas de *support vectors*. Dessa forma, os *support vectors* são determinantes para o modelo estimado por SVM⁹.

A questão central é que, para se usar o *maximal margin classifier* é preciso que as observações possam ser separadas por um hiperplano. Quando isso não é possível (existem observações que não estão do lado certo da margem do hiperplano) então é preciso usar um método que flexibilize o *maximal margin classifier* para encontrar um hiperplano separador ótimo que permita que algumas observações não sejam classificadas corretamente pelo hiperplano separador. Ou seja, que o hiperplano tenha algum grau de tolerância à erros de classificação¹⁰. Segundo James et al. [2013] esse método mais flexível deveria levar em conta: (i) robustez das estimações e (ii) melhor classificação da maioria das observações. O SVM flexibiliza o *maximal margin classifier* e atende a esses critérios.

Na situação base, onde a variável dependente é binária, o SVM compete diretamente com modelos de classificação como modelos logit e regressão em árvore.

No entanto, é possível usá-lo tanto para variáveis dependentes em vários níveis como para variáveis dependentes contínuas (VSM regression).

Usando como referência o caso da variável dependente binária ($Y \in \{-1, 1\}$), o problema do *maximal margin classifier*, definido na equação (2.17), é encontrar o hiperplano separador ótimo (dado pelo vetor de coeficientes β) de forma a maximizar a margem desse hiperplano (M).

$$\begin{aligned} \underset{\{\beta_0, \beta\}}{MAX} \quad & M \\ \text{s.t.} \quad & \|\beta\| = 1 \\ & y_i (x'_i \beta + \beta_0) \geq M \quad , i = 1, \dots, M \end{aligned} \tag{2.17}$$

Onde a restrição ($y_i (x'_i \beta + \beta_0) \geq M$) garante que todos os elementos de cada lado do hiperplano estão a pelo menos uma distância M desse hiperplano.

O problema dado pela equação (2.17) pode ser reescrito na forma da equação (2.18), onde usamos $M = \frac{1}{\|\beta\|}$. Essa é uma forma conveniente porque define um problema de otimização convexo (Hastie et al. [2009]).

$$\begin{aligned} \underset{\{\beta_0, \beta\}}{MIN} \quad & \|\beta\| \\ \text{s.t.} \quad & y_i (x'_i \beta + \beta_0) \geq 1 \end{aligned} \tag{2.18}$$

Assim, seguindo a idéia de flexibilizar o problema do *maximal margin classifier*, para permitir que algumas observações não estejam do lado certo do hiperplano, redefinimos o problema da equação (2.18) na forma da equação (2.19).

$$\begin{aligned} \underset{\{\beta_0, \beta\}}{MIN} \quad & \|\beta\| \\ \text{s.t.} \quad & y_i (x'_i \beta + \beta_0) \geq 1 - \xi_i \quad \forall i \\ & \xi_i \geq 0 \quad \forall i \end{aligned} \tag{2.19}$$

⁹ Mais ainda, métodos baseados em hiperplano separador que usam o support vectors são robustos a outliers (Hable and Christmann [2011])

¹⁰ A noção de soft margin que implementa esse conceito foi desenvolvido por Cortes and Vapnik [1995].

2.3 Seleção de modelos e Model Averaging

Nas seções anteriores desse capítulo tratamos de dois grandes grupos de métodos de estimação. Aqui vamos discutir como melhor escolher modelos entre diferentes estimações e como combinar resultados de diferentes modelos para obter melhores previsões. Enquanto a discussão sobre critérios de escolha de modelos é mais madura, a literatura sobre como combinar modelos está mais na fronteira de pesquisa.

Em uma abordagem computacional, o problema da escolha do modelo mais adequado para responder determinada pergunta se torna mais complexo de ser solucionado com o aumento do número de preditores a serem considerados. No entanto, é importante destacar que a classificação ou critério de classificação de um modelo como melhor ou não depende da pergunta que o modelo deve responder.

Um dos critérios mais bem aceitos, independentemente da pergunta, na literatura é o desempenho do modelo *out of sample*. Isso porque não existe, necessariamente, relação entre o desempenho dos modelos *in sample* (usando a amostra de treinamento) e *out of sample* (usando a amostra de validação). Ou seja, o modelo com menor erro de treinamento não necessariamente teria o menor erro de validação. Isso é um problema, pois geralmente o que se quer é estudar ou prever com dados ainda não observados. Nesse sentido, Guyon et al. [2010] destaca que, independente da corrente de pensamento, o procedimento de validação cruzada provavelmente é considerado o método mais eficiente de avaliação de modelos¹¹.

Assim, em um mundo ideal, com observações suficientes para se fazer estimações e verificação dos modelos, a literatura de aprendizado de máquina sugere dividir a base de dados em três partes. A primeira parte, que corresponde a 50% da amostra, é utilizada para a estimação (aprendizado) e chamada de amostra de treinamento. A outra metade da amostra é subdividida em duas outras partes, a amostra de validação e a amostra de teste. A amostra de validação é usada para calcular os erros de previsão dos modelos concorrentes e escolher o melhor modelo. A amostra de teste é usada para verificar o desempenho de previsão do modelo selecionado na amostra de validação. A amostra de treinamento é o que a econometria chama de *in sample* e a amostra de validação e a de teste é o *out of sample*.

No entanto, no mundo real o economista geralmente possui menos observações nos dados que gostaria. Assim, a escolha do critério de avaliação depende, além da pergunta a ser respondida, do tamanho da amostra.

Tendo em mente essas duas dimensões, Rao et al. [2001] descreve uma série de critérios divididos em vários grupos, como métodos baseados em testes de hipóteses, critérios de informação e abordagem bayesiana dentre outros.

Nessa tese vamos abordar um conjunto mais restrito de técnicas de seleção de modelos. Para uma cobertura mais abrangente do assunto sugerimos Rao et al. [2001].

Basicamente, a tese aborda dois grupos de métodos de seleção. O primeiro são os critérios de informação *in sample*, que controlem para a complexidade do modelo ou *overfitting*. O segundo utiliza procedimentos de validação cruzada para ajustar a medida de erro através da reutilização da amostra. Esses últimos processos estão entre as ferramentas mais importantes da teoria de aprendizado de máquina.

Existe um terceiro que não será abordado aqui que são os métodos bayesianos.

¹¹ Os autores dividem as correntes teóricas sobre seleção de modelos em frequentistas, bayesianos e práticos. Indo mais além, os autores argumentam que o problema de seleção de modelos é um assunto com pouca margem para discussão nas três correntes. Para os frequentistas o consenso seria que o melhor modelo é aquele que minimiza o erro das previsões. Para os bayesianos a melhor abordagem é ponderar as previsões de vários modelos. Por fim, para os práticos a abordagem mais eficiente para escolha do melhor modelo é a validação cruzada.

Tabela 1 – Critérios de informação para seleção de modelos

Critério de Informação	Fórmula - modelos Gaussianos	Fórmula - modelos gerais	Critério de decisão
Bayesian information criteria (BIC)	$BIC = \frac{1}{n} (RSS + \log(n) d \hat{\sigma}^2)$	$BIC = -2 * ll + \log(n) * d$	menor melhor
Akaike information criteria (AIC)	$AIC = \frac{1}{n \hat{\sigma}^2} (RSS + 2d \hat{\sigma}^2)$		menor melhor
Estatística C_p de Masllow	$C_p = \frac{1}{n} (RSS + 2d \hat{\sigma}^2)$		menor melhor
R^2 ajustado	$\bar{R}^2 = 1 - \frac{RSS/(n-d-1)}{TSS/(n-1)}$		maior melhor
Fator de Bayes (BF)		$BF = \frac{Pr(X M_a)}{Pr(X M_b)}$	$BF > 1 \Rightarrow M_a$

NOTA: A tabela resume os critérios de informação mais utilizados, a fórmula de cálculo e o critério de decisão, ou julgamento para escolha dos modelos. Onde, RSS é a soma dos quadrados dos resíduos e é dado por $RSS = (Y - \beta X)^T (Y - \beta X)$, d é o número de preditores, n é o tamanho da amostra, TSS é a soma total dos quadrados dado por $TSS = (y - \bar{y})^T (y - \bar{y})$, ll é a log verossimilhança.

2.3.1 Critérios de informação

Os chamados critérios de informação e estatísticas descritas nesse tópico usam informação *in sample*. Essas informações sobre o ajustamento (*fit*) do modelo *in sample* não são medidas adequadas para avaliar o desempenho, ou capacidade preditiva, *out of sample* desses modelos. No entanto, funcionam bem para o objetivo de comparar a adequação aos dados *in sample*.

Existem, na literatura, muitas formas de avaliar o ajustamento (*fit*) de um modelo aos dados. As medidas mais simples, como o tradicional R^2 , ignoram a complexidade do modelo. No entanto, isso não é desejável, pois modelos como um número excessivo de parâmetros tendem a ter *overfitting*. *Overfitting* é a situação onde uma sobreparametrização do modelo faz com que esse tenha um ajustamento aos dados *in sample* muito bom, mas que esse não consiga reproduzir essa eficiência *out of sample*.

Assim, os critérios de informação mais usados na literatura, que controlam para a complexidade dos modelos, são o *Bayesian information criteria* (BIC), *Akaike information criteria* (AIC), a estatística C_p de Masllow, o R^2 ajustado e o fator de Bayes.

A tabela (1) resume o cálculos dessas estatísticas e o critério de decisão.

Note que, pela tabela (1), o ajustamento de um modelo, dado pelo critério de decisão das 4 primeiras estatísticas listadas, é inversamente proporcional ao número de parâmetros estimados (d), à estimativa de variância ($\hat{\sigma}^2$) e à magnitude dos erros (RSS). No caso da fórmula geral, o ajustamento do modelo é diretamente proporcional a *log-verossimilhança* do modelo estimado (ll).

O fator de Bayes é um critério de informação diferenciado em relação aos demais listados. Ele compara modelos 2 a 2. Seja o modelo M_a e o modelo M_b , o critério de Bayes seleciona o modelo M_a se $BF > 1$ e M_b caso contrário. No entanto, $\log(Pr(X|M_a)) = BIC_a$ se assumirmos uma função perda de $-2 * \log(Pr(X|\theta_a, M_a)) = -2 * ll$.

Apesar de similares, os critérios de informação do *BIC* e *AIC* possuem algumas diferenças. Enquanto o critério do *BIC* é consistente quando $n \rightarrow \infty$, o critério de *AIC* não. Isso quer dizer que, para um conjunto de modelos onde o modelo real faz parte, quando aumentamos a amostra o *BIC* tende a selecionar o modelo real. Por outro lado, o *AIC* tende a escolher modelos com maior complexidade.

2.3.2 Amostra de validação

Nesse tópico vamos discutir os métodos que utilizam dados *out of sample* para testar e avaliar os modelos. Essas abordagens que usam uma amostra de validação são mais eficientes por controlarem para o problema

do viés de seleção. Esse viés ocorre quando se usa a amostra total como amostra de treinamento ou se utiliza critérios de informação *in sample* para avaliar os modelos. Dessa forma, como geralmente o que se quer é usar esses modelos em dados que ainda não ocorreram e não foram observados, os procedimentos que usam o erro de teste tendem a fazer melhores escolhas ou julgamentos.

Como mencionado na introdução do capítulo, o ideal é particionar a amostra total em amostra de treinamento, teste e validação. No entanto, nem sempre isso é possível e o pesquisador deve usar métodos de reamostragem que reutilizam a amostra.

O método que talvez seja o mais eficiente da literatura para calcular o erro de teste, ou erro de validação, é a validação cruzada. A validação cruzada consiste em dividir a amostra de k grupos, usar $k - 1$ grupos para a estimação do modelo e 1 parte para a validação. Esse procedimento é feito para até que todos os k grupos sejam usados como grupos de validação.

O erro de validação é igual à média dos erros de previsão das amostras de validação do processo. Ou seja, o procedimento de validação cruzada calcula ou estima o erro de teste usando a amostra de treinamento.

$$MSE_{CV} = \frac{1}{N} \sum_{n=1}^N \left(y_i - \hat{f}^{-k(i)}(x_i) \right)^2 \quad (2.20)$$

Onde $\hat{f}^{-k(i)}(x_i)$ é a estimativa de y_i usando como amostra de treinamento a amostra total menos a $k - \text{ésima}$ parte (que foi utilizada como amostra de validação).

No caso de se definir o número de grupos igual ao número de observações ($K = N$), chamamos o procedimento de *leave-one-out*.

O número de grupos a ser escolhido no processo de validação cruzada é um ponto importante. Os dois casos extremos são quando $k = 2$, tal que metade da amostra é da amostra de treinamento e a outra de validação, e quando $k = N - 1$, também conhecida como *leave one out cross validation* (LOOCV), pois em cada amostra de treinamento teria $N - 1$ observações. Para um número k grande (por exemplo, próximo de N), a variância do erro das previsões é maior, mas o viés é menor. Com isso o LOOCV seria praticamente uma estimativa não viesada do erro de teste. O contrário ocorre para quanto o número de grupos é menor que no limite seria $k = 2$. Nesse sentido, ignorando o custo de processamento, o melhor procedimento é o LOOCV, mas como nem sempre isso é possível, geralmente se usa $k = 5$ ou $k = 10$.

Um dos métodos mais populares para reutilizar a amostra e gerar amostras de validação é o **bootstrap**. Este é um processo de gerar B amostras aleatórias, com reposição, a partir da amostra original de treinamento. O tamanho dessas novas amostras é, geralmente, igual ao tamanho da amostra original, mas também podem ser usadas amostras menores.

No bootstrap, o cálculo do erro de validação deve ser feito, para cada sub amostra gerada, apenas nas observações que não foram incluídas nessa amostra de forma equivalente à equação (2.20). Isso porque, caso se utilize observações que fizeram parte da amostra de treinamento, então se pode viesar os erros de previsão e incorrer em problemas de *overfitting*. O mesmo raciocínio para o cálculo do erro de validação se aplica aos métodos derivados do bootstrap como o *bagging*, o *boosting* e o *random forest*.

2.3.3 Model Averaging e Ensemble

Nos dois tópicos anteriores, descrevemos como usar informação *in sample* e *out of sample* para escolher modelos. Aqui vamos discutir como combinar modelos. Os termos *model averaging* e *ensemble* vão ser usados como sinônimos.

Model averaging (ou *ensemble learning*) pode ser entendido como um processo de modelagem de uma variável objetivo (aprendizado supervisionado) a partir de previsões dessa variável feitas por outros modelos (*base learners*). Podemos pensar que os procedimentos de *model averaging* envolvem ponderar os resultados ou projeções de diversos modelos. Isso é feito principalmente com os objetivos de reduzir o viés da amostra de treinamento e a variância dos resultados dos modelos. Ou seja, o objetivo final é obter previsões melhores que as feitas por modelos individuais menores.

Assim, partindo de uma amostra de treinamento, uma forma de construir diferentes modelos é estimar a mesma especificação em várias amostras. E para isso o procedimento básico é o *bootstrapping*. O *bootstrapping* consiste em fazer amostragens com reposição a partir de uma amostra de treinamento.

Um outro método relacionado ao tema é o **bagging** ou *bootstrap aggregation*. O *bagging* envolve calcular médias entre modelos estimados com diferentes amostras de *bootstrap* no sentido de aumentar a eficiência de um estimador. Assim, o *bagging* envolve: (i) gerar múltiplas amostras, (ii) estimar modelos para cada uma dessas amostras e (iii) tirar média das previsões de cada modelo. O *bagging* é principalmente útil para métodos não lineares como árvores.

Para um número B de amostras de *bootstrap*.

$$\hat{f}_{bag}(x) = E(Y|X)_{bag} = \frac{1}{B} \sum \hat{f}^{*b}(x)$$

Onde $\hat{f}^{*b}(x)$ é a estimacão feita com a amostra do bootstrap b . Note que o *bagging* é um método de model averaging.

O **boosting** é um outro método de *model averaging* baseado no *bootstrap*. Envolve estimacões repetidas onde observacões não classificadas recebem um peso maior em cada repetição. A estimacão final é então uma média entre as estimacões repetidas. A diferenca do *boosting* para a o *bagging* é que o *bagging* faz uma média com pesos iguais para todos os modelos. O *boosting* geralmente é usado com árvores de decisão onde pode aumentar de forma significativa a performance preditiva. Também é usado para gerar modelos aditivos.

Uma abordagem com um raciocínio mais complexo e que merece destaque é a de Friedmand and Popescu [2008] que propõe a criação de “regras” a partir de dados e que a combinacão dessas “regras” gera modelos com poder preditivo muito competitivo.

Apesar de interessante, essas abordagens de *model averaging* não serão abordadas e aprofundadas nas aplicacões práticas dessa tese.

3 A relação de causa efeito do investimento e produto dentro dos canais de transmissão dos juros, poupança, investimento e produto

O investimento é um dos fatores mais importantes para o crescimento econômico, tanto em termos teóricos como empíricos. Em termos de teoria, o investimento é o principal determinante da acumulação de capital da economia. Por sua vez, a acumulação de capital é importante não só em termos de crescimento, mas também porque a riqueza de um país é definida, em grande parte, pelo capital acumulado. Em termos de trabalhos empíricos, o investimento é uma variável consistentemente significativa em explicar o crescimento dos países no longo prazo. Usamos isso como ponto de partida para estimar a relação estrutural de transmissão entre o produto, o investimento, a poupança e os juros de cada país individualmente em um painel com 101 países. Com base nessa relação estrutural que chamamos de “perfil”, procuramos agrupar os países com dinâmicas internas similares. Fazemos isso com o objetivo de identificar o porquê e como os países se assemelham ou diferenciam em termos de funcionamento interno das economias e que lições podemos tirar disso. Montamos o “perfil” com base em uma estrutura de equações simultâneas que estuda a conexão entre o produto, o investimento, a poupança e os juros. Na verdade, testamos vários sistemas de equações simultâneas, com diferentes relações estruturais entre as variáveis. Mais especificamente, testamos todas as combinações de relações estruturais, contemporâneas, entre as variáveis endógenas segundo uma hipótese de ligação entre os grupos de variáveis do trabalho. Nossos resultados têm três conclusões principais. A primeira é que nossas estimações indicam que, a relação de causalidade do investimento para o produto é mais freqüente que a direção oposta. Nossa segunda conclusão é que existe muita heterogeneidade na dinâmica de funcionamento interno dos países e nos canais de transmissão de juros, poupança, investimento e produto. A terceira conclusão é que, apesar dessas heterogeneidades, foi possível encontrar, em nossos resultados, uma característica robusta que agrupa os países. Esse principal critério para agrupar as economias ou a característica que mais assemelham os países é como a economia reage a variações do humor do investidor internacional. Esse efeito foi medido e verificado diretamente ou através dos fluxos de capitais internacionais e da taxa de câmbio.

3.1 Introdução

O crescimento econômico, direta ou indiretamente, desafia os *policy makers* há séculos. Mais que isso, alguns podem argumentar que o crescimento econômico é um dos objetivos finais, explícito ou não, dos governantes¹.

Na literatura, o início relevante do debate sobre crescimento econômico é a dualidade entre crescimento endógeno e exógeno. No fundo, ambas contam histórias muito similares, mas com visões e mecanismos

¹ Ou uma das principais ferramentas para atingir esse objetivo visto que o crescimento econômico tem efeito benéfico direto em toda a atividade do país.

diferentes. Basicamente, na abordagem exógena, os determinantes do crescimento de longo prazo, que no caso do modelo neoclássico é basicamente o progresso tecnológico, são exógenos. Por outro lado, na abordagem endógena se busca modelar esses determinantes do crescimento dentro do processo de decisão dos agentes.

Na verdade, parte dessas diferenças de visões do mesmo problema tem origem em fragilidades ou críticas à outra abordagem. Em específico, os modelos de crescimento endógenos vieram, em meados dos anos 80, como resposta às dificuldades dos modelos de crescimento exógenos em explicar as diferenças das dinâmicas de crescimento observadas empiricamente nos países. A década que se seguiu ao trabalho de Romer [1986] foi de intensa produtividade na literatura de crescimento econômico.

Da teoria de crescimento exógeno, e dos seus resultados e projeções, se desenvolveu o conceito de convergência da renda dos países. A questão da convergência é um tópico central dentro do debate de modelos de crescimento exógenos e endógenos. Segundo esta, se os países tiverem a mesma taxa de crescimento populacional, mesmos parâmetros de gostos e tecnologia, então eles devem ter a mesma renda por trabalhador de *steady state*². Se isso é verdade, então, com o tempo os países tendem a convergir incondicionalmente para um mesmo nível de renda, com os países mais pobres convergindo mais rápido que os mais ricos. No entanto, a teoria neoclássica de crescimento exógeno e a de crescimento endógeno não exigem que esses parâmetros sejam iguais para os países e que por isso a convergência é condicional. Apesar de muito plausível, a teoria da convergência, tanto incondicional quanto condicional, foi muito criticada por ter pouco suporte dos dados (Rodrik [2011]). No entanto, ao contrário do que aconteceu antes de meados da década de 90, é possível verificar um movimento generalizado no mundo de convergência do produto dos países para o produto dos países ricos (figura (1)).

A importância da discussão sobre convergência, incondicional ou condicional, independe de estarmos considerando casos de crescimento endógeno ou exógeno. Essa importância vai além da academia. Os resultados e argumentos dessas teorias motivaram os países do mundo a melhorar e desenvolver o ambiente econômico, macro e micro. O objetivo é criar e melhorar as condições para o crescimento, seja através de um maior produto de *steady state*, seja através de uma maior velocidade e estabilidade para atingir esse nível. De acordo com esses resultados dessas teorias, para se aumentar o produto de *steady state* pode-se usar políticas que diminuam taxa de impaciência da sociedade ou aumentem a propensão marginal a poupar. Para aumentar a velocidade de convergência, a literatura sugere políticas de aumento à eficiência do investimento e do incentivo ao progresso tecnológico, como direito de propriedade e investimento na qualidade do capital humano. Os custos operacionais e de mão de obra influenciam o nível de *steady state* e a velocidade de convergência. Na verdade, nível de *steady state* do produto por trabalhador e velocidade de convergência estão interligados.

Nesse sentido, em termos empíricos, essas condições para melhorar o potencial de crescimento dos países incluem, principalmente, estabilidade econômica (Lin and Kim [2014]), funcionamento institucional (Acemoglu et al. [2002] e Berg et al. [2012]), combate a corrupção (Evrensel [2010]), investimento no desenvolvimento da produtividade da mão de obra (Strulik et al. [2013]) e da infraestrutura. Logicamente, o sucesso dessas políticas é condicional às particularidades do país e ao momento do tempo (Rodrik [2005]), além de depender das distorções da economia (Easterly [2005], Lindauer and Pritchett [2002] e Sala-I-

² Sempre que, no decorrer do trabalho, for mencionado “renda”, nós vamos estar nos referindo à renda “por trabalhador”. O mesmo se aplica quando mencionarmos alguma variável “per capita”. Na verdade vamos estar nos referindo a variável “por trabalhador”.

Martin et al. [2004]) e de choques exógenos, que são responsáveis por uma parte significativa dos ciclos de negócios (Easterly et al. [1993]).

Com o objetivo de responder às mencionadas limitações dos modelos de crescimento exógenos, os trabalhos de modelos de crescimento endógenos testaram muitas variáveis controláveis pelos *policy makers* como determinantes do produto ou renda de um país. Essas variáveis possuem como principal objetivo tentar explicar as forças motoras do progresso tecnológico, qualidade institucional e eficiência do investimento de um país. Essa abordagem tem um apelo mais racional do que a suposição de que esses determinantes são totalmente exógenos e fora do controle dos *policy makers*.

Muitas dessas variáveis, apontadas como significativas nesses trabalhos, foram criticadas, por outros trabalhos, ou por não apresentarem robustez a testes de sensibilidade, ou por possuírem exceções significativas, ou por não possuírem uma relação de causa e efeito muito clara com o produto. Com respeito à crítica de que algumas variáveis não apresentam robustez a testes de sensibilidade queremos dizer que esses resultados são devidos a coincidências econométricas (Levine and Renelt [1992], Sala-I-Martin [1997] e Sala-I-Martin et al. [2004]), amostrais ou de especificação por exemplo. A crítica de exceções significativas destaca que para determinada regra, obtida a partir de determinado resultado significativo, é possível encontrar casos importantes que contradizem esse resultado, como os casos recentes da Índia (Madsen et al. [2010]) ou da China (Rodrik [2010]). O problema, mencionado na terceira crítica, da falta de clareza satisfatória na relação de causa e efeito é clássico na econometria e se aplica em várias outras áreas.

Um bom exemplo de variável significativa nesses trabalhos é a integração do país com o mercado internacional (Barro [1998], Sala-I-Martin et al. [2004], Mayda and Rodrik [2005] e Hausmann et al. [2005]). A abertura da economia permite que o país absorva tecnologia e *know-how* de países mais desenvolvidos. Esse nível de abertura de uma economia, medida das mais diversas formas, é uma variável importante e consistente para explicar variações do produto em trabalhos de painel de países (Levine and Renelt [1992], Rodriguez and Rodrik [2001] e Vamvakidis [2002]). No entanto, existem autores que destacam que esses resultados devem ser vistos com cautela, ou por existirem exceções significativas a regra, ou por apontarem inconsistências metodológicas nessas estimações.

A educação é outro exemplo importante, que se encaixa tanto na primeira quanto na terceira crítica. A educação foi apontada como a maior falha ou omissão dos modelos clássicos de crescimento (Barro [1998]). No entanto, o significativo investimento em educação realizado ao redor mundo desde o “descobrimento” dessa variável antes ignorada tem gerado resultados aquém do esperado (Bosworth and Collins [2003]).

No que se refere à terceira crítica, de que não é clara a relação de causa e efeito de algumas variáveis com o produto, nós destacamos o investimento. Em termos de teoria, o investimento é uma variável importante para explicar o crescimento desde os modelos de crescimento exógeno porque é o principal determinante da acumulação de capital da economia. Por sua vez, a acumulação de capital é importante não só em termos de crescimento, mas também porque a riqueza de um país é definida, em grande parte, pelo capital acumulado³. Em termos de trabalhos empíricos, o investimento é uma variável consistentemente significativa em explicar o crescimento dos países no longo prazo (Barro [1991] e Hausmann et al. [2005]). No entanto, apesar desse embasamento tanto teórico quanto empírico, existem, na literatura, dúvidas a respeito da relação de causalidade entre produto e investimento (William [1999], Podrecca and Carmeci

³ Um país com maior produtividade da mão de obra vai crescer mais, mas um país com maior nível de capital acumulado vai ser mais rico e proporcionar uma estrutura melhor para seus habitantes.

[2001] e Sala-I-Martin [2002]). Como o investimento tem um papel importante, tal como mencionado, na dinâmica do produto de um país, essa relação de causa e efeito é uma questão importante tanto para a literatura como para os *policy makers* dos países.

Sendo assim, dado a relevância do assunto, essas dúvidas na relação entre investimento e produto são as motivações principais e o ponto de partida do trabalho. Com isso, o esclarecimento dessa associação entre essas variáveis é o objetivo principal do trabalho.

A partir desse *link* entre produto e investimento montamos uma estrutura de equações simultâneas que nos permite estudar outras conexões, entre outras variáveis, da economia.

Na verdade testamos vários sistemas, com diferentes relações estruturais entre as variáveis. Essas múltiplas estimações têm o objetivo de definir e encontrar os *links* mais significativas, entre as variáveis consideradas no trabalho, para cada país. Mais especificamente, testamos todas as combinações de relações estruturais, contemporâneas, entre as variáveis endógenas segundo uma hipótese de ligação entre os grupos de variáveis do trabalho. Esses grupos de variáveis controlam para o produto e o investimento, para a poupança e para os preços da poupança e do investimento.

Esses sistemas de equações foram estimados por máxima verossimilhança para cada país da amostra. A partir dessas estimações escolhemos, dentre todos os modelos estimados para cada país, os 100 melhores de acordo com o critério do BIC. Para fazer essas estimações nosso principal problema foi a disponibilidade de séries temporais longas o suficiente. Na verdade, os dados estavam disponíveis para no máximo 20 anos em uma base anual, o que leva, na melhor das hipóteses, a 20 observações. No entanto, para alguns países importantes da amostra, como o Brasil, a disponibilidade de dados é mais restrita. Sendo assim, adotamos uma estratégia de utilizar dados simulados a partir de uma distribuição de probabilidade conjunta estimada por máxima verossimilhança para cada país. Essa estimativa usou a massa de dados observada onde o ajustamento (*fit*) das distribuições foi maximizado utilizando como critério o BIC.

Nossos resultados têm três conclusões principais. A primeira é que nossas estimações indicam que, a relação de causalidade do investimento para o produto é mais frequente que a direção oposta. Esse resultado está de acordo com a teoria tradicional. No entanto, por haver casos de países onde a relação de causalidade aponta para a direção do produto para o investimento, as dúvidas sobre a relação de causa e efeito que motivaram esse artigo se justificam.

Nossa segunda conclusão é que existe muita heterogeneidade na dinâmica de funcionamento interno dos países e nos canais de transmissão de juros, poupança, investimento e produto. Essas diferenças no funcionamento interno das economias são explicadas pelos efeitos de choques e por distorções internas e qualidade das instituições. Nesse sentido, a presença dessas diferenças compromete o desenvolvimento de teorias robustas e gerais. Essas diferenças, evidenciadas em nossos resultados, também apontam para outras relações descritas na teoria de uma forma, mas que os dados mostram que essas a dinâmica podem ser outra.

A terceira conclusão é que, apesar dessas heterogeneidades, foi possível encontrar, em nossos resultados, uma característica robusta que agrupa os países. Esse principal critério para agrupar as economias ou a característica que mais assemelham os países é como a economia reage a variações do humor do investidor internacional. Esse efeito foi medido e verificado diretamente ou através dos fluxos de capitais internacionais e da taxa de câmbio. Além do mais, relacionada a essa análise, verificamos que, quanto mais integrado o país é ao mercado internacional, maior é a eficiência do investimento em influenciar o

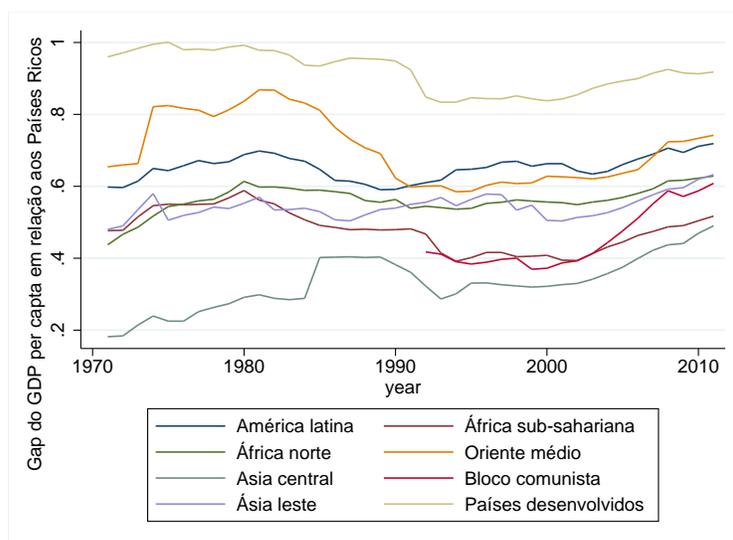
produto.

Especificamente no que se refere ao Brasil e aos países desenvolvidos, enquanto que os ciclos econômicos do Brasil dependem muito dos fluxos de capitais internacionais, os ciclos dos países ricos são pouco influenciados pelo humor dos investidores internacionais.

Dito isto, nosso trabalho contribui com a literatura no sentido de que nossos resultados exploram uma questão significativamente importante tanto para a academia quanto para questões práticas dos *policy makers*. No entanto, nossas análises vão além do objetivo principal do trabalho e fornecem evidências sobre os mecanismos de transmissão dos juros, poupança, investimento e produto. Mais ainda, ao invés de derivar resultados genéricos ou médios, como em trabalhos de painéis de países, nós exploramos as particularidades de cada país individualmente. Sendo assim, todas essas avaliações são enriquecidas pela comparação de diversos países com suas especificidades, que representam diferentes experiências.

O capítulo está estruturado da seguinte forma. Após essa revisão teórica e motivação do artigo, tratamos da metodologia e estimação. A metodologia se divide em construção da distribuição conjunta dos dados por país, definição e seleção dos sistemas de equações simultâneas e processo de agrupamento dos países de acordo com o funcionamento da economia. Na seção sobre resultados nós tratamos primeiro especificamente das equações do produto e do investimento, dentro dos sistemas de equações, e em segundo do processo de clusterização dos perfis dos países. Por fim vamos descrever as conclusões finais.

Figura 1 – Hiato do produto dos países do mundo



NOTA: A figura mostra o hiato do produto dos países do mundo em relação ao produto dos países ricos, informados pelo Banco Mundial, ao longo do tempo. As linhas mostram a média simples dos gaps dos países individuais, dentro de cada grupo destacado na legenda, ao invés de ponderá-la pelo tamanho do produto dos países. Nele é possível ver que antes de meados dos anos 90 havia pouca evidência de que suportasse a teoria da convergência da renda dos países. Esse fato fundamentou muitas críticas a esse argumento. No entanto, após meados dos anos 90 pode ser visto um movimento praticamente generalizado de convergência do gap dos produtos dos países do mundo.

3.2 Metodologia e Estimação

Nossa metodologia e estimação para abordar o problema do trabalho consistem em estimar, para um conjunto de hipóteses, todas as combinações de estrutura de sistema de equações simultâneas, escolher os melhores modelos, montar o perfil dos países e agrupar os países em *clusters* e em fatores comuns. A partir desses *clusters*, teremos não só uma boa idéia, em termos globais, da relação entre produto e investimento, como das similaridades do funcionamento interno das economias na sua relação entre investimento, poupança e preço do investimento e da poupança.

Dessa forma, nós vamos tentar estabelecer, em um primeiro momento, a relação de causalidade entre investimento e produto para todos os países da amostra. Em um segundo momento, nós teremos informações sobre, não só a relação entre investimento e produto, mas a relação entre estas primeiras variáveis e a poupança e os preços de poupar e investir.

Para definir essas ligações entre essas diversas dimensões de uma economia, nós vamos estimar, para cada país, uma série de modelos com relação contemporânea entre as variáveis. Dado o elevado número de parâmetros em cada estimação e um número pequeno de observações disponíveis na amostra nós vamos simular um conjunto de dados, a partir da distribuição conjunta dos dados observados, grande o suficiente para atender nossas necessidades. Essa distribuição conjunta das variáveis, de cada país, é estimada por máxima verossimilhança e selecionada dentre um conjunto finito de possibilidades.

Feito isso, o próximo passo é montar o perfil dos países. Para tanto nós calculamos, para cada coeficiente que poderia ser estimado, a média dos coeficientes dos 100 melhores modelos de cada país, que fossem diferentes de 0 em pelos menos 50 modelos⁴. Também incluímos mais uma regra que atribui valor igual a 0, no perfil do país, para aquele coeficiente que inverte o sinal em menos de 2 desvios padrão a partir da média dos coeficientes significativos⁵.

Além dos coeficientes médios do sistema de equações, de acordo com as regras acima, o que nós classificamos como perfil dos países também inclui (i) o BIC médio dos 100 melhores modelos, (ii) o desvio padrão dos BIC dos 100 melhores modelos, (iii) a renda por trabalhador do país e (iv) população do país entre 14 e 64 anos. Apesar de não fazer parte do conjunto de coeficientes dos sistemas de equações dos países, nós julgamos essas quatro informações como significativas para agrupar países com dinâmicas internas similares⁶.

Com o perfil de cada país, nós fizemos algumas análises de *clusters* para agrupar os países por similaridade no perfil e, conseqüentemente, da dinâmica da economia.

3.2.1 Dados, variáveis e amostra

Nossa amostra é composta por 101 países com dados observados em base anual para o período de 1990 até 2012. O período de tempo foi definido em função da disponibilidade de dados para as variáveis utilizadas

⁴ Se um coeficiente, para uma variável, não fosse significativo em pelo menos metade dos 100 melhores modelos para aquele país, então o coeficiente não é consistente e não representa uma relação robusta na dinâmica do país.

⁵ Se a distribuição dos coeficientes significativos estimados, para uma variável em uma equação, inclui o valor zero para um intervalo de até 2 desvio padrão, a partir da média, para esse coeficiente, então consideramos que esse coeficiente é igual a 0. Isso porque se a afirmação acima é verdadeira, então não podemos estabelecer a relação, entre a variável explicativa e a endógena no sistema, como positiva ou negativa.

⁶ Quando nos referirmos a “perfil dos países” no decorrer o trabalho vamos omitir essas 4 variáveis e nos referir apenas aos coeficientes da sistema de equações simultâneas com objetivo de simplificar a exposição. No entanto, é importante destacar que essas variáveis fazem parte do “perfil dos países”.

para os países julgados como interessantes, ou seja, que deveriam fazer parte da amostra. Se utilizássemos dados de antes de 1990, não poderíamos incluir, nas análises, os países do antigo bloco comunista. Isso não seria desejável, a nosso ver, pois além de serem países importantes mundialmente em termos de renda, também formam o grupo que mais reduziu o gap do produto em relação aos países ricos mostrado na figura (1).

Esses 101 países foram escolhidos por terem pelo menos 14 observações completas das variáveis do trabalho. Uma observação completa é um ano, período, em que todas as variáveis foram observadas em primeira diferença. O *threshold* de 14 observações foi definido porque esse é o número de observações que o Brasil tem disponível e consideramos que esse país deveria fazer parte do trabalho.

O trabalho usa 7 variáveis endógenas e 4 exógenas, em primeira diferença, nos sistemas de equações, tal como descrito na tabela (2). A principal fonte de dados é o Banco Mundial.

Tabela 2 – Variáveis do trabalho por grupo

Variável	Grupo	Medida	Descrição	Fonte
produto (<i>pib</i>)	1	anual	produto interno bruto	Banco Mundial
investimento (<i>invest</i>)	1	anual	formação bruta de capital fixo	Banco Mundial
poupança externa (<i>nx</i>)	2	anual	saldo da balança comercial	Banco Mundial
investimento direto estrangeiro (<i>fdi</i>)	2	anual	componente da conta capital	Banco Mundial
crédito doméstico (<i>cred</i>)	2	anual	crédito doméstico fornecido pelo sistema financeiro	Banco Mundial
câmbio (<i>e</i>)	3	média anual	moeda local em relação ao dolar americano	Banco Mundial
juros (<i>r</i>)	3	anual	taxa de juros bancária *	Banco Mundial
VIX médio (<i>VIXm</i>)	exo	média anual	média do VIX no ano	CBOE***
VIX desvio padrão (<i>VIXsd</i>)	exo	std anual **	desvio padrão do VIX no ano	CBOE***
produto mundial (<i>pib_m</i>)	exo	anual	produto do mundo	Banco Mundial
Inflação Mundial (<i>pi_m</i>)	exo	anual	deflator do produto do mundo	Banco Mundial

NOTA: A tabela mostra as variáveis utilizadas nas estimações do trabalho. A base temporal de mensuração é anual. Todas as variáveis foram utilizadas nas estimações em primeira diferença. * A taxa de juros do país foi escolhida como a taxa de juros de crédito do sistema bancário porque essa é a série de juros mais completa disponível no Banco Mundial, ao contrário de outras como os juros de títulos públicos ou juros base da economia. ** Desvio padrão anual. A base de dados do VIX, tanto para o cálculo da média quanto do desvio padrão, é diária. *** Chicago Board Options Exchange.

Assim, pela tabela (2), uma observação completa, no tempo t do país c é dada pelo vetor descrito na equação (3.1). Um *drop*, para cada país c , da distribuição conjunta das variáveis desse país, também será um vetor equivalente a esse, mas sem o indexador do tempo (t) e sim um indexador do número do *drop* (k).

$$(pib_{c,t}, i_{c,t}, nx_{c,t}, fdi_{c,t}, cred_{c,t}, e_{c,t}, r_{c,t}, VIXm_{c,t}, VIXsd_{c,t}, pib.m_{c,t}, pi.m_{c,t}) \quad (3.1)$$

As 7 primeiras variáveis da tabela (2) são as variáveis que controlam para as dinâmicas internas do país de interesse no estudo. As quatro últimas, grupo *exo*, controlam para fatores internacionais, como a aversão ao risco ($VIXm_{c,t}$ e $VIXsd_{c,t}$), ciclos mundiais ($pib.m_{c,t}$) e alterações nos preços internacionais ($pi.m_{c,t}$). Essas alterações nos preços podem tanto influenciar o preços das exportações e importações, como indicar períodos de maior instabilidade mundial.

As variáveis econômicas ($pib_{c,t}, i_{c,t}, nx_{c,t}, fdi_{c,t}, cred_{c,t}, pib.m_{c,t}$) foram calculadas per capita, para a população entre 14 e 64 anos. No caso das variáveis $pib_{c,t}, i_{c,t}, nx_{c,t}, fdi_{c,t}$ e $cred_{c,t}$ utilizamos a população do país e para $pib.m_{c,t}$ usamos a população mundial.

Um último aspecto que merece ser comentado é que não deflacionamos as variáveis econômicas. Não deflacionamos porque, no período analisado alguns países importantes, como Brasil e Rússia, por exemplo, atravessaram períodos inflacionários significativos. Esses períodos apresentaram taxas de inflação muito

elevadas que tendem a distorcer e subavaliar as variáveis econômicas deflacionadas desses países. Assim, para controlar para as variações de preços dos países ao longo do tempo, incluímos a taxa de câmbio doméstica (e_c), que é a razão dos preços de duas moedas. Essa taxa de câmbio está no grupo das variáveis mais exógenas, dentre as endógenas, do sistema de equações. Isso quer dizer que ela entra como variável explicativa nas equações de todas as demais variáveis endógenas (em qualquer combinação utilizada). Isso é suficiente para controlar para as variações de preços nas variáveis econômicas.

3.2.2 Distribuição conjunta e simulações

A distribuição conjunta das variáveis, que são usadas nas estimações, é obtida ao se combinar as distribuições marginais independentes de cada variável. Essas distribuições marginais são escolhidas dentre um conjunto de distribuições contínuas utilizando o BIC como critério de melhor ajustamento (*fit*) e são combinadas através de uma função de Cópula. Segundo o teorema de Sklar, se as distribuições marginais são contínuas e uniformemente distribuídas, então existe apenas uma função cópula C que garante a igualdade entre a distribuição conjunta que representa os dados, F , e as distribuições marginais de cada variável (F_1, \dots, F_i).

A idéia inicial de cópulas foi desenvolvida por Sklar em 1959. Segundo este, uma função distribuição multivariada F pode ser decomposta em duas partes: (i) nas funções distribuições marginais F_i e (ii) em uma função cópula C que descreve a relação ou dependência entre as distribuições marginais.

$$F = C(F_1, \dots, F_i),$$

Como a função distribuição de probabilidade (CDF) de uma variável é uniformemente distribuída entre 0 e 1, podemos transformar qualquer distribuição em outra aplicando a inversa da CDF objetivo em uma distribuição uniforme gerada a partir da CDF da distribuição de origem. O caminho oposto também é possível. Assim, se podemos calcular a função distribuição acumulada, o teorema de Sklar se aplica a qualquer distribuição marginal utilizada.

Dessa forma, estimamos por máxima verossimilhança, para cada distribuição contínua testada, os parâmetros desta distribuição utilizando a massa de dados observada. A partir da distribuição de probabilidade calculamos o BIC. A distribuição marginal de cada variável para cada país é definida (escolhida), dentro do conjunto de distribuições testadas, com base no critério de melhor BIC ($BIC = -2 * \ln(LL) + k * \ln(n)$, onde LL é a verossimilhança, k é o número de parâmetros estimados e n é o número de observações)⁷.

No entanto, antes de realizar as estimações das distribuições marginais, aplicamos duas transformações nas variáveis (já per capita se for o caso). Primeiro nós retiramos a tendência de forma linear ao regressir cada variável, de cada país, contra o “ano” ou “tempo”⁸. Segundo nós diferenciamos as variáveis. Com isso os dados tendem a ser centralizados no 0 e com uma distribuição relativamente bem comportada nas duas direções. Dessa forma é mais fácil obter boas estimativas das distribuições marginais, da função cópula e conseqüentemente da distribuição conjunta.

⁷ As distribuições contínuas que testamos são: Beta, Birnbaum-Saunders, Exponential, Extreme value, Gamma, Generalized extreme value, Generalized Pareto, Inverse Gaussian, Logistic, Log-logistic, Lognormal, Nakagami, Normal, Rayleigh, Rician, t location-scale, Weibull.

⁸ A “tendência” é o coeficiente (β_{tempo}^i) de uma variável que mede tempo (no caso o “ano”) em uma regressão da variável objetivo (y_t^i), que quer tirar tendência, sobre o tempo ($y_t^i = \beta_0^i + \beta_{tempo}^i * ano + \varepsilon_t^i$). Assim, a variável sem tendência seria $y_t^{i,*} = \beta_0^i + \varepsilon_t^i$

Antes de estimar a função de Cópulas transformamos todas as distribuições estimadas no passo anterior em distribuições normais. Fizemos isso calculando, a partir da distribuição marginal escolhida, a função densidade acumulada em cada ponto observado da amostra. A distribuição normal é obtida aplicando a inversa da função distribuição acumulada da normal nessa função em cada ponto.

A partir da distribuição conjunta de cada país fizemos 10.000 *drops* iguais ao vetor da equação (3.1) para serem usados na estimação do sistema de equações simultâneas. O número 10.000 foi escolhido porque com esse número ao mesmo tempo (i) verificamos que é grande o suficiente para que praticamente todos os coeficientes do sistema sejam significativos e (ii) é pequeno o suficiente para garantir a velocidade dos códigos envolvidos em todos os processos.

3.2.3 Modelo estrutural e identificação

O modelo estrutural tem como objetivo principal estudar a relação de causalidade entre produto e investimento. Essas duas variáveis chamamos de grupo 1 e são as variáveis mais endógenas do trabalho. Isso quer dizer que essas duas variáveis são explicadas por todas as demais. O sistema como um todo possui 7 variáveis endógenas e 4 exógenas.

Para explicar o investimento e o produto usamos a poupança do país e os preços do investimento e da poupança. O conjunto de variáveis que controlam para a poupança do país chamamos de grupo 2. Também usamos um conjunto de variáveis que controlam para os preços do investimento e da poupança que chamamos de grupo 3.

As variáveis do grupo 2 são exógenas às variáveis do grupo 1, mas são endógenas às variáveis do grupo 3. Isso quer dizer que, por um lado, a poupança do país explica o investimento e o produto e, por outro, é explicada pelos preços do investimento e da poupança.

As variáveis do grupo 3 são exógenas às variáveis do grupo 1 e 2, ou seja, os preços da poupança e do investimento explicam o produto, o investimento e a poupança. Por outro lado não é explicado por nenhuma dessas variáveis, ou grupo de variáveis.

Seja P o grupo 3 que controla para os preços do investimento e da poupança, S o grupo 2 que controla para a poupança e I o grupo 1 que inclui o investimento e o produto. Nas equações abaixo são mostradas, de forma esquemática, as relações de exogeneidade e endogeneidade entre os grupos. As variáveis a direita das setas são as endógenas na relação e as da esquerda são as exógenas. Assim, o grupo 3 explica o grupo 1 e 2 (equação (3.2)), enquanto o grupo 2 explica o grupo 1 (equação (3.3)).

$$P \rightarrow S \rightarrow I, \quad (3.2)$$

$$P \rightarrow I, \quad (3.3)$$

Independente das relações de endogeneidade e exogeneidade entre os grupos, descritas acima, testamos todas as combinações de endogeneidade e exogeneidade das variáveis dentro do mesmo grupo. Isso quer dizer que, dentro de cada grupo, testamos todas as combinações de coeficientes a serem estimados.

Por fim temos quatro variáveis que são exógenas a todas as outras variáveis e fazem parte do grupo *exo*. Essas variáveis do grupo *exo* são exógena ao sistema e ao país. São variáveis que controlam para os

ciclos mundiais. A tabela (2) mostra a composição de cada grupo e descreve as variáveis utilizadas no trabalho.

A especificação do modelo estrutural (equação (3.4)) começa com uma matriz triangular superior onde cada linha é uma equação de uma variável endógena do sistema. Essa é a matriz A_c de dimensão 7×7 (variáveis endógenas) no sistema de equações (3.4), onde c indexa o país. A diagonal principal dessa matriz (A_c) é formada por uns (números 1) que representam a variável endógena dependente da equação. Os componentes acima dessa diagonal são os coeficientes estruturais a serem estimados. Optamos por uma matriz triangular superior para simplificar as combinações testadas. É possível obter a identificação dos parâmetros estruturais com outras especificações, mas essa flexibilização iria aumentar muito o número de combinações testadas e conseqüentemente o tempo de processamento.

Além desta matriz (A_c) existe uma matriz com os coeficientes das variáveis do grupo *exo* em cada equação que também serão estimados. Essa matriz é a B_c de dimensão 4×7 (4 variáveis exógenas em 7 equações de variáveis endógenas) no sistema de equações (3.4), onde c indexa o país.

$$A_c Y_{c,t} = B_c X_{c,t} + \varepsilon_{c,t}, \quad (3.4)$$

A matriz A_c triangular superior faz com que o sistema seja exatamente identificado. No entanto, essa especificação não necessariamente é a melhor. Assim, testamos até 3 coeficientes iguais a zero na parte superior da matriz desde que a matriz A_c respeite a condição de *rank* para que o sistema seja identificado⁹. O problema de aumentar o número de coeficientes iguais a 0 na matriz A_c é que aumenta o problema de sobre-identificação e isso não é desejado. A escolha de 3 coeficientes iguais a 0 foi em função também do tempo de processamento. Um número maior iria aumentar muito o tempo de processamento de estimação das combinações.

3.2.4 Fatoração, seleção de variáveis e clusterização dos betas

De posse dos coeficientes estimados (betas), para cada variável dependente em cada equação do sistema (3.4) para os 100 melhores modelos para cada país, calculamos a média desses coeficientes para aqueles que tiveram valores diferentes de 0 em mais de 50% desses 100 melhores modelos.

Com o perfil de cada país na relação entre produto, investimento, poupança e preço do produto e investimento (dado por um vetor equivalente à equação (3.1) fixo no tempo para cada país), calculamos uma matriz com as distâncias euclidianas entre os perfis de todos os países.

Os *clusters* foram construídos por hierarquização com o método de Ward. O método de Ward consiste em otimizar (minimizar) a variância dentro de cada *cluster*. Para construção do *cluster*, por hierarquização, iniciamos o processo no primeiro passo com n *clusters* de 1 indivíduo (país), onde n é o número de países. No segundo passo, computamos a soma do quadrado das distâncias euclidianas para todas as combinações 2 a 2, dos perfis dos países. Isto consiste na construção de uma matriz triangular $n \times n$. Dentro dessa matriz escolhemos o par que tem a menor soma do quadrado das distâncias. Escolhemos essa primeira dupla e formamos o primeiro *cluster*. Assim, ao fim do segundo passo, temos $n - 1$ *clusters*,

⁹ A condição de *rank* é necessária e suficiente para garantir a identificação em um sistema de equações simultâneas. Segundo esta, dadas as restrições em uma determinada equação (coeficientes iguais a 0 ou fixos), não se deve impor restrições excessivas nas demais equações do sistema, para as variáveis com restrição na equação de referência. De forma mais direta, para cada equação do sistema, a matriz formada pelas colunas, nas demais equações, relativas às variáveis excluídas dessa equação, deve ter determinante diferente de 0.

com 1 *cluster* com 2 países e o resto com 1 país. No terceiro passo realizamos o mesmo procedimento para todas as combinações de 2 a 2 (com um *cluster* com 2 países) e formamos $n - 2$ *clusters*. O algoritmo continua até termos 1 *cluster* com n (todos os) países.

Nosso objetivo aqui é agrupar os países com perfil similar do sistema de equações simultâneas, ao mesmo tempo em que destacamos os grupos com perfis diferentes.

Escolhemos o método de Ward, porque, dentro da literatura, ele é um dos dois que, tradicionalmente, apresenta os melhores resultados. Uma crítica ao método de Ward é que ele tende a gerar *clusters* com o mesmo número de indivíduos. Uma boa alternativa, e o outro método que na literatura apresenta os melhores resultados, é o *average linkage* que pode destacar grupos com minorias. Este método calcula a distância entre as médias dentro de cada grupo. Mas isso também não é muito desejável, pois existe uma tendência a sobrevalorizar as individualidades dos países. Na verdade, o uso do procedimento de *average linkage* em nossas estimativas não gerou resultados interessantes.

Esse procedimento é feito, tal como descrito no tópico de resultados, para 3 conjuntos distintos de variáveis, ou subconjuntos do perfil total (com todos os coeficientes) dos países.

Além da análise de *clusters*, nós vamos usar a análise de fatores para definir grupos de países similares. Apesar das duas técnicas poderem ser utilizadas para agrupar indivíduos, elas o fazem utilizando diferentes critérios. Enquanto que na análise de fatores estamos preocupados com as correlações entre as variáveis dos perfis dos países, na análise de *clusters* nós vamos focar em uma medida de distância, que em nosso caso é a distância euclideana. A análise de fatores agrupa os indivíduos com base na correlação e a análise de *cluster* com base na medida de distância ou similaridade. Dessa forma, os resultados, em termos de agrupamento de países, na análise de fatores e de *clusters* não necessariamente são os mesmos. No entanto, como vamos usar variáveis normalizadas (numa mesma escala) tanto na análise de *clusters* como de fatores, nós esperamos que os resultados da análise do co-movimento e das distâncias sejam similares.

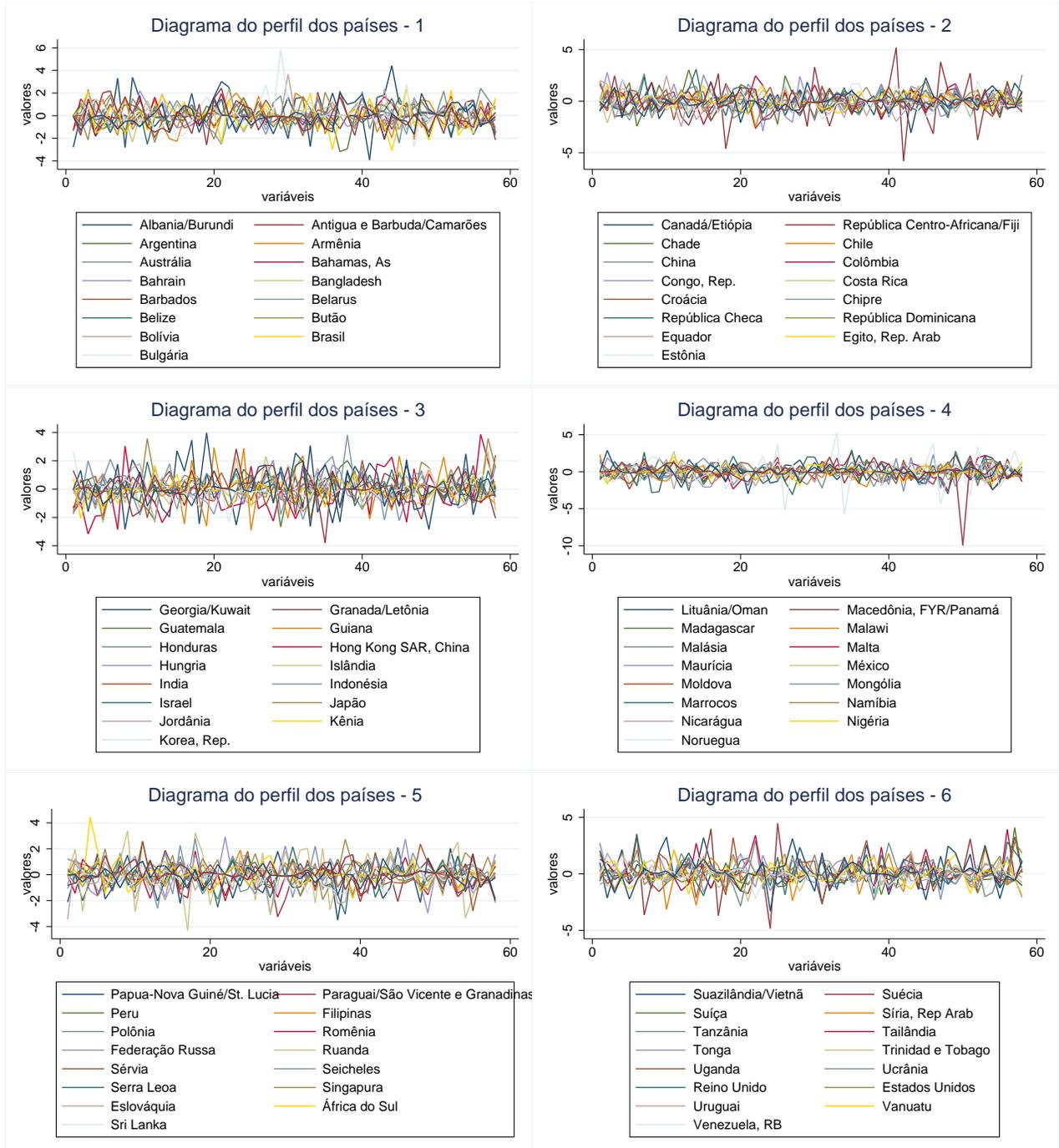
O problema do número de *clusters* é particularmente importante porque, dentro da análise de *clusters* sua definição é discricionária e fonte de críticas. Para contornar o problema e objetivar a solução, nós vamos utilizar o número de fatores com autovalores maiores que 1, estimados na análise de fatores com o mesmo conjunto de dados, para definir o número de clusters nesse conjunto de dados. Faremos isso porque, ao usar variáveis com a mesma escala, esperamos que os resultados nessas duas análises fossem similares.

Por serem baseados em medidas de distância, os resultados da análise de *clusters* são particularmente sensíveis a *outliers*. Nesse sentido, com o objetivo de identificar, visualmente, países com valores extremos nas variáveis, mostramos na figura (2) o perfil dos países em diagramas. No eixo horizontal estão as variáveis e no vertical os valores. Assim, cada linha mostra o comportamento de cada país ao longo das variáveis. Note que, devido a limitações do número de países, algumas cores são usadas para 2 países.

Assim, de acordo com a figura (2) podemos identificar como países com valores atípicos (linhas destoantes): Belize e República Africana Central (Diagrama 1), Guiana (Diagrama 2), Noruega (Diagrama 4) e Suécia (Diagrama 5). De fato, nas análises feitas esses países tendem a estar sozinhos em *clusters* separados.

Um último aspecto importante relacionado a essas duas análises é o peso ou a importância de cada variável para definir se um país faz parte de cada fator e de cada *cluster*. No caso da análise de fatores esse peso vai ser dado diretamente e objetivamente pelos *fator scoring*. Para a análise de *clusters* nós não

Figura 2 – Diagrama de Perfil dos países



NOTA: As figuras mostram o diagrama do perfil dos países. No eixo horizontal estão as variáveis e no eixo vertical estão mensurados os valores delas. Cada linha representa um país e mostra os valores, para esse país, de todas as variáveis consideradas ao longo do eixo horizontal. Como as variáveis foram normalizadas então é possível compará-las e uma mesma escala. O objetivo é identificar países outliers, que seriam representadas por linhas com valores extremos nas variáveis, que podem distorcer a análise de clusters.

temos os pesos. Assim, nós vamos usar a média dos valores de cada variável, dos países pertencentes ao *cluster*, para identificar as características importantes para cada *cluster*.

3.3 Resultados

Para avaliar os resultados das estimações dos sistemas de equação (3.4) e das análises de fatores e *clusters*, vamos organizar a exposição em tópicos. Primeiro vamos analisar as equações de pib_t e $invest_t$ nos sistemas de equações. Essa análise está diretamente relacionada ao nosso objetivo principal.

Depois, vamos tentar agrupar o perfil dos países por similaridade e fazer as análises de *clusters* e de fatores propriamente ditas. Essa abordagem tem como objetivo identificar similaridades no comportamento e dinâmica dos países.

Tanto a primeira quanto a segunda análise são feitas com base no perfil de cada país. O perfil de cada país mostra a média de cada coeficiente do sistema de equações para os 100 melhores modelos, quando o coeficiente é significativo em pelo menos 51 modelos. É importante destacar que um coeficiente é igual a 0, no perfil do país, ou porque ele não foi significativo em nenhum dos 100 melhores modelos ou ele foi significativo em menos de 51 modelos.

A tabela (3) mostra o total de combinações testadas. Para construir o total de combinações realmente testadas primeiro construímos todas as combinações de estruturas da matriz, A_c (equação (3.4)), com até 3 zeros na parte superior e respeitando a ordem de *rank*. Depois disso definimos todas as combinações de ordenamento das variáveis endógenas seguindo a ordem de causalidade dos grupos descritas no tópico “Modelo Estrutural e Especificação”. Por fim tivemos que limpar os sistemas de equações repetidos. Isso porque, ao permitir até 3 coeficientes iguais a zeros na parte superior da matriz A_c , então, mesmo que mudemos a ordem das variáveis endógenas em $Y_{c,t}$, na equação (3.4), existe a possibilidade que ocorra sistemas de equações iguais.

Tabela 3 – Relações entre Produto e Investimento

Estrutura	Combinações
Matriz superior A_c^*	986
Ordem das variáveis endógenas	24
Total	23.664
Líquido de repetições	17.779

NOTA: A tabela mostra o total de combinações testadas para cada país. A primeira linha mostra o total de combinações de estruturas da matriz A_c , triangular superior com até 0 zeros na parte superior, que respeita a ordem de rank. A segunda linha mostra o total de combinações da ordem das variáveis endógenas respeitando a regra de formação da ordem de endogeneidade das variáveis. A linha com “total” mostra o total de combinações por país. A última linha mostra o total efetivo de combinações testadas por país, controlada para as repetições de equações.

3.3.1 Relação entre Investimento e Produto

As nossas estimações têm muitos resultados e para focar em nosso objetivo principal, vamos isolar nessa seção as equações do investimento e do produto. Essas duas equações são mostradas na (4), da tabela (5) e tabela (6) para o perfis dos países.

Nessas tabelas temos, para cada país, duas linhas. Em cada linha estão discriminados os coeficientes das variáveis explicativas para as equações do produto (pib_t) e do investimento ($invest_t$). Nas colunas estão as variáveis explicativas indicadas no cabeçalho. Como essas duas variáveis são as mais endógenas do sistema de equações, então todas as demais variáveis do trabalho foram estimadas como explicativas.

Vale destacar que apenas indicamos os coeficientes significativos. Quando um coeficiente não foi significativo deixamos a célula em branco ao invés de usar o valor 0.

O fato de haver ou coeficientes de pib_t explicando $invest_t$ ou vice versa se deve a estrutura triangular da matriz A_c na equação (3.4).

Essas tabelas possuem muita informação o que dificulta tirar muitas conclusões ou informações delas. O que podemos destacar é que, na relação de causa efeito entre investimento e produto, existem mais coeficientes que definem uma relação de causalidade do investimento para o produto (diferentes de 0) do que o contrário. Na verdade essa conclusão pode ser vista mais facilmente na figura (3).

Uma segunda conclusão que se pode tirar da tabela (4), da tabela (5) e tabela (6) é que o perfil dos países é muito diferente, o que dificulta ainda mais tirar conclusões a partir dessas tabelas. Se olharmos apenas para a relação de causalidade entre investimento e produto, vemos que existem casos em que pib_t explica $invest_t$, vice versa e casos em que não existe relação entre essas duas variáveis.

Na figura (3), tal como comentado, podemos ver a relação de causalidade entre pib_t e $invest_t$ de forma mais visual. A primeira observação que se pode ser feita é que, tal como comentado, existem mais coeficiente diferentes de 0 de pib_t na equação de $invest_t$ do que o contrário.

No entanto, o objetivo dessas figuras é avaliar a relação entre a causalidade do pib_t e $invest_t$ com a sensibilidade dessas duas variáveis aos ciclos mundiais, controlado pelas variações no produto mundial ($pib.wld_t$) no sistema de equações. Na figura do lado direito, analisamos a equação do investimento. No eixo vertical temos o coeficiente do pib_t na equação do $invest_t$ ($invest(pib)$) e no eixo horizontal nós temos o produto mundial ($pib.wld_t$) na equação do investimento ($invest(pib.wld)$). Na figura do lado direito usamos a mesma lógica, mas na equação do produto.

O que mais chama a atenção nessas duas figuras é a existência de casos em que relação entre produto e investimento, ou vice versa, é negativa para países importantes. No caso do produto causando o investimento vale destacar essa relação inversa para a Austrália, Índia, Argentina e Brasil. Isso faria sentido para países onde a economia tem uma dinâmica menos organizada como Brasil, Argentina e Índia. No caso do investimento causando o produto, os casos importantes onde essa relação é negativa são a China, a Rússia e o Canadá. Uma forma de ver esses resultados inversos do esperado é o possível efeito de distorções na economia ou choques depois dos anos 90. Por outro lado, uma outra explicação coerente para esse comportamento é que o efeito mais significativo dessas relações de causalidade é diferido no tempo. Assim, ao estimar o efeito contemporâneo capturamos apenas uma parte do impacto de uma variável na outra.

Uma outra questão que chama atenção é que, quando significativo, o efeito do produto no investimento ($invest(pib)$) é, na grande maioria das vezes, positivo. Por outro lado, no caso do investimento causando o produto ($pib(invest)$) essa relação é menos evidente pois existem muitos casos positivos, mas também tem muitos casos negativos.

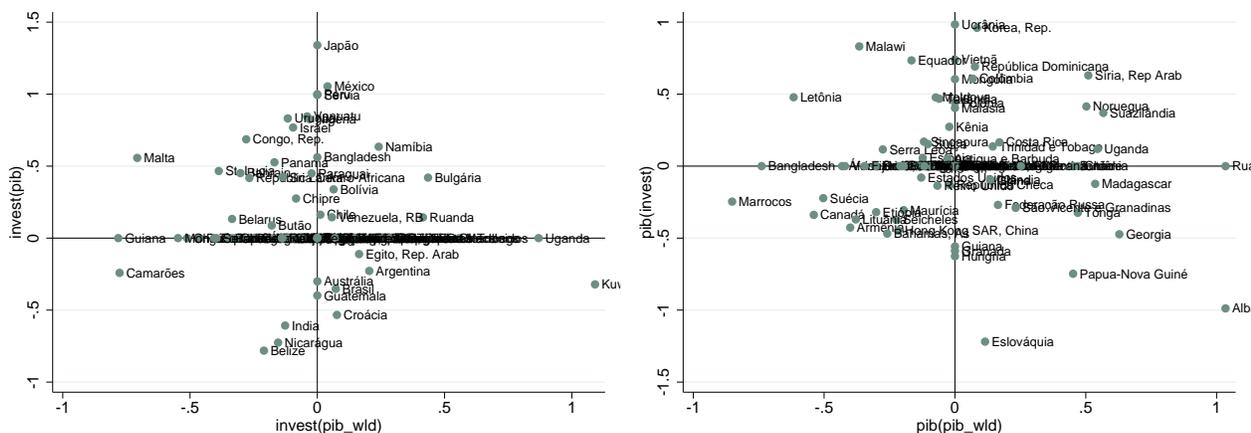
Também é importante notar que, na maioria dos países, existe uma relação positiva entre os ciclos dos países ($pib(invest)$) e do mundo ($pib(pib_wld)$). Isso quer dizer que, na maioria dos casos, quando o mundo

está crescendo, os países estão crescendo também e vice versa. No caso do investimento ($invest(pib_wld)$) isso também é verdade, mas em menor escala, pois existem, proporcionalmente, muitos casos em que essa relação tem sinal negativo.

Uma forma interessante de analisar a relação de cada par de variáveis é estabelecer uma linha de tendência em cada uma das duas figuras. No caso da equação do investimento ($invest(pib)$ e $invest(pib_wld)$) essa linha é ligeiramente negativamente inclinada. Isso quer dizer que quanto mais sensível o investimento é aos ciclos mundiais, menos sensível ele é ao ciclo do país. Ou seja, quanto mais sensível a questões externas, menos sensível o investimento é a questões internas.

Por outro lado, para a equação do produto ($pib(invest)$ e $pib(pib_wld)$) essa linha de tendência é ligeiramente positiva. Assim, assumindo a relação mais frequente em que o investimento causa o produto, quanto mais sensível, ou conectada, aos ciclos mundiais uma economia é, mais eficiente (mais sensível $invest(pib)$) é o investimento em afetar o produto. Essa é uma conclusão importante, pois a eficiência do investimento é um dos determinantes do crescimento, pois mede o quanto do investimento novo é convertido em capital na economia.

Figura 3 – Relação entre Investimento e Produto e a sensibilidade aos ciclos econômicos mundiais



NOTA: A figura mostra os coeficientes, no perfil dos países, do investimento na equação do produto (lado esquerdo) e do produto na equação do investimento (lado direito). O objetivo dessas figuras é avaliar a relação entre a causalidade entre pib_t e $invest_t$ com a sensibilidade dessas duas variáveis aos ciclos mundiais, controlado pelas variações no produto mundial no sistema de equações. Na figura do lado direito, analisamos a equação do investimento. No eixo vertical temos o coeficiente do pib_t na equação do $invest_t$ ($invest(pib)$) e no eixo horizontal nós temos o produto mundial ($pib.wld_t$) na equação do investimento ($invest(pib.wld_t)$). Na figura do lado direito usamos a mesma lógica mas na equação do produto.

3.3.2 Similaridades nas dinâmicas totais da economia

Nesse tópico vamos tentar agrupar os países por similaridades nas dinâmicas internas de suas economias. Essas similaridades são feitas com base nos perfis dos países. Para tanto vamos fazer três análises diferentes de agrupamentos em três subgrupos de variáveis utilizando duas metodologias diferentes. Com isso fazemos seis análises de agrupamentos e as metodologias utilizadas será a análise de cluster e de fatores.

A primeira, chamada de análise (a), aplica um procedimento de seleção dos coeficientes, das variáveis explicativas nos sistemas de equações, que representem o núcleo das similaridades dos perfis dos países. Isso quer dizer que vamos eliminar, da análise, aquelas variáveis (coeficientes) que destacam singularidades bem específicas de alguns países. Para fazer essa classificação e identificação das variáveis vamos utilizar

análises de fatores. Isso quer dizer que a classificação vai ser feita com base na variância conjunta das variáveis. Em específico, usamos um algoritmo que, a cada etapa, elimina o coeficiente (variável) que possui o menor nível de variância explicada pelos fatores estimados (com autovalores maiores que 1). Esse procedimento continua até que não existam variáveis com menos de 40% da variância explicada pelos fatores estimados. Dado o elevado grau de heterogeneidade nos perfis dos países, essa abordagem tem como objetivo definir aquelas variáveis que descrevem dinâmicas similares ou comparáveis dos países.

A segunda análise, chamada de análise (b), usa todos os coeficientes sem distinção, enquanto que, a terceira análise, chamada de análise (c), usa apenas os coeficientes das variáveis endógenas dos sistemas. Isso exclui as variáveis relativas ao mercado mundial (produto e inflação mundial e as relacionadas ao VIX) pertencentes ao grupo *exo* da tabela (2).

Dessa forma, todas as três análises começam com uma análise de fatores com o objetivo de identificar as variáveis que melhor definem, ou discriminam, cada grupo, assim como determinar o número relevante de grupos (fatores ou *clusters*) em cada análise. A tabela (7) mostra o número de fatores com autovalores maiores que 1 em cada análise. Esse número de fatores com autovalores maiores que 1 será usado para definir o número de *clusters* em cada análise de *cluster* e de fatores em cada análise de fatores.

Tabela 7 – Autovalores das análises de fatores

Fator	Análise de fatores		
	(a)	(b)	(c)
fator 1	1.85345	1.95741	1.34925
fator 2	1.25967	3.23106	1.26594
fator 3	1.39893	2.63802	1.2629
fator 4	1.99625	2.06521	1.24874
fator 5	1.82617	1.98948	1.20906
fator 6	1.52548		1.17216
fator 7	1.68969		1.16057
fator 8	1.14732		1.1561
fator 9	0.98756		1.11794
fator 10	1.14766		1.08467
fator 11	1.41989		1.07376
fator 12	1.07856		1.01204
fator 13	1.52496		0.97183
fator 14	1.19771		0.96797
fator 15	1.0483		0.94752
fator 16	0.9463		0.94283

NOTA: A tabela apresenta os autovalores das análises de fatores no perfil dos melhores sistemas de equações aplicados a cada país. Essa análise de fatores teve como principal objetivo escolher o número de clusters a ser definido na análise de clusters. A análise (a) se refere ao resultado do procedimento que exclui, a cada etapa, a variável (coeficiente do sistema de equações) com menor variância explicada pelos fatores estimados. Esse procedimento de escolha se encerra quando não houver nenhuma variável com menos de 40% da variância explicada pelos fatores estimados. A análise (b) usa todas as variáveis e a (c) retira os coeficientes das variáveis exógenas do sistema de equações.

3.3.2.1 Análise de fatores

A tabela (8) descreve os resultados dessas análises de fatores, assim como destaca as principais variáveis discriminantes de cada fator. Também destacamos alguns dos principais países de cada grupo, que nesse caso são fatores. Uma variável é chamada de discriminante de um grupo se possui um peso (*factor scoring*)

significativo, em relação às demais variáveis, na classificação dos países daquele grupo.

De acordo com essa tabela, podemos ver que, pela análise (a) e pela análise (b), os fatores ou grupos são formados, principalmente, baseados nos coeficientes das variáveis de VIX do sistema de equações. Isso quer dizer que, utilizando o critério da variância conjunta do perfil dos países, a principal característica que agrupa os países é como a dinâmica interna do país responde, ou reage, a variações na aversão ao risco internacional. Essa é uma conclusão importante porque a análise (a) é considerada a mais robusta, pois exclui os coeficientes que destacam particularidades dos países e mantém apenas aqueles coeficientes que descrevem um co-movimento do perfil dos países.

Quando retiramos as variáveis relacionadas ao mercado mundial, análise (c), verificamos que as variáveis de investimento direto estrangeiro (fdi_t) e câmbio (e_t) correspondem a mais de 2/3 das variáveis explicativas relevantes para discriminar os grupos. Isso reforça o argumento de que a principal característica que diferencia, ou agrupa, as economias é como a dinâmica interna responde ou reage ao mercado internacional.

Tabela 8 – Descrição dos principais clusters e das variáveis que os caracterizam

Fator	Análise (a)		Análise (b)		Análise (c)	
	Variáveis	Países	Variáveis	Países	Variáveis	Países
fator 1	$e : VIX.sd$	África do Sul Bolívia	$nx : VIX.sd$ $fdi : VIX.sd$	Suazilândia Singapura	$nx : r$	Argentina
fator 2	$fdi : VIX.m$	Brasil Geórgia Noruega	$e : VIX.m$ $fdi : nx$	Brasil México Geórgia	$cred : nx$	Reino Unido
fator 3	$nx : VIX.m$	Canadá Suécia	$cred : VIX.sd$	Argentina Austrália Federação Russa Reino Unido	$invest : fdi$	Bolívia
fator 4	$nx : VIX.sd$	Armênia	$nx : VIX.m$	Canadá Japão Estados Unidos	$pib : e$	Filipinas África do Sul Croácia
fator 5	$cred : VIX.m$	Israel	$r : VIX.sd$	Chile Venezuela	$pib : nx$	Canadá Chile
fator 6	$r : VIX.sd$	Chile	$invest : VIX.sd$	Israel Korea	$invest : cred$	China Venezuela
fator 7	$invest : VIX.sd$	Federação Russa România			$cred : e$	Japão
fator 8	$cred : VIX.sd$ $invest : pi.wld$	Eslováquia Marrocos			$fdi : e$	Israel México
fator 9	$r : pi.wld$	Japão Austrália Indonésia México Uruguai			$nx : cred$	Colômbia Federação Russa Tailândia
fator 10	$fdi : VIX.sd$	Hungria Argentina			$cred : r$	Austrália Índia
fator 11	$BIC.m$	Estados Unidos Suíça			$pib : r$ $invest : nx$	Noruega Costa Rica
fator 12	$pib : VIX.m$	China República Checa Bulgária			$gdp : invest$ $fdi : cred$	Korea
fator 13	$nx : e$	Vietnã Colômbia			$nx : fdi$ $r : e$	Albânia
fator 14	$invest : pi.wld$ $cred : pi.wld$	Reino Unido Venezuela			$invest : e$	Geórgia
fator 15	$pib : pib.wld$	Índia			$pib : fdi$	Brasil Estados Unidos
fator 16	$invest : VIX.m$ $invest : VIX.sd$	Korea Tailândia			$cred : fdi$	Singapura

NOTA: A figura mostra o resumo do resultado das análises de fatores feitas para selecionar variáveis, identificar as variáveis mais importantes na discriminação de cada grupo e definir o número de grupos relevantes na análise de cluster. A primeira coluna se refere ao número de fatores com autovalor maior que 1 naquela análise. Para cada análise ((a), (b) ou (c)) separamos os resultados em duas colunas. A primeira de “variáveis” mostra as variáveis mais importantes para discriminar aquele grupo. Na coluna “países” destacamos alguns países relevantes naquele grupo.

3.3.2.2 Análise de clusters

Agora vamos focar nos resultados das análises de *clusters*. Aqui também utilizamos os subconjuntos de variáveis da análise (a), (b) e (c).

Na figura (4) mostramos os dendogramas das três análises de *clusters*, com o número de *clusters* definido pela tabela (7). A primeira conclusão que se tira é, tal como descrito na seção anterior, o elevado grau de heterogeneidade dos grupos. Segundo que, tal como visto na tabela (7), a análise que usa menos fatores é aquela com mais variáveis (análise (b)).

Independentemente do número de fatores com autovalores maiores que 1 descritos na tabela (7), o que mais se destaca é a existência de dois grandes grupos de países nas 3 análises. Obviamente, para um conjunto diferente de variáveis, o formato desses grupos é diferente. Mas ainda sim se destacam dois grupos.

No caso da análise (a), usando um processo de seleção de variáveis, o primeiro grupo vai de G1 até G12 e o segundo vai de G13 até G16. Nesse primeiro grupo se destacam Brasil, Korea, Austrália, México, Rússia, Argentina, Chile, China, Índia, Indonésia e Uruguai. No segundo estão os principais países desenvolvidos como Japão, Estados Unidos, Canadá e Reino Unido.

Na análise (b), que usa todas as variáveis, permanecem no primeiro grupo Brasil e Korea com a adição de Reino Unido e Japão. De resto os países principais dos grupos são os mesmo descritos para a análise (a).

Na análise (c), essa divisão em dois grupos é menos evidente e as distâncias (medidas no eixo vertical) são menores. Dos países citados acima, apenas Japão e Chile estão no primeiro grupo, todos os demais estão no segundo grupo.

Sendo assim, verificamos que, um conjunto diferente de variáveis nas análises modifica de forma significativa os resultados. A análise (b) é a menos criteriosa e por isso julgamos que os resultados são menos importantes. A análise (a) é a mais refinada e a análise (c) é a que reúne apenas as variáveis de interesse.

Na tabela (9) e na tabela (10) mostramos o perfil dos países dentro de cada *cluster* gerados nas análises (a) e (c), respectivamente.

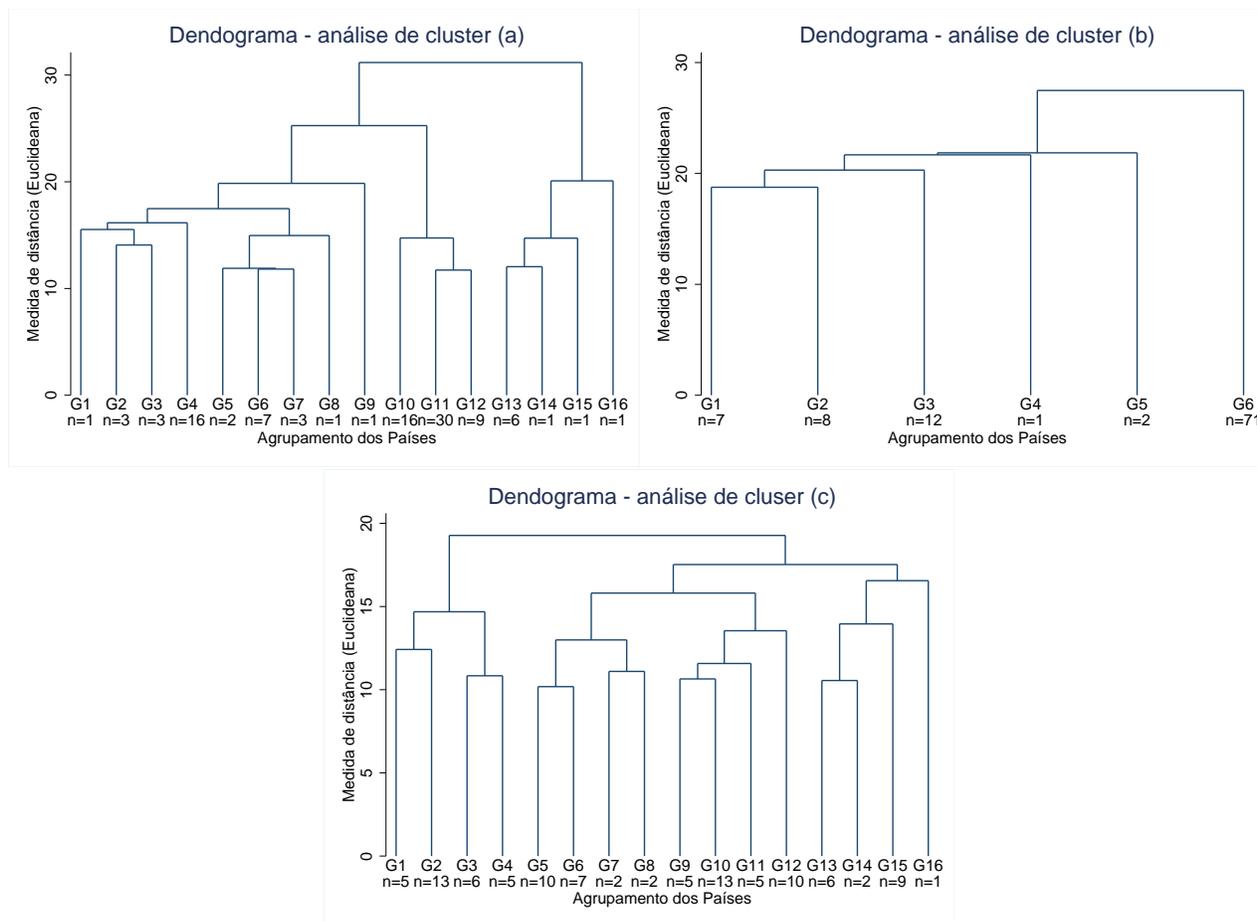
Existe, nessas duas tabelas, um total de 32 *clusters*. Assim, para objetivar a exposição, vamos focar nas características gerais no cluster do Brasil e nas dos países desenvolvidos.

De um modo geral, assim como na análise de fatores, as características mais evidentes para diferenciar os *clusters* na análise (a) são as relacionadas ao VIX ($VIX.m_t$ e $VIX.sd_t$), câmbio (e_t) e poupança externa (nx_t) como variáveis explicativas. Na análise (c) a principal variável explicativa para discriminar os clusters é a poupança externa (nx_t) seguida do investimento externo (fdi_t). Dessa forma, a história aqui é a mesma da descrita na análise de fatores.

No caso do Brasil, pela análise (a) ele está no *cluster* 2 e na análise (c) ele está no *cluster* 15. O *cluster* 2, na análise (a), é caracterizado, principalmente por três características principais. Primeiro é a elevada dependência do produto e do investimento ao VIX ($pib(VIX_m)$ e $invest(VIX_m)$ respectivamente). Segundo é a balança comercial que é muito influenciada pela taxa de câmbio ($nx(e)$)¹⁰. Terceiro é que o

¹⁰ Vale destacar aqui que a taxa de câmbio é medida pela USD sobre o BRL. Assim, uma relação negativa como a apontada no *cluster* 2, que dizer uma valorização do real (BRL) ou uma desvalorização do dollar americano (USD) tem um efeito sobre a balança comercial muito menor do que o que acontece no resto do mundo. Vale notar que as variáveis são medias

Figura 4 – Dendogramas das Análises de Clusters



NOTA: A figura mostra o dendograma as 3 análises de Clusters feitas, com os perfis dos países, pelo método de Ward. No eixo vertical está a distância euclidiana que foi a medida de distância utilizada. Nós não mostramos as figuras até a base porque, devido ao elevado número de países, o eixo horizontal não ficaria claro. Nesse sentido fizemos o ponto de corte no número de fatores definidos na tabela (7). A base das figuras mostra o número de países em cada grupo, ou cluster.

investimento estrangeiro, o crédito e a taxa de juros são muito menos influenciados pelos ciclos mundiais que o resto do mundo.

O cluster 15, pela análise (c), é caracterizado por duas características principais. Primeiro que o PIB é muito mais sensível aos fluxos de capital internacional (nx_t e fdi_t) que o resto do mundo. Em segundo que o investimento e, conseqüentemente o PIB, são muito influenciados pelos juros domésticos (r_t).

No caso dos países desenvolvidos, na análise (a), os clusters G13 a G16, são caracterizados por baixa sensibilidade à aversão ao risco mundial ($VIX.m_t$).

Tantos os resultados para o Brasil como para os países desenvolvidos fazem muito sentido. Enquanto que os ciclos econômicos do Brasil dependem muito dos fluxos de capitais internacionais, os ciclos dos países ricos são pouco influenciados pelo humor dos investidores internacionais.

em relação à media mundial.

3.4 Conclusão

Os debates sobre crescimento econômico após meados dos anos 80 ganharam novos contornos. Nessa época, em resposta às deficiências dos modelos neoclássicos, Romer e outros economistas apresentaram os primeiros modelos de crescimento endógenos. A partir de então a literatura procurou identificar os determinantes do crescimento que poderiam ser controlados por políticas.

Muitas variáveis foram testadas e muitas teorias de relações de causa efeito foram desenvolvidas. Algumas, no entanto, não encontraram suporte adequado dos dados e experiências empíricas.

Um exemplo importante é a educação e o investimento na qualidade do capital humano. Essa variável teria um efeito direto e positivo no progresso tecnológico de um país que por sua vez é o principal determinante do crescimento nos modelos neoclássicos. Esse *link* foi apontado, por um tempo, como a maior omissão dos modelos de crescimento exógenos. No entanto, o investimento expressivo em educação que se observou ao redor do mundo desde então não foi acompanhado por um desenvolvimento tecnológico e, conseqüentemente, crescimento econômico em muitos países tal como esperado.

A abertura comercial seria outro exemplo. Apontada como a mais importante variável, ou uma das mais importantes, para explicar o crescimento em trabalhos de regressão de crescimento, existem trabalhos na literatura que argumentam que esses resultados devem ser vistos com cautela e questionam a robustez dessa variável. Em tese, uma maior abertura comercial e integração com o mercado internacional permitem ao país absorver tecnologia e know-how de países mais desenvolvidos. Na verdade essa teoria é válida e a dinâmica é observada nos países do leste asiático, que poderiam ser classificados como as melhores referências em termos de sucesso de estratégia de crescimento. No entanto, é possível encontrar outros casos e situações no tempo em que essa relação de causa e efeito não é observada.

Uma relação de causa e efeito que também é questionada é do investimento e do crescimento econômico. Essa relação é especialmente importante na literatura devido ao papel do investimento na dinâmica do produto. O investimento é o principal determinante da acumulação de capital na economia. Enquanto que um país com mais qualidade da mão de obra irá crescer mais, um país com maior acúmulo de capital vai ser mais rico e proporcionar melhor qualidade de vida à população.

É em função dessa importância do investimento para o crescimento de um país e da conseqüente relevância das dúvidas sobre a relação de causa e efeito dessas variáveis que nós vamos explorar, nesse artigo, o tema.

Para tratar do assunto, nós utilizamos sistemas de equações simultâneas com controles endógenos para o produto, o investimento, a poupança e os preços da poupança e do investimento. A partir desses sistemas de equações simultâneas nós obtivemos mais informação que a relação de causalidade direta entre produto e investimento. Com essas informações nós podemos estudar e compara as dinâmicas de investimento, poupança e produto nas economias¹¹.

Nós utilizamos uma amostra de 101 países que possuísem observações suficientes nas variáveis utilizadas para o período de 1990 a 2012. Como utilizamos dados anuais e nem todos os países possuíam todas as observações em todo o período, nós não teríamos observações suficientes para estimar os parâmetros de um sistema de equações com sete variáveis endógenas. Dessa forma, para contornar o problema de

¹¹ É importante notar que nossas variáveis foram diferenciadas de forma que o que modelamos nos sistemas de equações são as variações. Assim, quando falamos em causalidade entre produto e investimento, por exemplo, estamos nos referindo à causalidade em termos de taxas de crescimento.

insuficiência de dados, nós estimamos a distribuição conjunta desses dados para cada país e simulamos uma amostra grande o suficiente para estimar os parâmetros necessários.

Nossos resultados indicam que existem mais países em que o investimento causa o produto, para o período considerado, do que na direção oposta. Esses resultados corroboram com a teoria tradicional de que o investimento causa o crescimento.

Quando olhamos para o sistema de equações como um todo, verificamos que existe muita heterogeneidade na dinâmica de produto, investimento, poupança e juros dos países e que não se pode definir uma teoria global. Esse comentário também se aplica, em específico, às dinâmicas de produto e investimento. Com isso, nossos resultados indicam que o canal de transmissão entre juros, poupança, investimento e produto é significativamente diferente entre os países.

Na verdade, verificamos que o que mais aproxima as dinâmicas internas dos países é como o país responde à aversão ao risco internacional. Seja essa aversão medida diretamente, seja através dos fluxos de capitais internacionais e da taxa de câmbio. Se tivermos que apontar uma característica que diferencia os países, essa seria a melhor opção para a década de 90 e anos 2000. Essa conclusão não tem relação nenhuma com os resultados tradicionais das regressões de crescimento que apontam o grau de abertura internacional como uma variável importante e consistente para explicar as variações do produto. O que indicamos aqui é que a forma como o país reage a movimentos no humor internacional é uma característica importante para agrupar os países. Esses resultados estariam mais de acordo com a idéia de que o investidor internacional avalia os países em blocos.

Especificamente no que se refere ao Brasil e aos países desenvolvidos, enquanto que os ciclos econômicos do Brasil dependem muito dos fluxos de capitais internacionais, os ciclos dos países ricos são pouco influenciados pelo humor dos investidores internacionais.

Também é importante mencionar que notamos que alguns resultados atípicos, como relação inversa da esperada entre variáveis do sistema, ressaltam distorções internas e efeitos de choques em algumas economias. De fato, na literatura, boa parte dos ciclos das economias é atribuída a choques. Também existe uma linha de pesquisa importante na literatura que atribui à qualidade das instituições, e distorções internas, a eficiência de políticas de crescimento e o bom funcionamento interno dessa.

Tabela 4 – Relações entre Produto e Investimento

País	Equação	Variáveis Explicativas										
		pib_t	$invest_t$	nx_t	fdi_t	$cred_t$	e_t	r_t	$vix.m_t$	$vix.sd_t$	$pib.wld_t$	$pi.wld_t$
Albania	pib_t		-0.99			-0.64	-0.25	-0.23	0.98	-0.96	1.03	0.46
	$invest_t$			-0.30			-0.15	-0.08	0.69	-0.30		0.40
Antigua e Barbuda	pib_t	0.05	-0.22	0.31	0.53			-0.05	-0.36	-0.19	-0.03	0.43
	$invest_t$		0.23	0.04	-0.42			0.13	0.31	-0.08	-0.03	-0.30
Argentina	pib_t		-0.28		-0.20	-0.50	-0.20		0.08	0.30	0.27	0.16
	$invest_t$	-0.23	0.31	0.31	-0.30				0.19	-0.15	0.20	0.04
Arménia	pib_t		-0.43			0.80	-0.08			-0.52	0.54	-0.40
	$invest_t$			0.56	-0.20	-0.71	-0.89					
Austrália	pib_t		0.31	-0.15	-0.45	0.36	-0.41			-0.11	0.09	0.19
	$invest_t$	-0.30	0.37	-0.13	0.48		0.24	-0.18				
Bahamas, As	pib_t		-0.47	0.43	0.15	-0.60			-0.08		-0.26	
	$invest_t$			0.64				0.31	-0.15	-0.24		0.05
Bahrain	pib_t		-0.42				-0.14	-0.29	-0.42		-0.12	0.70
	$invest_t$	0.45		0.41		-0.44	-0.22	-0.23			-0.30	
Bangladesh	pib_t			0.21	-0.32	-0.25	0.31	-0.15	-0.10		-0.74	
	$invest_t$	0.56			-0.34	0.38		0.20				0.19
Barbados	pib_t				-0.43			0.34			-0.20	
	$invest_t$			-0.65	-0.13			-0.27	-0.35	-0.10		0.30
Belarus	pib_t		0.07	0.63	0.15	0.21	0.52	0.12	-0.04	-0.21	-0.32	
	$invest_t$	0.13		-0.05	0.06	0.49	0.15	0.33	0.08	-0.33	-0.45	
Belize	pib_t		0.25		-0.13		-0.43	-0.03	-0.45	-0.13	0.12	
	$invest_t$	-0.78			-0.18			-0.04		-0.21	0.16	
Butão	pib_t		-0.07	-0.72	0.55	0.59	0.59	-0.27	-0.51	0.45	0.18	
	$invest_t$	0.09		-0.80	-0.10		-0.42	0.19	-0.01	0.10	-0.18	-0.62
Bolívia	pib_t		-0.24	0.68	-0.11	0.11	-0.30	-0.14	-0.16	-0.11	-0.17	
	$invest_t$	0.34		0.61		-0.18	0.12	-0.16	-0.06	0.06		
Brasil	pib_t		0.19	0.62	-0.54	-0.48	0.25	0.16	-0.57	-0.05	0.13	
	$invest_t$	-0.35		0.58	0.27		0.38		-0.06	0.07	0.04	
Bulgária	pib_t		-0.51	0.15	0.54	0.59		0.46	-0.23	0.21	-0.24	
	$invest_t$	0.42		-0.45	-0.25		-0.76	-0.44	-0.23	0.22	0.44	0.13
Burundi	pib_t		-0.37	-0.16		0.10	-0.10	-0.34	-0.19	-0.15	-0.29	
	$invest_t$		0.14	0.35	-0.52	0.06	0.30	0.19	-0.47		0.38	
Camarões	pib_t		-0.50	0.41		1.07	0.70	0.07	0.53	0.04	0.56	
	$invest_t$	-0.24		-0.48	-0.29	-0.51	0.12		-0.64	0.36	-0.78	0.36
Canadá	pib_t		-0.34	0.32	-0.50	-0.48	-0.50		0.12	0.60	-0.54	-0.15
	$invest_t$				-0.45		-0.25	-0.31		-0.13		-0.14
República Centro-Africana	pib_t		0.23	-0.25	0.78	0.31	-0.36		0.59	-0.09	0.20	
	$invest_t$	0.42		-0.49	-0.67	-0.19		0.96	-1.51	-0.27	-0.33	
Chade	pib_t		-0.17	0.32	0.17	0.96	-0.71	-0.22	-0.26	0.47		
	$invest_t$		0.20	0.95	0.32	-0.23	-0.50	-0.25	-0.07	-0.33	0.39	
Chile	pib_t		0.33	0.07	0.25	0.41	-0.06	0.14	-0.33	-0.10	-0.06	
	$invest_t$	0.16		-0.22	0.05	0.38	-0.42		0.06	0.20	0.01	0.08
China	pib_t		-0.10	0.14	-0.23	-0.30		0.13	0.79	-0.56	0.14	0.07
	$invest_t$		0.18	0.41	0.54	-0.25	-0.12	-0.32	0.24	-0.51	-0.38	
Colômbia	pib_t		0.61	0.21	0.31		-0.17	-0.03	0.04	0.07	0.06	
	$invest_t$			0.30	0.31	-0.06	-0.88	-0.14	0.12	0.09	0.18	
Congo, Rep.	pib_t		0.81	-0.12	-0.21	0.69	-0.10		0.15	-0.18	-0.16	
	$invest_t$	0.69		-0.18	0.19	-0.24	-0.31	0.02	0.50	-0.28	-0.11	
Costa Rica	pib_t		0.17	-0.61	0.46	0.58	0.21	-0.29	0.03	0.17	-0.27	
	$invest_t$			0.48	-0.33	0.29	-0.10	0.38	0.05	-0.07	-0.10	0.15
Croácia	pib_t		-0.28	0.26	-0.44	0.63	0.31	-0.25	0.08	0.09	0.36	
	$invest_t$	-0.53		0.35	0.35	0.33	-0.48	-0.34	0.59	0.08	0.21	
Chipre	pib_t		0.25	-0.21	-0.19	-0.82	-0.07	-0.03	0.10		-0.06	
	$invest_t$	0.28		-0.09	-0.11	0.95	0.26	0.03	0.05	-0.08	-0.38	
República Checa	pib_t		-0.13		-0.26	-0.26	0.07	-0.15	0.71	-0.35	-0.02	-0.06
	$invest_t$		0.10	0.38	1.12	-0.18	0.13	0.10	-0.45	0.29	0.34	
República Dominicana	pib_t		0.69	-0.10	-0.17	-0.62	-0.04	0.13	-0.14	0.08	0.20	
	$invest_t$			0.44	0.15	0.14		-0.16				
Equador	pib_t		0.73	0.47	-0.57		0.16	-0.11	0.15	-0.54	-0.17	0.78
	$invest_t$			-0.83	0.17	-0.62	-0.37	0.09	-0.44	0.26	-0.41	0.12
Egito, Rep. Arab	pib_t		0.46	0.19	0.13	0.34		0.05	0.19	-0.11	-0.14	
	$invest_t$	-0.11		0.31	0.20	-0.08	0.58		-0.07	0.19	0.16	-0.29
Estônia	pib_t		0.06	-0.18	-0.08	0.79		-0.18	0.12	0.21	-0.12	-0.19
	$invest_t$			0.35	0.14	-0.36		0.27	-0.49	-0.07	-0.15	0.24
Etiópia	pib_t		-0.32		0.38	-0.56	0.58	-0.22	-0.44	0.06	-0.30	0.04
	$invest_t$		0.24	-0.09	-0.10	-0.06	0.56	-0.15	0.05	0.22	-0.61	
Fiji	pib_t		-0.21	0.72	-0.30	0.03	0.26	-0.27		-0.35	-0.07	
	$invest_t$		0.23	0.08	0.08	-0.06	0.59	-0.28	-0.10	-0.06	0.31	
Georgia	pib_t		-0.47		-0.14	-0.38		0.27		0.63	0.33	
	$invest_t$		0.16	-0.37	-0.68	1.01	0.41	1.25	-0.80	0.28	-0.23	
Granada	pib_t		-0.59	-0.25					0.19			
	$invest_t$		0.26	0.23	0.47					0.15	-0.25	-0.28
Guatemala	pib_t				-0.12	0.46	-0.15		0.23	-0.21	-0.28	
	$invest_t$	-0.40		-0.74	0.23	-0.11	0.17		0.20			
Guiana	pib_t		-0.56	0.19	-0.28	0.12	0.61		-0.24			
	$invest_t$		0.05	-0.12		0.16	-0.66	0.72	-0.23	-0.78	0.74	
Honduras	pib_t		-0.24		-0.48	0.54	-0.28	-0.12	0.07	-0.14	0.09	
	$invest_t$		-0.31	0.13	0.11	-0.56	0.23	-0.20	-0.05	0.28	0.08	
Hong Kong SAR, China	pib_t		-0.45	-0.27	-1.04	-0.67	-0.83		-0.85	0.99	-0.22	0.32
	$invest_t$			-0.29	-0.67	-0.50		-0.29				
Hungria	pib_t		-0.63		0.05	-0.33	0.60		-0.18	-0.18	-0.30	-0.53
	$invest_t$		0.56	0.06								
Islândia	pib_t		-0.09		-0.46	0.31	-1.07		0.16	-0.35	0.14	-0.41

Tabela 5 – Relações entre Produto e Investimento (continuação)

País	Equação	Variáveis Explicativas										
		pib_t	$invest_t$	nx_t	fdi_t	$cred_t$	e_t	r_t	$vix.m_t$	$vix.sd_t$	$pib.wld_t$	$pi.wld_t$
Indonésia	pib_t			-0.26	0.58	0.11	0.39	0.67	0.25	0.30	-0.14	-0.57
	$invest_t$			0.29	0.17	0.66	-0.27	-0.20	-0.06	0.25	-0.14	-0.15
Israel	pib_t			0.14	0.17	-0.09	0.25	-0.04	-0.12	0.19	-0.19	-0.28
	$invest_t$	0.77			0.17	-0.08		0.09	0.18	0.16	-0.10	
Japão	pib_t			0.04	0.05	0.09	-0.97	0.04	0.01	0.01	-0.09	-0.06
	$invest_t$	1.34		0.30		0.38		0.32	0.24	0.26		0.28
Jordânia	pib_t			-0.37	-0.12	-0.43	0.06		-0.21	0.18	0.45	0.57
	$invest_t$			-0.80	0.29	0.05	0.18	-0.02	-0.25	-0.41	-0.10	-0.29
Kênia	pib_t		0.27	-0.60	-0.08	-0.10	-0.64		-0.30		-0.02	
	$invest_t$			0.64		0.06	0.39	-0.22	0.38	-0.05	0.05	-0.09
Korea, Rep.	pib_t		0.96		-0.20	0.18	0.14	0.17	0.05	0.22	0.08	-0.09
	$invest_t$			0.43	0.16		-0.61	-0.16	0.15	0.16	-0.18	0.12
Kuwait	pib_t			0.38	-0.55	0.16	-0.43	-0.17	0.52	-0.97		0.39
	$invest_t$	-0.32				-0.08	-0.14	-0.06	-0.49		1.09	
Letônia	pib_t		0.48			-0.19		0.19	0.27		-0.62	-0.49
	$invest_t$			0.41			-0.13		0.46	-0.15		0.30
Lituânia	pib_t		-0.37			0.44	0.38	0.36	-0.28	0.23	-0.38	
	$invest_t$			-0.41	0.54	0.19	-0.21		-0.23	0.14	0.21	0.18
Macedônia, FYR	pib_t			-0.04	-0.09	-0.06	-1.07	-0.09	0.12	-0.04	0.01	-0.03
	$invest_t$			-0.35	0.23	0.27	-0.38	0.38	0.41	-0.41	0.15	-0.13
Madagascar	pib_t		-0.12	0.10	-0.38	0.11	0.45		-0.16	-0.19	0.54	0.13
	$invest_t$			0.33	0.33	0.17	-0.68		0.29	0.16	-0.07	0.05
Malawi	pib_t		0.83	-0.42	-0.24	-0.31	0.29		0.40	-0.20	-0.37	-0.39
	$invest_t$			0.63	0.21	0.24	0.26	-0.43	-0.38		0.25	
Malásia	pib_t		0.40	-0.12	-0.34	0.15	0.21		-0.02	0.23		-0.19
	$invest_t$			-0.28	-0.44	-0.33	-0.08	0.40	0.02	0.20	-0.07	-0.24
Malta	pib_t			-0.18		0.62			0.29	0.18		
	$invest_t$	0.56		0.13	0.21	-0.33			-0.34		-0.71	-0.38
Maurícia	pib_t		-0.31	0.54	0.22	-0.21	0.53		0.02	0.29	-0.20	-0.26
	$invest_t$			0.33	0.24	-0.25	-0.81	-0.54	-0.07	0.41	-0.19	0.10
México	pib_t			-0.04	-0.09	0.46	-0.68	-0.41	-0.17	-0.11	-0.11	0.05
	$invest_t$	1.05		-0.18	0.30	0.08	0.35		-0.15	-0.07	0.04	0.20
Moldova	pib_t		0.48	0.09		-0.32	-0.72	-0.12	0.04	0.09	-0.07	0.08
	$invest_t$			0.03	-0.24	0.15	-0.50	0.01	-0.17	-0.11	0.03	0.29
Mongolia	pib_t		0.60	-0.19	-0.34	-0.28	0.63		-0.35			-0.21
	$invest_t$			-0.25		0.38	0.68	0.27		0.55	-0.55	-0.21
Marrocos	pib_t		-0.25	-0.16	0.39		0.11	-0.55	0.78	-0.96	-0.85	0.40
	$invest_t$			0.28	-0.74	0.43	-0.57		0.31		0.56	
Namíbia	pib_t				-0.22	0.09	0.52	-0.10	-0.06	0.16		
	$invest_t$	0.63		-0.51		0.22		-0.13		-0.09	0.24	-0.25
Nicarágua	pib_t			0.12	-0.12	0.40	0.35	0.31	0.14	-0.29		0.54
	$invest_t$	-0.73		0.56	-0.58		-0.06	-0.05	-0.02	0.25	-0.15	0.26
Nigéria	pib_t			-0.49	0.07	0.06						0.18
	$invest_t$	0.83		0.09			0.10	-0.21	0.02	0.44	-0.03	-0.29
Noruega	pib_t		0.41	-0.40	-0.46			0.52	0.18	0.68	0.50	
	$invest_t$			0.54	0.30	-0.43	-0.25		0.63	-0.75	-0.29	-0.33
Oman	pib_t			0.83	0.14				0.18		-0.30	0.30
	$invest_t$			-0.72		0.30		0.31		-0.44	0.29	-0.11
Panama	pib_t					0.49			0.53	-0.09	0.11	0.37
	$invest_t$	0.53		0.14	0.26		0.07		-0.30	0.06	-0.17	-0.34
Papua-Nova Guiné	pib_t		-0.75		-0.67	-0.16	0.88	-0.55	-0.31		0.45	-0.23
	$invest_t$				-0.57	-0.12	0.36	-0.59	0.26	0.11	-0.35	
Paraguai	pib_t				-0.13	0.09	-0.70		-0.09	0.10	-0.12	-0.27
	$invest_t$	0.45		-0.25		0.69	0.18	-0.28	-0.35	0.13	-0.02	0.16
Peru	pib_t			-0.20	-0.21	0.11	0.63		0.04	0.48	-0.22	-0.42
	$invest_t$	1.00		-0.18	0.10			0.32				
Filipinas	pib_t			0.34	0.14		0.67	-0.34	-0.04	0.22		-0.17
	$invest_t$			-0.21		0.21	0.32	0.24	-0.21	-0.03	0.19	-0.09
Polónia	pib_t		0.44	0.28	0.20	0.24	0.32		0.08			0.09
	$invest_t$			-0.08		0.62	-0.19	0.45	-0.15	-0.04	-0.46	-0.23
Romania	pib_t			-0.42	0.32	0.18	0.32	-0.08	-0.35	0.21	-0.11	-0.12
	$invest_t$			-0.61		-0.25	0.16	-0.37	-0.72	0.62	-0.26	-0.38
Federação Russa	pib_t		-0.27	0.25		-0.43	-0.73	-0.19	-0.21	-0.18	0.16	0.45
	$invest_t$			0.26	0.05	-0.60	0.22	0.30	-0.49	0.53	-0.06	-0.03
Ruanda	pib_t			0.56	-0.92	0.20	0.61	-0.18	-0.17	0.33	1.03	-0.89
	$invest_t$	0.14		-0.39	0.25	0.39			-1.66	1.09	0.41	-0.68
Sérvia	pib_t											
	$invest_t$	0.99										
Seicheles	pib_t		-0.37		-0.35	0.07	-0.53	-0.28	-0.24	0.54	-0.22	-0.31
	$invest_t$			-0.11	-0.75	0.23	-0.59	0.68	-0.22	0.94	-0.11	-0.13
Serra Leoa	pib_t		0.12		0.11	-0.06	0.52	0.15	-0.38	0.33	-0.28	0.13
	$invest_t$			0.40	-0.08	-0.30	0.08	-0.21	-0.11	0.10	-0.40	-0.08
Singapura	pib_t		0.17		-0.12	0.24		0.63	-0.21		-0.12	-0.02
	$invest_t$			0.37	0.18		0.55	-0.25	0.12	-0.15	0.32	
Eslováquia	pib_t		-1.22	0.36		-0.24	-0.21	0.34	-0.05	-0.37	0.12	-0.58
	$invest_t$			-0.04	-0.16	-0.68	-0.21	0.13	0.20	-0.38	-0.10	0.08
África do Sul	pib_t			0.10	-0.33	1.61	0.99	-0.42	0.20	0.47	-0.43	-0.28
	$invest_t$			0.16	0.15	-0.30	0.22		-0.05	0.22	-0.21	-0.44
Sri Lanka	pib_t			-0.29		0.36	-0.05		-0.19	0.12	-0.30	-0.20
	$invest_t$	0.42		-0.17	-0.23	0.15		-0.09	0.28		-0.14	-0.11
St. Lucia	pib_t				-0.22			0.17		-0.36		0.20
	$invest_t$	0.47		-0.05	0.23	0.26			-0.44	0.16	-0.39	-0.17
São Vicente e Granadinas	pib_t		-0.29	-0.19	-0.07	-0.09		0.48	0.03	0.05	0.23	0.38

Tabela 6 – Relações entre Produto e Investimento (continuação)

País	Equação	Variáveis Explicativas										
		pib_t	$invest_t$	nx_t	fdi_t	$cred_t$	e_t	r_t	$vix.m_t$	$vix.sd_t$	$pib.wld_t$	$pi.wld_t$
Suécia	pib_t		-0.22	0.32	-0.11		0.11	1.10	-1.08	-0.32	-0.50	-0.25
	$invest_t$			0.30	-0.21	0.26	0.22	1.17	-1.43	0.11	-0.25	-0.45
Suíça	pib_t		0.15	0.09			0.55		0.02	-0.02	-0.10	-0.25
	$invest_t$			-0.21	-0.13	-0.18	0.36	0.08	-0.06	0.07	-0.13	-0.26
Síria, Rep Arab	pib_t		0.63		-0.46				-0.13		0.51	-0.98
	$invest_t$			0.43		-0.97		0.45	0.21	-0.48		
Tanzânia	pib_t			-0.06	-0.59	0.07	0.47	0.39	0.03	0.02	-0.04	-0.17
	$invest_t$			0.70	0.18	0.15		-0.32	-0.04	0.02	-0.34	-0.83
Tailândia	pib_t		0.47	0.26		0.67	-0.33			-0.05	-0.06	0.02
	$invest_t$			-0.49	-0.23	0.98		-0.33	0.08	0.11	-0.09	-0.02
Tonga	pib_t		-0.32	-0.16		0.63	-0.04	0.35	-0.33	0.47		
	$invest_t$			0.15	0.43	0.26	-0.64		-0.12	-0.39	0.09	0.43
Trinidade e Tobago	pib_t		0.14	-0.32		0.65	0.35		0.26	0.39	0.14	-0.23
	$invest_t$			0.22		-0.18	-0.15	0.10	-0.51	-0.21	0.36	-0.10
Uganda	pib_t		0.12	-0.25	-0.15	-0.11	0.47	0.05	0.03	-0.44	0.55	0.39
	$invest_t$			-0.05	0.47	-0.74	0.19	0.26	-0.04	-0.60	0.87	-0.23
Ucrânia	pib_t		0.98			-0.09			0.26	-0.35		0.10
	$invest_t$			0.48	0.44		-0.25	0.27				
Reino Unido	pib_t		-0.14	0.09	-0.21		-0.36	1.06		-0.18	-0.07	0.03
	$invest_t$			0.34	0.27	-0.24	-0.75	0.06	-0.32	0.33	-0.14	0.27
Estados Unidos	pib_t		-0.08	0.07	0.22	-0.06		0.67	-0.29	-0.14	-0.13	-0.09
	$invest_t$			0.07	0.05	-0.10		0.20	-0.11	0.03	0.45	-0.07
Uruguai	pib_t			0.53	0.15	-0.07	-0.24	-0.04	-0.16	0.24	-0.14	-0.20
	$invest_t$	0.83			-0.10	0.08	0.58	0.06	-0.06	0.05	-0.12	-0.07
Vanuatu	pib_t			0.28	0.31	-0.27	-0.56	-0.23	0.33	-0.22	-0.15	-0.08
	$invest_t$	0.85		0.22	-0.50	0.18	0.51		-0.25	0.24	-0.04	-0.42
Venezuela, RB	pib_t			0.30	0.09	0.57	0.21	-0.22		0.26	-0.42	0.11
	$invest_t$	0.14		-0.11		0.60	0.35	0.04	0.08		0.06	0.08
Vietnã	pib_t			0.74		-0.55	-0.06	-1.01	-0.14	0.14		0.08
	$invest_t$				0.30	0.59	0.20	0.48	0.15	-0.02	0.04	-0.16

NOTA: A tabela mostra o perfil dos países da amostra para as equações do produto (gdp_t) e investimento ($invest_t$) no sistema de equações. O perfil dos países são coeficientes médios dos 100 melhores modelos para cada país que foram significativos em pelo menos 51 modelos. As colunas mostram os coeficientes, do perfil de cada país, de cada variável explicativa utilizada.

Tabela 9 – Valor médio das variáveis normalizadas em cada cluster - análise (a)

variável	c 1	c 2	c 3	c 4	c 5	c 6	c 7	c 8	c 9	c 10	c 11	c 12	c 13	c 14	c 15	c 16
<i>pib(invest)</i>	-2.79	-0.69	0.67	1.11	0.23	-0.67	-0.31	-0.02	-0.02	-0.23	-0.15	0.08	-0.25	-1.28	-0.65	1.14
<i>pib(e)</i>	-0.68	-0.43	-0.75	-0.56	1.29	1.21	0.75	1.07	0.46	-0.07	0.38	-0.46	-0.97	-1.86	0.05	-0.17
<i>pib(vix_m)</i>	3.32	1.06	1.26	0.26	-1.46	-0.38	-0.42	-0.57	0.01	-0.13	0.14	-0.38	0.02	-2.85	-3.63	0.63
<i>pib(vix_sd)</i>	-2.80	-1.47	-1.14	-0.35	0.87	0.50	0.24	1.06	1.81	-0.36	0.25	0.16	0.02	3.03	-0.90	2.10
<i>pib(pib_wld)</i>	3.39	-0.24	0.28	0.10	-0.59	0.10	0.02	3.39	-0.24	-0.17	-0.13	0.19	-0.38	-0.65	-1.57	1.67
<i>invest(fdi)</i>	-0.12	-0.70	0.26	0.34	-0.45	-0.47	1.69	0.72	-2.34	0.12	0.23	-0.88	-0.30	-2.37	-0.83	0.87
<i>invest(r)</i>	-0.43	0.77	-0.56	0.02	0.33	-1.63	-0.42	-0.15	-0.15	-0.10	0.03	0.94	0.18	-1.15	3.96	-0.15
<i>invest(vix_m)</i>	1.96	1.51	-0.54	-0.03	0.12	-0.17	-0.74	-4.30	2.68	0.27	-0.10	-0.01	0.17	0.12	-3.69	1.79
<i>invest(vix_sd)</i>	-0.95	-0.91	0.21	-0.06	0.78	0.96	-0.04	3.23	-4.60	-0.40	-0.01	0.42	0.37	-0.05	0.27	-2.31
<i>invest(pi_wld)</i>	1.67	-0.04	-0.22	0.40	-0.83	0.68	1.18	-2.31	-1.03	0.65	-0.42	-0.80	0.18	0.19	-1.47	-1.01
<i>nx(fdi)</i>	3.02	-0.76	0.05	0.27	-0.89	0.39	-0.55	0.05	0.05	0.35	-0.34	0.30	0.11	-1.52	-1.75	0.05
<i>nx(e)</i>	-0.19	-1.63	0.95	0.16	1.47	-0.69	0.10	-1.05	-1.69	-0.45	0.23	0.10	0.38	-1.12	-0.19	2.12
<i>nx(r)</i>	-1.41	-0.72	-1.33	0.03	1.08	0.97	-0.45	1.47	-1.24	0.29	-0.02	0.55	-0.55	-1.00	-4.83	-0.68
<i>nx(vix_m)</i>	-0.19	0.12	-0.19	0.30	-0.87	-0.41	-1.65	1.47	-0.19	-0.41	0.03	-0.28	0.39	1.61	4.46	3.67
<i>nx(vix_sd)</i>	0.13	-0.35	1.11	-0.09	0.92	-0.45	1.68	-1.96	0.13	0.32	0.04	0.32	-0.83	-1.39	-0.66	-5.16
<i>nx(pi_wld)</i>	-0.16	-0.54	2.51	-0.20	-0.16	-0.66	0.15	1.69	-0.16	-0.02	0.09	-0.59	0.67	-1.00	-0.69	1.01
<i>fdi(nx)</i>	-0.06	1.05	2.45	0.30	-0.06	-0.03	0.63	-0.86	-2.44	-0.15	-0.40	0.15	-0.06	-0.06	-0.06	-0.06
<i>fdi(e)</i>	1.53	0.23	1.44	-0.26	1.34	0.28	-0.97	-0.10	-0.10	-0.31	0.14	-0.33	0.12	-2.11	1.32	-0.10
<i>fdi(vix_m)</i>	-2.73	1.84	-0.55	0.07	-0.71	-0.53	-0.47	2.02	-0.17	-0.20	0.04	0.00	-0.22	0.63	-0.17	5.30
<i>fdi(vix_sd)</i>	2.05	-1.01	0.45	-0.24	0.91	-0.34	0.28	-2.17	-0.08	0.53	-0.02	0.03	0.30	-0.08	1.49	-5.73
<i>fdi(pib_wld)</i>	-0.05	0.89	0.91	-0.39	-0.16	-0.20	0.74	2.03	-1.17	-0.82	0.24	0.64	0.14	-1.05	-0.05	-0.05
<i>fdi(pi_wld)</i>	1.99	-2.26	0.44	0.06	0.05	-0.16	0.93	-2.70	0.31	-0.44	0.41	-0.45	0.40	-1.02	-0.61	1.14
<i>cred(vix_m)</i>	-3.91	0.13	0.68	-0.45	1.42	-0.40	0.86	1.65	5.19	-0.12	0.16	0.04	-0.37	-1.80	-0.03	-0.03
<i>cred(vix_sd)</i>	1.61	0.20	0.29	0.19	-2.05	-0.04	0.29	-0.04	-5.79	0.45	-0.25	-0.15	0.34	1.78	1.07	-0.04
<i>cred(pi_wld)</i>	4.42	-1.73	-0.01	-0.09	-1.85	-0.08	-0.71	1.99	0.17	0.05	0.20	-0.23	0.26	2.26	-1.19	-1.06
<i>e(vix_m)</i>	0.79	1.64	1.80	0.07	1.73	0.04	-0.24	-0.55	-3.16	-0.22	-0.24	-0.11	-0.37	-2.33	1.15	3.84
<i>e(vix_sd)</i>	1.03	-0.66	-1.87	-0.10	-2.15	-0.03	0.41	0.17	3.81	0.25	0.21	0.34	-0.39	1.43	-0.68	-4.36
<i>r(vix_m)</i>	0.35	0.05	-0.05	0.61	0.85	0.03	-1.47	-1.64	2.73	-0.08	-0.05	-0.33	-0.30	-0.97	3.09	-3.26
<i>r(vix_sd)</i>	-1.94	0.74	-0.46	-0.19	-0.82	0.33	0.76	1.45	-3.75	-0.01	0.12	-0.14	-0.07	-1.17	-0.65	3.34
<i>r(pi_wld)</i>	0.79	-1.16	1.34	0.11	-0.50	-0.11	1.18	-3.34	0.08	0.06	-0.33	0.20	0.42	1.25	0.72	1.12
<i>BICm</i>	-1.67	-1.22	-1.53	-0.69	-1.00	-0.50	-0.63	-1.70	-1.58	0.73	0.69	0.18	0.08	-0.61	-0.19	-2.08
<i>BICsd</i>	-0.41	0.58	0.37	0.49	0.97	0.74	0.31	0.09	1.00	-0.50	-0.53	0.01	0.14	3.86	0.09	-0.30
<i>pib</i>	-0.57	-0.47	-0.13	-0.29	-0.60	-0.44	-0.50	-0.57	-0.57	-0.07	-0.26	-0.24	2.81	1.98	3.26	2.82

NOTA: A tabela apresenta o valor médio de cada variável dentro de cada cluster pela análise (a). As variáveis utilizadas nas análises de clusters e de fatores foram normalizadas em uma distribuição normal (0,1). A primeira coluna mostra as variáveis no formato [equação(variável explicativa)].

Tabela 10 – Valor médio das variáveis normalizadas em cada cluster - análise (c)

v	c 1	c 2	c 3	c 4	c 5	c 6	c 7	c 8	c 9	c 10	c 11	c 12	c 13	c 14	c 15	c 16
<i>pib(invest)</i>	-0.58	-0.02	0.51	-0.02	0.01	-0.33	-0.37	-1.70	-0.23	-0.52	-0.57	1.66	-0.02	0.66	-0.05	-0.02
<i>pib(nx)</i>	0.36	-0.16	1.44	0.58	-0.28	0.77	-1.15	-0.45	-0.29	-0.33	0.30	-0.45	-0.88	-1.50	0.70	0.01
<i>pib(fdi)</i>	0.07	0.16	-1.20	-0.05	-0.01	0.51	1.01	-2.59	0.00	-0.15	-0.29	-0.41	0.79	-0.09	0.83	0.16
<i>pib(cred)</i>	-0.53	0.03	0.27	0.86	0.59	-0.70	0.70	-1.20	-0.12	-0.17	0.09	-0.28	-0.36	0.89	0.06	1.33
<i>pib(e)</i>	-0.85	-0.18	0.60	-0.51	0.25	-0.34	0.54	-0.12	0.15	0.12	0.63	-0.69	0.61	-0.05	0.30	-0.17
<i>pib(r)</i>	0.10	-0.29	0.03	-0.56	-0.28	-0.59	-0.99	-0.98	0.10	-0.03	0.16	0.31	0.72	0.46	0.85	-0.08
<i>invest(pib)</i>	0.20	1.96	0.16	0.39	-0.54	-0.46	0.32	-0.28	-0.41	-0.28	-1.19	-0.28	-0.59	-0.28	-0.39	1.23
<i>invest(nx)</i>	-0.48	-0.22	-1.61	-0.92	0.77	0.47	-0.37	-0.53	0.22	0.49	0.71	1.04	-1.52	-0.21	-0.21	0.26
<i>invest(fdi)</i>	-0.58	-0.03	-0.18	-0.38	0.17	0.25	-1.76	-2.20	-0.07	0.35	-0.72	0.47	0.03	1.22	0.24	0.75
<i>invest(cred)</i>	-0.14	0.29	-0.09	0.93	-0.62	-0.50	0.55	-0.94	-0.66	0.56	-0.30	0.05	-0.13	-0.89	0.33	-0.06
<i>invest(e)</i>	1.20	0.21	-0.43	-0.09	0.07	-0.66	-1.70	0.52	0.26	-0.73	-0.26	0.03	0.21	1.87	0.51	0.05
<i>invest(r)</i>	0.03	-0.11	0.25	-0.04	0.50	-0.29	-0.91	-1.68	0.06	0.37	-1.03	-0.37	-0.65	0.80	0.90	0.09
<i>nx(fdi)</i>	1.18	0.33	-0.27	0.23	0.20	0.67	0.05	-0.73	-1.33	0.10	0.05	0.40	0.45	0.05	-1.71	-1.34
<i>nx(cred)</i>	1.80	-0.23	0.32	-2.09	0.44	0.62	0.03	-0.63	0.03	-0.10	-0.37	-0.30	-0.25	0.03	0.28	0.03
<i>nx(e)</i>	-0.66	-0.23	-0.61	-0.59	-0.47	0.08	-0.64	-1.47	0.39	0.43	0.56	0.60	0.20	0.35	0.35	-0.19
<i>nx(r)</i>	-0.59	0.13	0.90	-0.16	-0.08	-0.01	-0.79	-0.52	0.53	-0.04	1.34	-0.46	0.54	0.13	-0.90	0.41
<i>fdi(nx)</i>	0.19	-0.19	-0.20	-1.17	0.38	-0.06	3.92	0.87	-0.06	0.00	0.26	0.12	-1.23	0.80	-0.06	-0.06
<i>fdi(cred)</i>	-1.21	-0.21	0.32	0.20	-0.13	1.49	1.62	-0.51	-0.13	-0.14	-1.21	0.08	0.04	0.51	0.13	-0.13
<i>fdi(e)</i>	0.73	0.45	-0.57	-0.39	-0.31	-0.25	-0.10	-1.55	0.56	-0.15	1.72	0.03	-0.31	-2.63	0.16	-0.10
<i>fdi(r)</i>	-0.19	0.11	0.71	-0.72	-0.13	-0.01	0.10	-0.94	0.28	0.13	1.24	0.19	-0.83	0.10	-0.73	1.98
<i>cred(nx)</i>	-0.01	-0.12	-0.44	-0.01	0.19	-0.01	0.82	-1.75	-1.39	0.10	-0.47	0.18	0.89	2.77	-0.09	0.38
<i>cred(fdi)</i>	-0.21	0.17	0.02	-0.45	0.08	0.29	0.02	0.02	-2.49	-0.37	0.26	0.31	1.20	0.02	0.17	1.96
<i>cred(e)</i>	1.25	0.01	-0.22	-1.03	-0.36	-0.81	0.81	-0.90	0.24	0.47	-0.35	-0.27	0.30	0.37	0.45	-0.01
<i>cred(r)</i>	-0.35	-0.22	-0.64	-0.22	1.25	-0.18	0.91	0.92	-0.35	-0.38	0.51	-0.35	0.26	-1.19	0.22	1.12
<i>e(r)</i>	0.75	-0.16	0.03	-0.87	-0.75	-0.73	-0.40	1.01	0.07	0.70	0.53	0.22	-0.20	1.50	-0.23	-0.07
<i>r(e)</i>	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	-9.95

NOTA: A tabela apresenta o valor médio de cada variável dentro de cada cluster pela análise (c). As variáveis utilizadas nas análises de clusters e de fatores foram normalizadas em uma distribuição normal (0,1). A primeira coluna mostra as variáveis no formato [equação(variável explicativa)].

4 Política fiscal, mercado internacional e antecipação das flutuações do produto: uma abordagem usando aprendizado de máquina

O mundo viveu na primeira década do século XXI a crise econômica mais importante da história recente. A “grande recessão” que inicialmente se restringiu ao mercado financeiro teve efeitos significativos no lado real das economias em todo o mundo. Nesse cenário, duas forças motoras dos ciclos de negócios tiveram papel diferenciado em relação ao observado anteriormente. Estas são a política fiscal e o mercado financeiro. Tendo isso em mente, vamos usar esse quadro desenhado para testar um estimador em dois estágios, inspirado no relaxed lasso (Meinshausen [2007]), que tem como principal característica resumir um conjunto grande de informação para resolver um problema final. No caso do trabalho em questão, o problema é antecipar flutuações do produto dos países com informações disponíveis até determinado momento. Em específico, para verificar a relevância das duas forças motoras das flutuações do produto mencionadas anteriormente, comparamos a importância relativa de controles para estes grupos de informação com outros grupos de controles que são tradicionalmente tratados na literatura, teórica e empírica, como dentre os mais importantes. Ademais, além de resumir um volume grande de informação, nosso procedimento em dois estágios possui vantagens, sobre outros estimadores e procedimentos, no que se refere ao custo computacional e interpretabilidade das relações de causalidade (deriva facilmente o efeito marginal dos preditores). Os objetivos desse capítulo são dois. Primeiro é testar procedimentos que resumem de forma eficiente um volume grande de informação¹. Segundo é testar se houve mudanças significativas nos ciclos de negócios com a crise de 2008². Nosso trabalho tem duas conclusões principais. Primeiro que nosso estimador foi bem sucedido em resumir um volume grande de informação e teve um desempenho superior para previsão com dados novos a todos os demais utilizados no trabalho, inclusive os destacados na literatura de aprendizado de máquina por Fernández-Delgado et al. [2014]. Segundo que a interação com o mercado internacional foi uma característica muito mais importante para antecipar flutuações do produto que a política fiscal, o mercado financeiro e a produtividade da mão de obra durante o período analisado. Com isso, reforçamos os argumentos da literatura e do capítulo anterior dessa tese, mas rejeitamos as hipóteses do trabalho que novas dinâmicas econômicas observadas no período analisado alteraram significativamente a importância relativa das forças motoras do produto.

4.1 Introdução

A política monetária leniente e a fiscalização inadequada do chamado “shadow banking” nos EUA levou a uma bolha na economia norte americana. Com a crise de confiança que se instalou no início de 2007 no mercado “subprime” e culminou com a quebra do Lehman Brothers no 2º semestre de 2008, o mundo

¹ Esse é um problema recorrente e pertinente em muitas situações de economia e finanças.

² Fazemos isso com um raciocínio diferente do usado em econometria.

entrou em um período de crise. Essa crise teve um efeito contágio significativo com bancos e firmas tendo por um lado seu *funding* comprometido e por outro tendo que se desfazer de ativos ou corrigi-los nos balanços.

Como saldo final, a crise que inicialmente se restringiu ao sistema financeiro se espalhou para o lado real onde foi sentida pela sociedade em praticamente todo o mundo. Essa transmissão e potencialização se deram por diferentes canais, como restrição no canal de crédito, políticas de contingenciamento por parte de empresas, redução da renda disponível dos agentes da economia (demissões, reduções salariais, etc.).

Diante desse quadro recessivo, governos, autoridades e acadêmicos foram pressionados para propor estratégias para tratar a situação. Como os países não foram afetados da mesma forma e no mesmo momento, governos do mundo usaram diferentes instrumentos seja para restringir o contágio, principalmente através dos balanços de bancos e firmas, seja para reativar a economia.

O primeiro conjunto de medidas estava relacionado à política monetária. Bancos centrais usaram taxas de juros, redesconto e *quantitative easing* para estimular a economia, reativar o canal de crédito, fornecer liquidez para o sistema e retirar ativos comprometidos do sistema bancário³.

Em um segundo momento, a política fiscal ganhou espaço relevante no debate econômico para reativar as economias. Essa mudança de postura entre economistas e policy makers merece destaque porque a política fiscal tradicionalmente recebe um papel secundário e mais passivo, principalmente no meio acadêmico, enquanto que a política monetária é a política “ativa” do governo para atingir objetivos como estabilidade de preços ou reduzir volatilidade do produto⁴.

Dessa forma, em função dos eventos, choques e mudanças de políticas e comportamento dos agentes, observamos dinâmicas internas das economias mundiais (forças motoras, canais de transmissão, políticas, etc) após 2007 bem diferentes do registrado nas décadas anteriores e descritas nos manuais de economia.

Dentro das dinâmicas internas da economia, estamos particularmente interessados nas que explicam as variações do produto.

Tendo isso em mente, um resultado tradicional e robusto da literatura, tanto em termos teóricos como em trabalhos empíricos, é o destaque da interação do país com mercado internacional como um dos principais determinantes dos movimentos do produto e do crescimento (Barro [1991], Barro [1998], Sala-I-Martin et al. [2004] e Mayda and Rodrik [2005])⁵. De fato, os resultados do capítulo anterior chamam atenção para esse aspecto⁶.

Por outro lado, a literatura atribui uma importância menor à política fiscal nas flutuações do produto⁷. Smets and Wouters [2007] e Schmitt-Grohe and Uribe [2012], por exemplo, apresentam resultados que

³ Usamos o termo “quantitative easing” de forma genérica para descrever programas de compra de ativos do sistema bancário e expansão do balanço dos bancos centrais.

⁴ A literatura sobre essa relação entre política ativa e passiva é extensa e o assunto não será tratado aqui em maiores detalhes.

⁵ Argumentamos “de forma robusta” porque esse é um resultado consistente em trabalhos empíricos. No entanto, destacamos que não é uma regra genérica e, como muitas outras em economia, depende de questões específicas indiossincráticas de cada caso e momento no tempo.

⁶ Tal como definido na introdução da tese, o que chamamos de interação com o mercado internacional inclui questões como por exemplo, como a economia acomoda ou reage a choques do mercado internacional, como esses choques se propagam pela economia, como a economia compete no mercado internacional e como absorve know-how e tecnologia para alterar sua produtividade interna.

⁷ Logicamente, em casos (países) e momentos específicos na história, a política fiscal teve papel importante sobre o produto. Assim como outras variáveis econômicas, demográficas, culturais, etc. Mais uma vez, quando usamos esse argumento sobre

indicam que choques nos gastos do governo são pouco relevantes para os ciclos de negócios⁸.

Dessa forma, diante do argumentado sobre as características do período analisado, temos, em tese, dois importantes motores da economia mundial, que podem ser traduzidos em duas hipóteses a serem testadas e uma estrutura para balizar nossas análises. Esses dois motores da economia mundial em destaque no período são a política fiscal e o mercado financeiro e a estrutura para balizar nossas análises são as forças motoras do produto tradicionalmente destacados na literatura teórica e empírica. Esse é o ambiente que usaremos para testar nosso procedimento para escolha de decisão ótima no espaço de decisões possíveis.

Nosso estimador resume um volume grande de informação, reduzindo o espaço de possibilidades para uma situação computacionalmente factível. Ademais, além desta, possui principalmente quatro vantagens destacadas aplicadas ao problema de “big data” e ao que se propõe testar nesse capítulo. A primeira é que os resultados obtidos são robustos por testarem um número elevado de hipóteses. Segundo que apesar de testar muitas hipóteses, o procedimento é otimizado, reduz significativamente o custo de processamento no espaço de possibilidades e fornece um teste de consistência dos resultados. Terceiro que ele identifica o que é realmente importante (variáveis signals) ou mostra a importância relativa de cada preditor. Quarto que, dependendo da configuração utilizada, ele deriva facilmente o efeito marginal dos preditores. Vamos propor e utilizar uma abordagem baseada em técnicas de ML para identificar variáveis signals dentro de um universo grande de preditores. Nosso procedimento é inspirado no relaxed lasso (Meinshausen [2007]) e, na aplicação feita, combina estimadores de MQO, lasso e regressão em árvore.

Vamos usar indicadores do mercado de trabalho, investimento, da produtividade e da interação do país com o mercado internacional (que são teorias tradicionais em economia para explicar o produto) para estimar a importância relativa da política fiscal e do mercado financeiro. Para controlar para erros de mensuração na informação a ser modelada, o que poderia comprometer o testes das hipóteses, consideramos diversas dimensões da mesma variável. Em específico, aplicamos transformações não lineares, interações (para coeficientes variáveis) e decomposição de variáveis em componentes que controlam para diferentes dimensões da mesma variável.

A metodologia adotada é particularmente interessante para situações quando se têm muitos preditores ou informação disponível, como no mercado financeiro. Um exemplo importante é quando o objetivo final é a previsão com dados novos. Nesses casos, é importante usar o maior volume de informação disponível e considerar muitas possibilidades para minimizar erros de previsão.

Testamos nosso procedimento em três painéis de países distintos, cada qual maximizando a disponibilidade de informação de um determinado grupo de informação. Assim, nossas estimações foram feitas em três bases de dados distintas.

Nosso trabalho tem duas conclusões principais. Primeiro que nosso procedimento foi bem sucedido em resumir um volume grande de informação e teve um desempenho superior para previsão com dados novos a todos os demais utilizados no trabalho, inclusive os destacados na literatura de aprendizado de máquina por Fernández-Delgado et al. [2014]. Segundo que a interação com o mercado internacional foi uma característica muito mais importante para antecipar flutuações do produto que a política fiscal, o mercado financeiro e a produtividade da mão de obra durante o período analisado. Com isso, reforçamos os

a importância relativamente menor da política fiscal, estamos nos referindo a literatura sobre a relação entre política ativa e passiva e o efeito sobre a estabilidade do produto e da economia no longo prazo.

⁸ No entanto, apesar desses argumentos, é importante destacar que a política fiscal possui dimensões difíceis de serem mensuradas.

argumentos da literatura e do capítulo anterior dessa tese, mas rejeitamos as hipóteses do trabalho que as novas dinâmicas econômicas observadas no período analisado alteraram significativamente a importância relativa das forças motoras do produto.

Inicialmente contextualizamos a base de dados, o período de análise que motivam e servem de pano de fundo para o trabalho. Em seguida vamos fundamentar a estrutura econômica do trabalho na literatura com foco nos determinantes dos ciclos de negócios do produto dos países. Na terceira parte descrevemos a base de dados, especificação econométrica, o procedimento de estimação e como vamos avaliar as estimações e estimadores envolvidos. No quarto tópico, descrevemos os resultados tanto em termos de eficiência dos estimadores como das hipóteses levantadas. Por fim concluímos com uma revisão do trabalho.

4.2 Ciclos de negócios, crise de 2008 e política fiscal

Para abordar o problema proposto, vamos usar abordagens de aprendizado supervisionado onde a variável objetivo a ser modelada é o produto dos países da amostra.

Nesse sentido, a primeira pergunta a ser feita seria o que determina as variações do produto? Para tanto, é interessante verificar o que nos conta o período analisado e a literatura.

Assim, considerando o período analisado, e as discussões sobre a necessidade da política fiscal assumir um papel ativo na economia (Blanchard et al. [2010] e Blanchard et al. [2013]), o primeiro preditor, ou grupo de preditores, são os controles para a política fiscal. O efeito da política fiscal pode se dar tanto através de choques na demanda agregada (gastos, transferências ou tributos) como via oferta agregada (quando o governo aplica distorções no setor produtivo)⁹.

No entanto, a literatura destaca outras alternativas que reunimos em dois grupos principais: a interação com o mercado internacional e o mercado de trabalho.

Variações no mercado de trabalho estão ligadas a variações no produto nas teorias econômicas mais básicas. Além disso, indicadores do mercado de trabalho estão entre os principais *early warning signals* usados por analistas. Aqui também podemos pensar na produtividade da mão de obra, que apesar de estar ligada a outras linhas de pesquisa, têm uma ligação conceitual com o mercado de trabalho¹⁰.

O segundo grupo, a interação da economia com o mercado internacional, é um aspecto na literatura destacado como uma das principais forças motoras do produto. Essa relação é ressaltada tanto em trabalhos de modelos teóricos como em estimações empíricas (Barro [1991], Barro [1998], Sala-I-Martin et al. [2004] e Mayda and Rodrik [2005]).

Para tanto, a transmissão se dá por três canais. Primeiro através de movimentos de capitais para acomodar choques internos. Segundo através da abertura comercial e dos termos de troca e seus reflexos sobre a produtividade e a competitividade interna da economia. Terceiro através dos movimentos cambiais que afetam a renda real relativa da economia.

No que se refere aos fluxos de capitais internacionais, em termos teóricos, em uma economia aberta, os consumidores tentam suavizar seu consumo ao longo do tempo e para tanto usam o mercado internacional

⁹ Uma das principais linhas de trabalho que estudam o efeito da política fiscal estimam variações do multiplicador da política fiscal (Chahrour et al. [2012]).

¹⁰ A produtividade da mão de obra é uma importante força motora do produto tanto no curto prazo como no longo prazo. Essa variável foi particularmente importante no Brasil no período analisado por estar associado com o crescimento da economia pós 2003.

para acomodar choques temporários que afetam a renda. Assim, limitações de um país no acesso ao mercado financeiro internacional têm efeito de tornar o consumo e o produto mais voláteis. É o que ocorre nas economias em desenvolvimento. Em específico, choques negativos na renda, em uma situação de mobilidade imperfeita de capitais, levam a redução do consumo, com aumento dos juros reais e déficit comercial. Em um cenário mais próximo do real, onde dividimos os bens consumidos numa economia entre *tradables* e *non-tradables*, um choque positivo na renda leva a um ciclo de boom e recessão. No primeiro momento ocorre um aumento da renda que leva a um boom de consumo, com déficit na balança comercial, superávit na conta capital e apreciação no preço relativo dos *non-tradables*. No segundo momento ocorre uma contração no consumo, reversão na balança comercial e fluxos de capitais e depreciação do preço relativo dos *non-tradables*.

Do ponto de vista empírico, os fluxos de capitais internacionais são cíclicos e tem efeito significativo nos movimentos do produto dos países. Na verdade esses fluxos possuem particularidades importantes de serem considerados, pois podem afetar a direção do impacto no produto dependendo dos controles envolvidos. Três fatos estilizados relacionados a particularidades dos fluxos são destacados na literatura. Primeiro que o fluxo de capital internacional tende a seguir de países menos desenvolvidos para países mais desenvolvidos. Segundo que o investimento direto tende a seguir o caminho oposto. Terceiro que o saldo líquido destes dois fluxos citados anteriormente tende a ir de países menos desenvolvidos para países mais desenvolvidos (Reinhardt et al. [2013] e von Hagen and Zhang [2014]).

Assim, o conceito de fluxo de capital internacional possui componentes com características específicas que não necessariamente apontam para a mesma direção. De fato, apesar do impacto sobre o produto dos países de todos esses componentes, Aizenman et al. [2013] destaca que o que importa para o crescimento são os investimentos diretos.

Nesse sentido, uma segunda pergunta interessante seria o que determina esses fluxos? Basicamente, a literatura argumenta que os movimentos de capitais internacionais ocorrem, em grande parte, por reajustes nos portfólios dos investidores (Caballero et al. [2008] e Tille and van Wincoop [2010]). Por sua vez, para explicar reajustes em portfólio de investidores, devemos recorrer à literatura de finanças. Em finanças, para gestão de portfólio, um dos principais pilares é a dicotomia risco e retorno¹¹.

Dessa forma, diante de tudo exposto até agora, a base do nosso raciocínio aqui são os determinantes do risco e retorno das carteiras dos investidores. Em específico, após testar diferentes variáveis Byrne and Fiess [2016] chamam a atenção para a importância do retorno idiossincrático dos títulos dos países, os preços de commodities e o retorno de títulos de longo prazo de economias desenvolvidas.

Por outro lado, outra visão do problema dos determinantes dos fluxos de capital internacional divide esses fatores em dois grupos. O primeiro são os fatores de *push* (fatores externos) e o segundo são os fatores de *pull* (fatores domésticos). Os resultados dessa literatura tendem a apontar que os fatores de *push* como mais importantes, ou seja, fatores exógenos aos países. No entanto, esse resultado não pode ser generalizado por ser condicional à amostra utilizada (Fernandez-Arias [1996], Chohan et al. [1998] e Kim [2000]).

Em resumo, fatores que afetam o risco e retorno dos ativos dos investidores com acesso ao mercado internacional, provocam reajustes nas carteiras, que levam a movimentos nos capitais internacionais e que

¹¹ Dependendo de como estruturamos o problema do investidor, poderíamos pensar em fatores que afetam as preferências dos investidores para um mesmo nível de risco e retorno.

Tabela 11 – Países da base de dados

País	País	País
1. Austrália	17. Japão	33. Brasil
2. Austria	18. Korea	34. Chile
3. Belgica	19. Luxemburgo	35. China
4. Canadá	20. Holanda	36. Colômbia
5. República Checa	21. Nova Zelândia	37. Índia
6. Dinamarca	22. Noruega	38. Indonésia
7. Estônia	23. Polônia	39. México
8. Finlândia	24. Portugal	40. Rússia
9. França	25. Eslováquia	41. África do Sul
10. Alemanha	26. Eslovênia	
11. Grécia	27. Espanha	
12. Hungria	28. Suécia	
13. Islândia	29. Suíça	
14. Irlanda	30. Turquia	
15. Israel	31. Reino Unido	
16. Itália	32. Estados Unidos	

NOTA: A tabela lista os países da amostra. Os países escolhidos são os da OCDE mais alguns países em desenvolvimento.

por sua vez afetam o produto dos países.

Esses movimentos de capitais internacionais foram responsáveis por aquecer economias mas também por causar efeitos negativos. De fato, os “sudden stop” foram particularmente problemáticos para países com economias mais frágeis ao longo do tempo. Nessa linha, durante a crise, muito se discutiu sobre medidas macroprudenciais para controlar esses fluxos (Alberola et al. [2016]).

Além dos fluxos de capitais internacionais e condições do mercado financeiro internacional, outra dimensão importante é a abertura comercial. Por exemplo, políticas expansionistas (monetária, tarifária, comercial, etc) que afetem os preços relativos intertemporais e termos de troca causam distorções na economia que levam a um ciclo de superávit comercial e boom de consumo em um primeiro momento, mas a uma recessão no momento seguinte. No entanto, a literatura destaca que em situações como estas, quando o efeito das distorções é proporcionalmente menor que o chamado “efeito renda” da abertura comercial, então a economia não sofreria do ciclo de boom e recessão. Esse efeito renda geralmente é associado a ganho de produtividade no setor de tradables e é estimado em trabalhos que estudam micro dados do setor industrial (Tybout et al. [1991], Tybout and Westbrook [1995], Krishna and Mitra [1998], Hay [2001] e Fernandes [2007]). Ainda no que se refere ao segundo canal de transmissão descrito nos modelos teóricos, os trabalhos empíricos destacam, as condições dos termos de troca (Schmitt-Grohe and Uribe [2015]) e os movimentos da balança comercial (Garcia-Cicco et al. [2010]).

Por fim, não podemos deixar de mencionar o investimento que foi tratado no capítulo anterior.

4.3 Metodologia e estimação

Para nossas estimações e exercícios vamos usar uma base de dados econômicos e financeiros de países. Os países da amostra, listados na tabela (11), são os membros da OCDE mais alguns países em desenvolvimento. Estes foram escolhidos por serem significativos em termos de economia mundial ou por ter uma boa disponibilidade de dados. A base combina dados de domínio público (Banco Mundial, OCDE, BIS, dentre outros) e dados de acesso mais restrito (Bloomberg).

O período de treinamento básico começa em 2001 e vai até 2015 (dependendo do caso a série pode ser mais extensa e começar antes). Esse é um período muito interessante para análise, pois contempla informação da economia mundial antes da crise de 2008, durante e da recuperação desta crise.

Em termos teóricos o período é interessante para análise por três motivos principais. Primeiro por causa da disponibilidade de dados (quantidade e qualidade) que permite avaliar muitas dimensões de eventos. Segundo que foi um período com muitos eventos de stress, ou seja, situações limite (melhorar a definição de “eventos de stress”). Terceiro que foi um período rico em políticas e eventos.

4.3.1 Especificação econométrica

Para abordar o assunto, dado um painel de países com muitos preditores em relação ao número de observações, o trabalho começa com uma equação em painel que explica a variação do produto dos países dada pela equação (4.1).

$$\Delta y_{c,t} = \alpha \Delta y_{c,t-1} + \beta X_{c,t-1} + \delta Z_{t-1} + \varepsilon_{c,t} \quad (4.1)$$

Onde $y_{c,t}$ é a variação do produto do país c no tempo t e $\Delta y_{c,t}$ é a primeira diferença de $y_{c,t}$. $X_{c,t-1}$ é uma matriz de variáveis específicas do país c que incluem observações tanto no tempo t como em defasagens. Essas variáveis incluem dados de contas nacionais, preços, mercado financeiro, risco, etc. Z_{t-1} é uma matriz que incluem variáveis constantes no cross section dos países tanto no período t como defasadas. Essas variáveis incluem variáveis mundiais, como medidas de aversão ao risco internacional e observações específicas de países que afetam o painel todo (como PIB da China, ou crédito nos EUA, etc), ou PCA em todas as variáveis.

No entanto, a equação (4.1) é uma equação de referência. Em termos práticos, o trabalho vai usar especificações diferentes. Isso porque nosso objetivo aqui não é escolher o melhor modelo para prever produto e sim testar um método para determinação de $X_{c,t-1}$, onde $X_{c,t-1}$ seria o conjunto de variáveis que realmente importam (variáveis signals). Dessa forma, nosso objetivo aqui é testar se o nosso método de seleção de variáveis e modelos consegue tem um bom desempenho relativo a abordagens mais tradicionais e parcimoniosas como sugerido em Bayoumi et al. [2004].

Assim, nossa equação principal para teste é a equação (4.2) e o processo de seleção será aplicado sobre a matriz $X_{c,t}$.

$$\Delta y_{c,t} = \beta X_{c,t-1} + \varepsilon_{c,t}^X \quad (4.2)$$

Para balizar as estimações e os resultados das estimações da equação (4.2), tal como foi argumentado na introdução dessa tese, não vamos usar os melhores modelos do mercado para previsão de produto por motivos já apresentados. Aqui optamos por usar duas abordagens tradicionais, derivadas da equação (4.2), que abordam fragilidades de uma especificação do tipo da equação (4.2) e os estimadores de ML em evidência na literatura atualmente.

No que se refere às especificações derivadas da equação (4.2), a primeira são modelos $AR(p)$ individuais por países (equação (4.3)). Esse é a referência mais desafiadora, pois captura as idiosincrasias de cada país ao contrário de uma equação de painel que estima um “efeito médio”.

Tabela 12 – Dados

Grupos de Dados	Variáveis	Tipo	N	Fonte
Contas Nacionais	pib e seus componentes, comércio internacional	país	202	OCDE
Preços	preços, câmbio e juros	país	6	OCDE
Contratos Futuros - Prêmio de risco	títulos públicos, câmbio e mercado futuro	país	3	Bloomberg
Mercado Financeiros	índices se bolsa e demais indicadores de negócios	país	1	Bloomberg
Endividamento	público e privado, qualidade do crédito	país	2	Bloomberg
Crédito do Setor Público	operações de crédito do governo em detalhes	país	37	Banco Mundial
Mercado Financeiro Internacional	operações financeiras com o BIS, reservas	país	15	BIS
Balança de Pagamentos	balança comercial de fluxo de capitais	país	60	OCDE
Mercado de Trabalho		país	3	OCDE
CDS		país	1	Bloomberg
Aversão ao Risco Internacional	VIX, etc	global	2	Bloomberg
Preços de Commodities		global	183	Banco Mundial

NOTA: Os dados usados nessa tese estão em base trimestral. A coluna Grupo de Dados descreve o grupo conceitual de variáveis e, geralmente, é agrupada por fonte de dados. A coluna Variáveis lista ou explica as variáveis que fazem parte desse grupo. A coluna Tipo classifica os dados do grupo como “país”, quando a conta diferencia o valor para cada país em cada instante no tempo, e “global” quando a variável é constante no cross section. A coluna N enumera o número de variáveis dentro desse grupo. A coluna Fonte indica a fonte de dados das variáveis listadas naquela linha.

A segunda são modelos que usam “f” fatores comuns da base de preditores (equação (4.4)). Esses modelos são desafiadores como referência porque qualquer abordagem determinística que limita a informação de $X_{c,t}$ a um grupo pequeno de preditores corre o risco de ser “irresponsável”. Uma abordagem que reúne toda, ou a maior parte de um grupo grande de variáveis signals é mais “responsável”.

$$\Delta y_{c,t} = \alpha_{0,c} + \sum_{i=1}^p \alpha_{c,i} \Delta y_{c,t-i} + \varepsilon_{c,t}^{AR} \quad (4.3)$$

$$\Delta y_{c,t} = \delta_{0,c} + \sum_{i=1}^f \delta_{j,c} Z_{j,t-1} + \varepsilon_{c,t}^Z \quad (4.4)$$

Já com relação aos estimadores em destaque na ML, escolhemos três baseados nos resultados de Fernández-Delgado et al. [2014]. São eles: o SVM/SVR (equação (2.19)), o deep learning (equações (2.9) - (2.11)) e o *random forest* descrito no subtópico específico no tópico sobre regressão em árvore no capítulo 2.

4.3.2 Dados

O período de treinamento básico começa em 2001 e vai até 2015 (dependendo do caso a série pode ser mais extensa e começar antes). É uma base trimestral mais curta que a do capítulo anterior pois possui como objetivo estudar características do período e está sujeita a limitação de dados para observações anteriores a 2003 (como o CDS).

Os dados da análise desse capítulo tiveram dois momentos. Em um primeiro momento construímos uma base agregando outras e tínhamos pouco conhecimento sobre suas imperfeições. No segundo momento construímos uma base com variáveis previamente definidas com o objetivo de destacar alguns aspectos.

A base de dados do primeiro momento é descrita na tabela (12) e, assim como os códigos para montagem de base derivados do algoritmo (5), foram usados em uma abordagem mais computacional do problema onde se tinha entre os objetivos principais explorar a base, encontrar padrões e testar a eficiência dos códigos.

Tabela 13 – Variáveis Base - nível e índices

Número	Nome	Código	Fonte	Medida
Contas Nacionais - fluxos internacionais (List NX)				
1	exportações líquidas	nx	OCDE	usd
2	importações de serviços	m_s	OCDE	usd
3	importações de bens	m_g	OCDE	usd
4	exportações de serviços	x_s	OCDE	usd
5	exportações de bens	x_g	OCDE	usd
6	produto	pib	OCDE	usd
7	reservas	res	OCDE	usd
8	conta derivativos fluxo de capitais	der	OCDE	usd
9	. curto prazo	fcp	OCDE	usd
10	. longo prazo	flp	OCDE	usd
11	. saldo	fsd	OCDE	usd
Governo - política fiscal (List G)				
	impostos			
12	. bens	t_g	OCDE	usd
13	. capital	t_k	OCDE	usd
14	. valor agregado	t_i	OCDE	usd
15	. produção	t_y	OCDE	usd
16	. renda	t_w	OCDE	usd
17	gastos do governo	c_g	OCDE	usd
Mercado de Trabalho e Consumidor (List L)				
18	emprego	emp	OCDE	%, $\frac{emprego.total}{pop.total}$
19	custo da mão de obra	wage	OCDE	usd
20	salário (índice)	wp_i	Bloomberg	usd, índice 2008.Q4=100
21	salário (variação)	wp_d	Bloomberg	$ln\left(\frac{wp_{i,t}}{wp_{i,t-1}}\right) * 100$
22	produtividade da mão de obra (índice)	a_i	Bloomberg	índice 2008.Q4=100
23	produtividade da mão de obra (variação)	a_d	Bloomberg	$ln\left(\frac{a_{i,t}}{a_{i,t-1}}\right) * 100$
24	produtividade da mão de obra estimada (índice)	a_e	OCDE	ver nota (*)
25	produtividade da mão de obra estimada (variação)	a_ed	OCDE	$a_{e,t} - a_{e,t-1}$
26	crédito ao consumidor	cred	Bloomberg	$ln\left(\frac{credito_{usd,t}}{credito_{usd,t-1}}\right) * 100$
27	confiança do consumidor	conf	Bloomberg	dispersão, indiferente=0
Outros (List O)				
28	renda disponível (índice)	yd_i	Bloomberg	usd, índice 2008.Q4=100
29	renda disponível (variação)	yd_d	Bloomberg	$ln\left(\frac{yd_{i,t}}{yd_{i,t-1}}\right) * 100$
30	consumo privado	c_c	OCDE	usd
31	formação bruta de capital fixo	i_i	OCDE	usd

NOTA: A tabela lista as variáveis de referência do trabalho que não fazem parte do mercado financeiro. Essa é uma base de referência porque essas variáveis serão usadas para gerar a base final de preditores através de transformações aplicadas nessas variáveis de referência. Estas variáveis são divididas em 4 grupos que controlam para fatores importantes para variações do produto no curto prazo para o período considerado. O campo **código** mostra o nome da variável nas estimações e campos **fonte** e **medida** são autoexplicativos. (*) A produtividade ($a_{e,t}$) é estimada usando dados trimestrais da OCDE na equação $ln\left(\frac{pib_t}{pop_{14-64,t}}\right) = \beta_i ln\left(\frac{i_{i,t}}{pop_{14-64,t}}\right) + \beta_l ln\left(\frac{emprego}{pop_{14-64,t}}\right) + a_{e,t}$.

Em um segundo momento, as variáveis foram filtradas e trabalhadas para testar características pertinentes ao objeto de aplicação empírica do trabalho.

A tabela (13) e a tabela (14) mostram as variáveis base utilizadas nesse segundo estágio. Essas variáveis são base ou de referência porque foram utilizadas em diversas transformações para gerar a base de preditores utilizada nas estimações. A tabela (15) descreve essas transformações.

Tabela 14 – Variáveis Base - mercado financeiro

Número	Nome	Código	Fonte	Medida
Taxas e índices (List F1)				
32	câmbio	e	bloomberg	$\frac{usd}{moeda.local}$
33	câmbio futuro (1 ano)	e_f	bloomberg	$\frac{usd}{moeda.local}$
34	juros do Banco Central	r	bloomberg	% no trimestre
35	juros - títulos públicos de 3 meses	r_3	bloomberg	% no trimestre
36	CDS	cds	bloomberg	pontos base da carteira do comprador
37	principal índice de bolsa de valores	sm	bloomberg	pontos
38	bolsa de valores futuro	sm_f	bloombeg	pontos
Spreads (List F2)				
39	câmbio	se		$set = e_{f,t} - e_t$
40	juros	sr		$sr_t = r_{3,t} - r_t$
41	bolsa de valores	s_sm		$s_{sm,t} = \ln \left(\frac{\Delta sm_{f,t}}{\Delta sm_t} \right) * 100$
42	juros dos EUA	sr_us		$sr_{us,t} = r_{us,t} - r_t$
43	bolsa de valores do EUA	s_sm_us		$s_{sm,us,t} = \ln \left(\frac{\Delta sm_t}{\Delta sm_{us,t}} \right) * 100$
44	CDS dos EUA	s_cds		$s_{cds,t} = cds_t - cds_{us,t}$

NOTA: A tabela lista as variáveis do mercado financeiro usadas no trabalho.

Tabela 15 – Transformações nas variáveis de referência

Transformação	Equação	Código	Listas
Variáveis em Diferença (List D)			
a. primeira diferença	$\frac{[v_t] - [v_{t-1}]}{[v_t] - [v_{t-1}]}$	$D[v_t]$	List NX, List G, List O, List F1, List F2
b. variação percentual	$\left(\frac{[v_t] - [v_{t-1}]}{[v_{t-1}]} - 1 \right) * 100$	$LN[v_t]$	List NX, List G, List O, List F1, List F2
Variáveis Nível (List LVL)			
c. razão do PIB	$\frac{[v_t]}{pib_t}$	$[v_t]Y$	List NX, List G, List L, List O
d. razão do PIB (variação)	$\Delta [v_t]Y$	$[v_t]YD$	List NX, List G, List L, List O
e. componentes	$[v_t] = b_0 + \sum_{i=1}^F (b_i \Delta [v_{fin,t}]) + \varepsilon_t^v$		$[v_{fin,t}] \in \{r_t, e_t, sm_t, cds_t, sr_t, se_t, sm_t, ssm_t, srus_t, ssmus_t, scdsus_t\}$
e.1. autônomo	$b_0 + \varepsilon_t^v$	$[v_t]Ba$	List NX, List G, List L, List O
e.2. juros	$b_r \Delta r_t$	$[v_t]Br$	List NX, List G, List L, List O
e.3. câmbio	$b_e \Delta e_t$	$[v_t]Be$	List NX, List G, List L, List O
e.6. bolsa de valores	$b_{sm} \Delta sm_t$	$[v_t]Bsm$	List NX, List G, List L, List O
e.7. cds	$b_{cds} \Delta cds_t$	$[v_t]Bcds$	List NX, List G, List L, List O
e.8. spread de juros	$b_{sr} \Delta sr_t$	$[v_t]Bsr$	List NX, List G, List L, List O
e.9. spread de câmbio	$b_{se} \Delta se_t$	$[v_t]Bse$	List NX, List G, List L, List O
e.10. spread do bolsa de valores	$b_{ssm} \Delta ssm_t$	$[v_t]Bssm$	List NX, List G, List L, List O
e.11. spread dos juros dos EUA	$b_{srus} \Delta srus_t$	$[v_t]Bsrus$	List NX, List G, List L, List O
e.12. spread da bolsa de valores dos EUA	$b_{ssmus} \Delta ssmus_t$	$[v_t]Bssmus$	List NX, List G, List L, List O
e.13. spread do CDS dos EUA	$b_{scdsus} \Delta scdsus_t$	$[v_t]Bscdsus$	List NX, List G, List L, List O
Variáveis Compostas (List CMP)			
f. razão de variáveis	$\frac{[v_{1,t}]}{[v_{2,t}]}$	$[v_{1,t}]TO[v_{2,t}]$	List LVL, LST L e LST D
g. variação da razão de variáveis	$\Delta ([v_{1,t}]TO[v_{2,t}])$	$[v_{1,t}]TO[v_{2,t}]D$	List LVL, LST L e LST D
Variáveis ao quadrado (List S)			
h. polinômio do segundo grau da variável	$[v_t]^2$	$S[v_t]2$	List CMP, List LVL, List D, List NX, List G, List L, List O

NOTA: A tabela lista as transformações aplicadas nas variáveis de referência. O primeiro grupo, chamado de List LVL, calcula relações de cada variáveis da tabela (13) (List NX e List G) em relação ao PIB (a. e b.) e decompõe essas variáveis em componentes (c.). O segundo grupo (list CMP) interage as variáveis do List LVL.

4.3.3 Variáveis Signals e Noise

Conceitualmente, a diferença entre uma variável signal e uma noise é que a primeira tem informação e a última apenas ruído. Em termos de modelagem de uma variável objetivo, como é no caso de aprendizado supervisionado, uma variável signal é aquela que tem informação explicativa sobre a variável dependente do modelo.

Esses conceitos são aplicáveis a qualquer caso ou área onde se estabelece uma relação de causa efeito entre variáveis. Esses termos são vistos de forma mais usual em áreas como processamento de dados e sinal. Nesses casos o analista ou pesquisador tem uma abordagem mais computacional ou estatística do problema. Isso porque em situações assim, não é simples estabelecer um modelo teórico com relação causa efeito na variável objetivo de modelagem e previsão.

Por outro lado, quando estamos no campo das ciências aplicadas como engenharia, física ou economia, geralmente temos modelos teóricos que definem ou indicam essas relações de causalidade. Com isso, temos priors de quais são variáveis signals e quais são noise antes de acessar dos dados.

Um exemplo comum em economia de aplicação dos conceitos de variáveis signals e noise é quando utilizamos algum filtro de sazonalidade em uma variável. Nesse caso, o comportamento não cíclico é o signal (informação não sazonal da variável) e o cíclico é o noise (parte da variável objetivo que não tem relação com, ou não guarda informação sobre, o movimento da variável independente de choques sazonais).

Outro exemplo, mais genérico, é a especificação de um modelo econométrico para uma variável explicativa. Neste, as variáveis explicativas significativas são signals e o termo de erro é uma variável noise. É com base nesse segundo exemplo que vamos definir o conceito de signal e noise utilizado nessa tese.

A nossa definição de variável signal é mais criteriosa que a definição acima. Em um ambiente de regressão tradicional, variável signal é aquela significativa e com elevado poder preditivo. Como definição mais geral, independente do estimador utilizado, variável signal é aquela com elevado poder preditivo em relação às demais.

Essa definição é baseada no procedimento de solução do estimador de lasso usando o algoritmo de *lars* (algoritmo (1)). No caso específico de regressões com muitos preditores, a diferença do que é signal e noise é especialmente problemática. Nesses casos, muitos coeficientes possuem distorções. Esse é o caso principal desse capítulo.

4.3.4 Montagem de base, variáveis signals e validação cruzada

A primeira etapa consiste em, a partir de uma base desconhecida, definir painéis balanceados ótimos. Por base desconhecida queremos dizer que não conhecemos a base em suas imperfeições (gaps nos dados principalmente e suas consequências). Por painéis balanceados ótimos queremos dizer, a partir de uma base como imperfeições, quais seriam os melhores painéis balanceados que poderia ser montados.

Nessa primeira etapa tivemos duas abordagens. A primeira exploratória descreve procedimentos mais computacionais para uma base desconhecida. A segunda que se baseia nos resultados de primeira, usa uma base conhecida com características definidas que queremos destacar.

4.3.4.1 Montagem de base

4.3.4.1.1 Primeira abordagem: etapa exploratória

Dessa forma, a primeira abordagem é criar uma base com muitos preditores e testar uma família de algoritmos que possuem como objetivo construir painéis balanceados ótimos. A descrição dessa base está na tabela (12).

O problema de montagem de base (ou busca) de painel balanceado dentro de uma base de dados desconhecida é mais complexo para base de dados maiores¹². Nesse sentido, a preocupação maior aqui é otimizar esse processo de busca de painéis balanceados dentro da base para restringir o trabalho a um sub-espaço pequeno do espaço de possibilidades total.

Os algoritmos utilizados foram baseados no algoritmo (5) que é a base da lógica do procedimento de construção de painéis balanceados a partir de uma base grande e desconhecida. Esses algoritmos podem ser divididos em três passos principais. O primeiro é construir uma matriz com frequência e demais informações das variáveis da base (pontos 1. e 2. do algoritmo (5)). O segundo é retirar variáveis que possuem baixa probabilidade de entrar no painel. Para tanto, usamos análise de fatores na matriz de frequência e informações a fim de identificar variáveis que possuam um comportamento (em termos de frequência) pouco similar às demais (ponto 3. do algoritmo (5)). Tendo aplicado filtros para excluir variáveis indesejadas em um primeiro momento, aplicamos análises de clusters na matriz de frequência e informação das variáveis remanescentes para agrupar indivíduos (países) com disponibilidade similar de frequência de variáveis (ponto 4. do algoritmo (5)).

No entanto, é importante destacar que o algoritmo (5) é aplicado estritamente em apenas um caso e que esse caso não gera os resultados mais interessantes. Em todas as outras aplicações esse algoritmo base foi flexibilizado, principalmente com critérios de busca, e adaptado para capturar aspectos interessantes a determinadas perguntas ou conjunto de preditores que se queira destacar. A principal vantagem do algoritmo (5), e suas variantes, é a automatização que podem ser usados em processos iterativos para cobrir espaços factíveis de possibilidades. Nesse sentido, a aplicação desses procedimentos é mais indicada em etapas exploratórias do trabalho onde não se tem um conhecimento aprofundado da base, como foi o caso da primeira abordagem da tabela (12).

4.3.4.1.2 Segunda abordagem: determinantes do produto

A segunda abordagem usa os resultados da primeira abordagem para definir um painel mais compacto com um conjunto menor de variáveis de referência. Com base nesse conjunto de referência, aplicamos transformações para obter novas informações e/ou retirar componentes dessas variáveis a fim de construir a base final de preditores.

Sendo assim, a segunda etapa inicia com uma base de referência. Apesar de trabalharmos aqui com uma base de dados menos “desconhecida” que a da primeira etapa, esta ainda possui imperfeições. Dessa forma, nessa etapa também tivemos que aplicar códigos para encontrar painéis balanceados ótimos.

¹² Chamamos uma base de desconhecida porque não sabemos onde estão as imperfeições (gaps) dessa base ou não é simples de encontrar e tratá-los.

Algorithm 5 Construção de painel balanceado

1. definição de um **período de referência**: t_{base} ;
 2. cálculo da **frequencia de cada preditor** no período de referência: $f_{c,t_{base}}$;
 3. **exclusão de variáveis não interessantes**: critério é a uniqueness (ou variância específica) em um PCA na base de frequencia (passo 2);
 4. **agrupar indivíduos similares**: análise de cluster na matriz de frequencia das variáveis do passo 3;
-

Essa base de referência segue o que foi discutido no tópico 2.2. Ciclos de Negócios, crise de 2008 e política fiscal. Dessa forma, dividimos esses conjuntos em cinco grupos de preditores. Cada grupo de referência contém variáveis para controlar para as informações que esse grupo representa.

O primeiro grupo de variáveis de referência (*List NX*) inclui informações de fluxos internacionais de bens, serviços e capitais (Contas nacionais - fluxos internacionais (tabela (13))). Esse grupo de informação é nosso principal ponto de referência para testar nossas hipóteses. O segundo grupo (*List G*) se refere à política fiscal (Governo - política fiscal (tabela (13))). O terceiro grupo se refere ao mercado de trabalho (Mercado de trabalho e mercado consumidor (tabela (13))). O mercado de trabalho (*List L*) é frequentemente apontado como uma fonte de *early warning signals* para a economia. O quarto grupo (*List O*) chamamos de outros e incluem variáveis como investimento que foi destacada no capítulo anterior. O quinto (*List F1* e *List F2*) é o mercado financeiro (tabela (14)) que sempre está em um lugar de destaque no debate econômico sobre crises e desempenho da economia.

Essas tabelas reúnem um total de 43 variáveis explicativas no grupo de referência. A partir dessas variáveis, realizamos uma série de transformações e interações para construir a base final de preditores. Essas transformações e interações usam as variáveis da List NX, List G e List O (21 variáveis) e são descritas na tabela (15).

Com o primeiro grupo de transformações criamos um grupo de variáveis chamado de List D reúne duas diferenciações de List NX, List G, List O, List F1 e List F2 que gera um conjunto de 66 preditores. Note que não incluímos List L porque vários de seus componentes já estão diferenciados.

O segundo é List LVL. Além da razão do produto e da primeira diferença dessa razão (*c.* e *d.*) vale chamar atenção para o que chamamos de “componentes” (*e.*). Com elas queremos decompor cada variável em 13 componentes e com isso incluir informação do mercado financeiro não diretamente, mas como a sensibilidade de variáveis reais. O autônomo (*c.1*) é o intercepto mais o termo de erro. Os demais são as sensibilidades às variáveis do mercado financeiro de List F1 e List F2. Essas transformações totalizam 300 variáveis.

A relação de variáveis compostas é formada por dois tipos de interações entre as variáveis de List LVL que totaliza 312840 variáveis.

Por fim, incluímos o polinômio de segundo grau de todas as variáveis consideradas no trabalho em List S em um total de 313216 preditores.

A base de preditores final é composta pelas variáveis de List NX, List G, List O, List L, List D, List LVL, List CMP e List S em um total de 626432 preditores.

4.3.4.2 Seleção de variáveis signals

Tal como definido para essa tese, variável signal é uma variável com elevado poder explicativo. Por sua vez, com elevado poder explicativo queremos dizer que o efeito marginal dessa variável é elevado em relação às demais variáveis em termos de valor absoluto.

Dessa forma, essa etapa do trabalho consiste basicamente em aplicar um processo de seleção de variáveis para extrair as variáveis signals do conjunto de preditores totais.

Para tanto, vamos usar três metodologias de ML que executam um processo de seleção de variáveis de forma otimizada. E por um processo de seleção de variáveis otimizada, queremos dizer que este não busca em todo o espaço de possibilidades tal como foi feito no capítulo anterior.

Esses três métodos são o MQO, lasso e regressão em árvore. O lasso e a regressão em árvore executam automaticamente seleção de variáveis tal como discutido no capítulo 2. O MQO será usado para identificar os preditores com maiores coeficientes relativos. Isso é possível em um ambiente onde os preditores possuem magnitude e medida comparável e os coeficientes não são viesados. A idéia é similar a forma como o lasso executa a seleção de variáveis.

Em específico, o fazemos estimações por MQO com preditores e variável objetivo (produto dos países (PIB)) normalizadas ($N(0, 1)$). O segundo é regressão lasso estimada por *lars* (algoritmo (1)). O terceiro é regressão em árvore estimada com algoritmo CHAID (algoritmo (4)).

Na verdade esses três métodos são usados como filtros em um processo que possui como objetivo organizar uma seleção justa sem distorções por (i) sobreposição de efeito e (ii) viés por multicolinearidade.

A esse processo chamados de algoritmo *seleção de variáveis signals* (SVS) e está descrito de forma esquematizada na figura (5).

O algoritmo *SVS* tem como input o conjunto de variáveis V e como output V^* , sendo composto por quatro processos. Esses processos são repetidos algumas vezes para refinar o conjunto final de V^* .

Dado um input V primeiro subdividimos esse conjunto em n conjuntos (processos). A seguir subdividimos cada V_n em outros conjuntos com baixa multicolinearidade dentro do grupo (processo 2). Por fim, aplicamos um método de identificação de variáveis signals dentro de cada V_{n,k_n} para gerar V_{n,k_n}^* tal como definido na equação (4.5) (processo 3)¹³. A última etapa consiste em reunir todos os V_{n,k_n}^* e gerar V^* (processo 4).

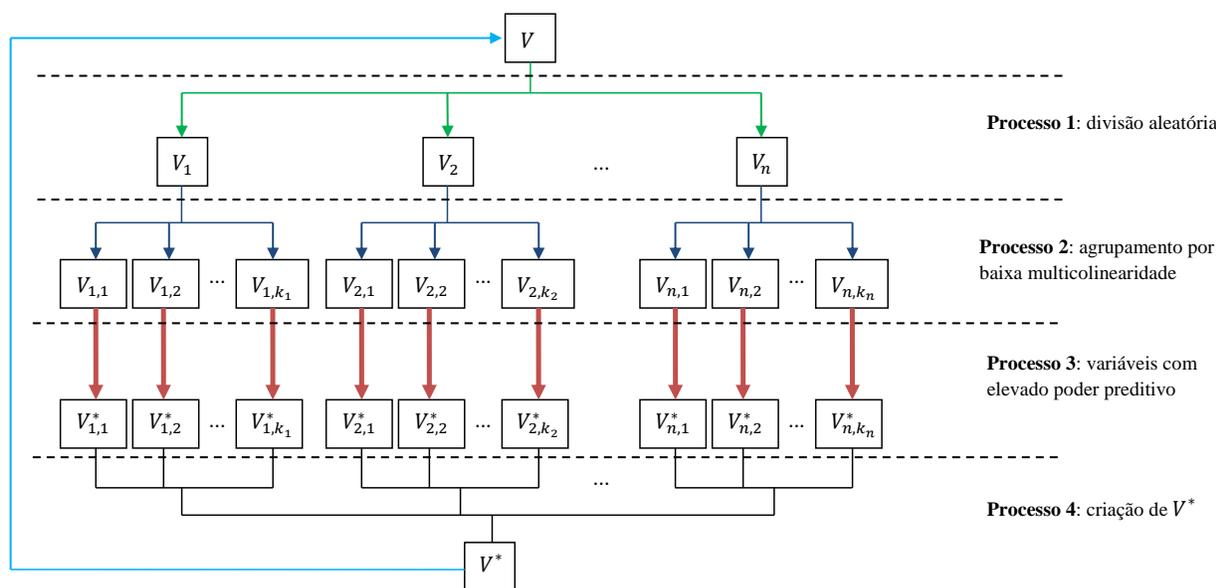
Especificamente para cada lista de variáveis testadas (com uma determinada estimação), as variáveis escolhidas como mais importantes fazem parte do conjunto V^* definindo na equação (4.5).

$$V^* = \left\{ v \in V \mid |b_v| \geq b^{max} - \sigma_{b,abs} \quad \forall \quad b_v \in B \right\} \quad (4.5)$$

Onde V é o conjunto de todas as variáveis (v) testadas em uma estimação (regressão) e B é o vetor de betas significativos dessas variáveis na regressão. V^* é o conjunto de todas as variáveis escolhidas como com elevado poder preditivo, $|b_v|$ é o coeficiente da variável v em valor absoluto, b^{max} é o maior beta (em valores absolutos) de B e $\sigma_{b,abs}$ é o desvio padrão de $|B|$. Assim, de acordo com a equação (4.5), uma variável v vai ser escolhida como com elevado poder preditivo em uma lista de variáveis ($v \in V^*$) se o valor absoluto do coeficiente dessa variável ($|b_v|$) for maior ou igual ao maior beta dessa regressão

¹³ No caso da regressão em árvore, a regra de identificação das variáveis signals dentro de V_{n,k_n} é diferente devido a natureza do algoritmo de solução utilizado (CHAID).

Figura 5 – Seleção de variável signal (SVS)



NOTA: A figura esquematiza o algoritmo de *seleção de variáveis signals (SVS)*. A ideia do código é, para um conjunto inicial de preditores V , selecionar um subconjunto de signals (V^*) em um processo sem distorções nos confrontamentos.

menos um desvio padrão do vetor de valores absolutos dos coeficientes dessa regressão ($b^{max} - \sigma_{b,abs}$). Em outras palavras, para uma determinada lista de variáveis V , as variáveis com elevado poder preditivo são aquelas com betas dentro de um intervalo entre o maior beta da regressão (b^{max}) e um desvio padrão, do vetor de betas, para baixo. No caso do painel com efeitos fixos, B vai ser formado por aqueles betas com $p - valor < 0.1$. Para a regressão lasso, B vai ser formado pelos coeficientes selecionados no ponto ótimo do path¹⁴.

A figura (5) esquematiza o algoritmo para identificação das variáveis signals de acordo com a definição de variável usada na tese.

4.3.4.3 Classificação por validação cruzada

A última etapa consiste em classificar e escolher modelos por validação cruzada. Esse é o “receituário tradicional” de ML. Esse procedimento de validação cruzada é usado em vários métodos para calibragem de parâmetros e modelos.

No entanto, por estarmos trabalhando com séries temporais, o que chamamos aqui de validação cruzada é diferente do que foi definido no capítulo 2. Talvez fosse mais adequado dizer que usamos como critério de classificação do erro *out of sample*. Isso porque, como estamos usando dados trimestrais em “painéis longos”, mesmo que sem o fator dinâmico da equação (4.1), devemos considerar a autocorrelação serial dos resíduos.

¹⁴ Ponto ótimo no path do lasso é aquele modelo (conjunto de preditores) com menor estatística Cp (tabela (1)) em um processo lars.

Algorithm 6 Classificação por validação cruzada (CVC)

1. particionar $V^* = V^{1,*} \cup \dots \cup V^{H,*}$;
 2. identificar $v^{h,*} \in V^{h,*} \therefore \varepsilon^{K,h,*} \leq \varepsilon^{K,i,*}$ tq $h \neq i$ e $v^{h,*}, v^{i,*} \in V^{h,*}$;
 3. estimar $\bar{v}^j = \bigcup_{l=1}^4 v^{l,*}$ tal que:
 - a) $\bar{v}^j \subset \bar{V}^*$;
 - b) $\bar{\varepsilon}^{K,j,*} \leq \bar{\varepsilon}^{K,m,*}$ tal que $j \neq m$ e $\bar{v}^{j,*}, \bar{v}^{m,*} \in \bar{V}^*$;
 - c) $\bar{V}^* = \{v^{1,*}, \dots, v^{L,*}\}$;
-

Dessa forma, como as observações estão ligadas por uma sequencia temporal, não podemos usar procedimentos de validação cruzada descritos no capítulo 2 ¹⁵.

Assim, vamos manter a idéia de usar o erro de validação para testar modelos, mas vamos calculá-lo de forma diferente do que foi apresentado. Em séries temporais, o procedimento usual para esses casos que é re-estimar o modelo $g(\cdot)$ para K amostras de treinamento cada uma indo de $t = 0$ até $t = N - 1 - k$, onde $k \leq K$. O erro *out of sample* (erro de validação) calculado dessa forma é dado pela equação (4.6).

$$\varepsilon^K = \sum_{k=1}^K (y^k - X^k \beta^k) \quad (4.6)$$

Onde y^K é o vetor de y observado que vai de $t = 1$ até $t = N - 1 - k$. X^k é a matriz de preditores que segue a mesma lógica. β^k é o vetor de coeficientes ($g_k(\cdot)$) que é o modelo estimado para uma amostra de treinamento que vai de $t = 1$ até $t = N - 1 - k$.

Tendo isso em mente, de forma objetiva, nessa etapa, vamos refinar V^* a fim de testar a eficiência da abordagem do passo anterior. Enquanto o algoritmo *SVS* usou critérios de informação *in sample*, tal como argumentamos acima, aqui vamos usar critérios de informação *out of sample*.

O procedimento consiste em, a partir de V^* , identificar H conjuntos ótimos de no máximo um número qualquer de I preditores. Por “ótimos” queremos dizer que cada conjunto h ($h \leq H$) possui o menor erro (equação (4.6)) dentro do subconjunto h de V^* ($V^{h,*}$).

Com isso temos, por fim, um conjunto ótimo de até 4 $v^{l,*}$, sendo que $v^{l,*}$ pode ser \emptyset no ponto 3. do algoritmo (6).

Esse conjunto final de preditores (\bar{v}) é um conjunto com um dos melhores desempenhos preditivos dentro do conjunto inicial V . Falamos “um dos melhores” porque cada estimação de \bar{v} gera um resultado diferente que tende a convergir para o subconjunto mais importante de preditores para um número elevado de repetições

Por fim, vamos comparar o desempenho preditivo *ou of sample* de \bar{v} com a média dos erros dos modelos *AR(p)* individuais por país (equação (4.3)), com equações com fatores comuns dentro de V^* (equação (4.4)) e com os estimadores de ML selecionados e mencionados anteriormente.

¹⁵ Note que esse argumento foi flexibilizado no capítulo anterior. Naquela situação foi possível porque além de não usarmos um termo autoregressivo na variável depende, os dados são anuais e por isso ignoramos a autocorrelação temporal dos resíduos.

4.4 Resultados

Os resultados do trabalho foram organizados para responder duas questões.

A primeira é se conseguimos resumir um volume grande de informação de forma eficiente com o nosso procedimento proposto para satisfazer um critério definido por uma determinada função perda. No caso em questão a função perda utilizada é o erro de validação que no nosso trabalho é o erro out of sample. Para responder a esta questão vamos comparar o desempenho do nosso preditor com outros três grupos de preditores em praticamente igualdade de condições¹⁶.

A segunda questão é se a política fiscal e o mercado financeiro estão entre os melhores preditores para antecipar flutuações do produto com dados novos de acordo com o que aprendemos no período considerado. Nessa segunda etapa também queremos saber se conseguimos definir um conjunto pequeno de “principais” forças motoras que seriam consistentes com nos nossos códigos. Se esse grupo existe, gostaríamos de saber se é possível tirar alguma lição deles.

A metodologia se resume em um primeiro momento definir alguns painéis balanceados ótimos. A partir de um determinado painel balanceado, temos um conjunto de preditores que chamamos de V . Vamos construir três painéis que maximizam a disponibilidade de algumas características que queremos destacar (política fiscal, mercado financeiro e internacional, e produtividade da mão de obra, respectivamente). Com isso temos V^1 , V^2 e V^3 , para os painéis 1, 2 e 3 respectivamente, que para facilitar e simplificar o entendimento chamamos genericamente de V .

A segunda etapa é, a partir de V , encontrar um subgrupo de variáveis signals que realmente são relevantes. Essa é uma etapa onde a velocidade de processamento é muito importante. Para tanto usamos o algoritmo da figura (5) para identificar $V_{code,step}^*$, onde $step$ é o número de iterações do algoritmo SVS , e $code$ é o processo de seleção de variável utilizado.

A terceira etapa consiste em tentar achar um subconjunto pequeno de preditores com elevado poder preditivo ($\bar{v}_{code,step}^j$) que simplificaria interpretações e aprendizado. Para balizar e quantificar o quão bom são os modelos parcimoniosos, usamos grupos de modelos de referência (equação (4.3) e equação (4.4)) e alguns estimadores da literatura de ML destacados anteriormente.

4.4.1 Eficiência preditiva com dados novos

Para abordar essa questão vamos estimar o erro out of sample (equação (4.6)) para o procedimento proposto ($\bar{v}_{code,step}$) e para três grupos de estimadores que utilizaremos como balizadores (ARs, fatores comuns e aprendizado de máquina), onde cada um representa um desafio diferente de acordo com o argumentado no tópico 4.1 Especificação Econométrica nesse capítulo.

Nosso estimador final é dado por $\bar{v}_{code,step}$ definido pelo algoritmo CVC (algoritmo (6)). Onde $code$ é o procedimento usado para identificar o conjunto de variáveis signals ($V_{code,step}^*$) pelo algoritmo SVS (figura (5)). Assim, $code \in \{mqo, lasso, tree, join\}$ onde $join$ são as variáveis signals identificados nos três procedimentos conjuntamente e $step$ é o número de iterações em SVS para definir $V_{code,step}^*$.

Todos os estimadores de painel (fatores comuns, melhores modelos ($\bar{v}_{code,step}$) e aprendizado de máquina) usaram como base $V_{code,step}^*$ pois nenhum deles poderia ser usado aplicado corretamente à base

¹⁶ Falamos em “praticamente” porque os modelos ARs não são estimadores de painel e usam base de dados de preditores diferente do usado para os outros casos.

inicial V de mais de 600.000 preditores¹⁷. Só isso já é uma vantagem do nosso estimador e representa um teste rigoroso para balizamento do erro out of sample. Isso porque, ao utilizar esses estimadores em $V_{code,step}^*$ e não V estamos comparando esses apenas ao segundo estágio do procedimento proposto, o que sobre estima a eficiência preditiva destes.

Os erros dos modelos ARs foram calculados com a raiz da média dos erros de cada país gerados por cada AR .

As tabelas (16), (17) e (18) mostram os resultados dos RMSE para os painéis ótimos 1, 2 e 3.

Nas tabelas mostramos os resultados para uma base V com tratamento para tendência e sem tratamento. Com “com tratamento para tendência” queremos dizer que retiramos a tendência e a sazonalidade. Com “sem tratamento” queremos dizer que mantivemos as variáveis tais como observados na natureza. Essa divisão foi uma preocupação do trabalho porque os modelos ARs tinham um poder preditivo sobre avaliado na presença de tendência e sazonalidade. Assim, uma comparação mais justa deveria ser feita em um ambiente sem tendência e sazonalidade¹⁸.

Nossos resultados indicam que o procedimento proposto no trabalho foi bem sucedido pois tem um resultado superior às nossas referências (menor erro de previsão com dados novos), inclusive os ARs .

Com isso queremos dizer que nossa escolha da forma funcional de $\bar{v}_{code,step}$, para a configuração utilizada, é uma decisão muito boa dentro do espaço de possibilidades das decisões possíveis.

Na verdade, de acordo com os exercícios efetuados não apresentados nessa tese, utilizando as lógicas por trás dos códigos, também conseguiríamos tomar decisões muito boas em outras configurações.

4.4.2 Política fiscal e mercado financeiro

Tendo em mente que nosso processo de tomada de decisão é eficiente, vamos argumentar com base nos $\bar{v}_{code,step}$ sobre as hipóteses levantadas no trabalho. Questionamos se o que aprendemos no turbulento início do século XXI pode ser relevante para antecipar flutuações do produto.

Assim, queremos saber se os controles para a política fiscal e o mercado financeiro estão entre os preditores integrantes dos melhores modelos ($\bar{v}_{code,step}$) para cada painel. Como o processo final foi bem sucedido, tal como discutido no tópico anterior sobre eficiência preditiva, então vamos assumir a análise do segundo estágio como suficiente.

Resumimos na tabela (19) a frequência dos preditores em cada $\bar{v}_{code,step}$ para os três painéis nas bases com tratamento de tendência (simples) e sem o tratamento (nenhum). Em cada linha identificamos um preditor com frequência relevante em todos os modelos. Os preditores que pertencem ao mesmo grupo de informação foram reunidos sob a denominação $[v_t](.)$. $[v_t](.)$ indica que o preditor $[v_t]$ (coluna “código” das tabelas (13), (14) e (15)) apareceu em $\bar{v}_{code,step}$ ou sozinho ou interagindo (coeficiente variável) no numerador¹⁹.

Nossos resultados indicam uma concentração de preditores, em termos conceituais, nos diferentes modelos escolhidos ($\bar{v}_{code,step}$) para os diferentes painéis utilizados

¹⁷ Supomos que o erro out of sample do estimador em $V_{code,step}^*$ é menor ou igual ao erro do mesmo estimador em V . Isso porque $V_{code,step}^*$ é formado por variáveis signals e V contém ruído.

¹⁸ Na verdade, uma comparação com modelos ARs não é justa em nenhuma situação (i) por não ser um estimador de painel com efeito médio e (ii) por não usar a mesma base de preditores.

¹⁹ Aqui não vamos mostrar os resultados para as variáveis que aparecem no denominador das interações, quando for o caso, porque os resultados foram muito dispersos. Assim, entendemos que as variáveis que apareceram no denominador, o fizeram apenas para ajustar o coeficiente da variável no numerador.

Tabela 16 – RMSE out of sample - painel 1

Modelo	Tratamento da Tendência		
	(a) simples	(b) sem tratamento	
Equações individuais			
$AR(1)$	1.161	0.949	
$AR(2)$	1.178	0.959	
$AR(3)$	1.184	0.925	
$AR(4)$	1.151	0.909	
Fatores comuns em $V_{code,step}^*$			
$PCA_{mgo,3}$	5	1.217	1.017
	10	1.219	1.039
	15	1.231	1.038
	20	1.234	1.053
	25	1.237	1.055
$PCA_{lasso,3}$	5	1.217	
	10	1.228	
	15	1.227	
	20	1.235	
	25	1.229	
$PCA_{tree,3}$	5	1.229	
	10	1.240	
	15	1.239	
	20	1.248	
	25	1.254	
$PCA_{join,3}$	5	1.266	
	10	1.240	
	15	1.236	
	20	1.264	
	25	1.256	
Melhores modelos ($\bar{v}_{code,step}$)			
$\bar{v}_{mgo,3}$	0.951	0.808	
$\bar{v}_{lasso,3}$	0.986		
$\bar{v}_{tree,3}$	0.969		
$\bar{v}_{join,3}$	0.977		
Aprendizado de Máquina			
$SVM_{mgo,3}$	1.059	0.833	
$RandomForest_{mgo,3}$	1.051	0.877	
$DeepLearnig_{mgo,3}$	1.111	0.885	

NOTA: A tabela mostra a raiz do erro *out of sample* (equação (4.6)) para modelos de equações individuais, fatores comuns e modelos escolhidos pelos algoritmo (6). Para os modelos $AR(p)$ o número mostra a raiz da média dos erros individuais. A segunda coluna para os modelos com fatores comuns mostra o número de fatores comuns (5, 10, 15, 20, 25). O RMSE é mostrado em 3 colunas para três tratamentos diferentes de tendência nas variáveis da base de dados. Vamos usar $V_{code,step}^*$ para $steps = 3$ e $code = \{mgo, lasso, tree, join\}$ onde *mgo*, *lasso* e *tree* são os signals identificados por *mgo*, *lasso* e *regressão em árvore*, respectivamente, e *join* são os signals identificados pelos três métodos (MQO, lasso e regressão em árvore).

Tabela 17 – RMSE out of sample - painel 2

Modelo	Tratamento da Tendência		
	(a) simples	(b) sem tratamento	
Equações individuais			
$AR(1)$	1.188	0.946	
$AR(2)$	1.199	0.951	
$AR(3)$	1.202	0.917	
$AR(4)$	1.176	0.890	
Fatores comuns em $V_{code,step}^*$			
$PCA_{mqo,3}$	5	1.265	1.076
	10	1.267	1.086
	15		1.092
	20		1.096
	25		1.096
$PCA_{lasso,3}$	5	1.265	
	10	1.264	
	15		
	20		
	25		
$PCA_{tree,3}$	5	1.311	
	10	1.306	
	15		
	20		
	25		
$PCA_{join,3}$	5	1.303	
	10	1.309	
	15		
	20		
	25		
Melhores modelos ($\bar{v}_{code,step}$)			
$\bar{v}_{mqo,3}$	1.002	0.845	
$\bar{v}_{lasso,3}$	1.007		
$\bar{v}_{tree,3}$	1.128		
$\bar{v}_{join,3}$	0.911		
Aprendizado de Máquina			
$SVM_{mqo,3}$	1.034	0.860	
$RandomForest_{mqo,3}$	1.069	0.899	
$DeepLearnig_{mqo,3}$	1.111	0.948	

NOTA: A tabela mostra a raiz do erro *out of sample* (equação (4.6)) para modelos de equações individuais, fatores comuns e modelos escolhidos pelos algoritmo (6). Para os modelos $AR(p)$ o número mostra a raiz da média dos erros individuais. A segunda coluna para os modelos com fatores comuns mostra o número de fatores comuns (5, 10, 15, 20, 25). O RMSE é mostrado em 3 colunas para três tratamentos diferentes de tendência nas variáveis da base de dados. Vamos usar $V_{code,step}^*$ para $steps = 3$ e $code = \{mqo, lasso, tree, join\}$ onde *mqo*, *lasso* e *tree* são os signals identificados por *mqo*, *lasso* e *regressão em árvore*, respectivamente, e *join* são os signals identificados pelos três métodos (MQO, lasso e regressão em árvore).

Tabela 18 – RMSE out of sample - painel 3

Modelo	Tratamento da Tendência		
	(a) simples	(b) sem tratamento	
Equações individuais			
$AR(1)$	1.149	0.944	
$AR(2)$	1.166	0.956	
$AR(3)$	1.171	0.923	
$AR(4)$	1.141	0.911	
Fatores comuns em $V_{code,step}^*$			
$PCA_{mqo,3}$	5	1.208	1.019
	10	1.211	1.029
	15	1.214	1.052
	20	1.245	1.052
	25	1.239	1.062
$PCA_{lasso,3}$	5	1.199	
	10	1.215	
	15	1.246	
	20	1.239	
	25	1.236	
$PCA_{tree,3}$	5	1.195	
	10	1.196	
	15	1.212	
	20	1.240	
	25	1.254	
$PCA_{join,3}$	5	1.195	
	10	1.198	
	15	1.191	
	20	1.217	
	25	1.237	
Melhores modelos ($\bar{v}_{code,step}$)			
$\bar{v}_{mqo,3}$	0.954	0.821	
$\bar{v}_{lasso,3}$	0.945		
$\bar{v}_{tree,3}$	0.983		
$\bar{v}_{join,3}$	0.983		
Aprendizado de Máquina			
$ SVM_{mqo,3}$	1.014	0.843	
$ RandomForest_{mqo,3}$	1.038	0.868	
$ DeepLearnig_{mqo,3}$	1.077	0.897	

NOTA: A tabela mostra a raiz do erro *out of sample* (equação (4.6)) para modelos de equações individuais, fatores comuns e modelos escolhidos pelos algoritmo (6). Para os modelos $AR(p)$ o número mostra a raiz da média dos erros individuais. A segunda coluna para os modelos com fatores comuns mostra o número de fatores comuns (5, 10, 15, 20, 25). O RMSE é mostrado em 3 colunas para três tratamentos diferentes de tendência nas variáveis da base de dados. Vamos usar $V_{code,step}^*$ para $steps = 3$ e $code = \{mqo, lasso, tree, join\}$ onde *mqo*, *lasso* e *tree* são os signals identificados por *mqo*, *lasso* e *regressão em árvore*, respectivamente, e *join* são os signals identificados pelos três métodos (MQO, lasso e regressão em árvore).

Enquanto que tanto a produtividade, como o investimento e os gastos do governo apareceram em nossas estimações finais ($\bar{v}_{code,step}$) apenas de forma marginal, os controles para interação do país com o mercado internacional representaram a quase totalidade dos preditores desses modelos escolhidos.

A tabela (19) mostra essa concentração para a interação com o mercado internacional ($nx(.)$, $x(.)$ e $m(.)$). Em específico, os relacionados à exportação ($nx(.)$ e $x(.)$) foram os mais presentes nos $\bar{v}_{code,step}$. Isso quer dizer que várias dimensões da interação do país com o mercado internacional foram apontadas como dentre as mais relevantes para determinar as variações futuras do produto dos países. Em específico, os relacionados à exportação ($nx(.)$ e $x(.)$) foram os mais frequentes. No entanto, se olharmos mais atentamente, vamos ver que os que controlavam para a abertura comercial (relaciona o indicador com o produto total - $nxY(.)$, $nxYD(.)$, $xY(.)$ e $xYD(.)$) foram o grupo mais frequente.

No que se refere ao mercado financeiro, verificamos que a sensibilidade do comércio internacional aos juros internos, ou spreads de juros ($[v_t]Br(.)$) foi a o preditor desse grupo mais presente nos $\bar{v}_{code,step}$. Mas como essa sensibilidade esteve restrita a um grupo dominador e foi tão frequente quanto a sensibilidade ao câmbio e spreads de câmbio, optamos por não tirar conclusões.

Falamos de forma genérica em vários aspectos, como $[v_t](.)$ ou interação com o mercado internacional ou até as interações/componentes, porque mais importante que a variável em si, tal como está mensurada, é o que ela representa em termos de idéia.

Assim, no que se refere às hipóteses que nos propomos a testar, não podemos dizer que a política fiscal, ou o mercado financeiro ou a produtividade da mão de obra estiveram entre os melhores indicadores para antecipar flutuações do produto dos países. Isso porque o efeito da interação com o mercado internacional foi relativamente muito superior ao dos demais grupos avaliados.

No entanto, é importante argumentar aqui que para essa conclusão seria interessante levar em conta o resultado do primeiro estágio ($V_{code,step}^*$). Enquanto o resultado do segundo estágio é mais restritivo, do ponto de vista de resumir e concluir sobre V , o resultado do primeiro estágio é mais genérico e fornece uma visão menos viesada do que é importante (signal).

Apesar disso, optamos por concluir apenas em cima do segundo estágio porque, com base em exercícios anteriores, os resultados seriam praticamente os mesmos e a análise dos resultados do primeiro estágio seria mais complexa.

4.5 Conclusão

Os ciclos da economia mundial no início do século XXI, tanto em termos de crescimento, como turbulência e recuperação, junto com a significativa disponibilidade de dados econômicos e financeiros nos fornecem um cenário extremamente interessante para testar nosso estimador proposto no trabalho. Isso porque, além de um ambiente interessante em termos de informação e disponibilidade de dados, a principal vantagem do processo é considerar e resumir um volume grande de informação.

Nosso procedimento é baseado no estimador de Meinshausen [2007] no qual estendemos os conceitos dos dois estágios do estimador em questão para aprimorar e flexibilizar a definição de variável signal do primeiro estágio, otimizar para poder aplicar a grandes bases de preditores e melhor interagir o conceito de signal do primeiro estágio com a relevância do problema final do segundo estágio.

Tabela 19 – Variáveis nos modelos $\bar{v}_{code,step}$ por painel

Painel	Painel 1				Painel 2				Painel 3				Total			
	trat. tend. variáveis	simples		sem	trat. tend. variáveis	simples		nenhum	trat. tend. variáveis	simples		sem				
	mqo	lasso	tree	join	mqo	mqo	lasso	tree	join	mqo	mqo	lasso	tree	join	mqo	
$a_e(.)$					1								1			2
$c_g(.)$			1													1
$fsd(.)$									1							2
$i_k(.)$	1		1								1	1				4
$m_g(.)$						1								1		2
$m_s(.)$			1		1	1		1	1	1						5
$nxBe(.)$						3	2		1	2						8
$nxBr(.)$		1	1	1	1	1	1			4	1	2	1	2	1	17
$nxBse(.)$						2	1	1		2						6
$nxBsrus(.)$						3	3		1	1						8
$nxYD(.)$					1	1	1		8	2		2			1	16
$nxY(.)$	3	3	2	3	2	2	6	2	5	1	3	2	3	2	5	44
$x_gBr(.)$		2									1			1	3	7
$x_gYD(.)$	5	1		1	2			1			2			1	1	14
$x_gY(.)$	1	1			1						2	2	1	1		9
$x_sBr(.)$	1	2	1		1					1				1	1	8
$x_sDY(.)$	4	2	1	2	4	1				1	2	3	1	2	1	24
$x_sY(.)$	5	3	6	1	2	3	1	0	1	1	6	2	4	4	1	40
destacadas	20	15	14	7	16	18	15	5	18	16	18	15	11	15	14	
total	20	16	14	8	17	20	17	5	18	17	20	17	11	15	14	

NOTA: A tabela mostra os preditores em cada modelo escolhido ($\bar{v}_{code,step}$) nas tabelas (16), (17) e (18). Agrupamos as variáveis com a mesma informação no numerador para simplificar a apresentação. Dessa forma $[v_t](.)$ indica que a variável $[v_t]$ apareceu no modelo ou isoladamente, ou como componente, ou no numerador de uma interação com outra variável. A última coluna (“total”) mostra o somatório de ocorrências de cada variável. A linha “trat tend” indica o tratamento aplicado a tendência das séries (“simples” retiramos a tendência linear e o efeito sazonalidade dos trimestres e “sem” indica que as séries não receberam tratamento para tirar tendência). A linha “destacadas” mostra o somatório dos preditores do modelo destacados na tabela. A última linha (“total”) mostra o total de preditores em cada modelo ($\bar{v}_{code,step}$).

Aplicamos nosso processo em equações para antecipar flutuações do produto dos países. Gostaríamos de saber se o que aprendemos com os dados no período, a que nos referimos, de muita informação é relevante para prever esses futuros movimentos do produto e quais lições podemos tirar de nossas estimativas. Em específico, destacamos a política fiscal, o mercado financeiro e a produtividade da mão de obra como importantes forças motoras do produto dos países e que tiveram um comportamento diferenciado do observado até então.

Para testar a eficiência do nosso processo proposto, usamos como balizadores modelos em painel que consideram informação comum no conjunto de variáveis signals ($V_{code,step}^*$) do primeiro estágio, estimadores em evidência na literatura de aprendizado de máquina e modelos individuais ARs .

Cada um desses grupos de estimadores representa desafios como balizadores das estimativas por diferentes razões. Em específico os ARs são, no nosso entender, os estimadores mais sobre-estimados por não serem modelos em painel (consideram efeitos individuais ao invés de efeitos médios dos painéis). Para aumentar o desafio e até por limitação operacional dos nossos balizadores, aplicamos os métodos de painel sobre a base de variáveis signals do primeiro estágio do processo e não sobre a base original que seria o mais justo.

Nossas estimativas mostram que nosso estimador possui um desempenho preditivo com dados novos superior a todos os balizadores utilizados. Isso quer dizer que conseguimos tomar decisões muito boas, em termos de especificação, no espaço de possibilidades de decisões.

Quando analisamos os preditores que fazem parte dos modelos selecionados, verificamos que o aspecto mais relevante para prever as variações futuras do produto são relacionadas à interação com o mercado internacional. Em específico o volume de exportações (bens e serviços) relativo ao tamanho do país.

Como estamos usando preditores que medem o desvio em torno da média do período analisado, queremos dizer que a melhor fonte de informação para saber se o país vai crescer ou não no próximo período é se ele está absorvendo mais renda internacional ou não do que fazia no passado. Chegamos a essa conclusão analisando muitas dimensões dos componentes de cada variável com o objetivo de controlar ou reduzir o problema de erro de especificação.

Por fim, vale mencionar que ao final do trabalho optamos por simplificar as análises pois, por diversos caminhos, chegamos aos mesmos resultados e conclusões. Primeiro que com a lógica do processo conseguimos tomar uma decisão muito boa dentro do espaço de possibilidades. Segundo que os preditores mais relevantes sempre são os relacionados à interação com o mercado internacional.

Dessa forma, é importante destacar que uma análise mais prudente seria feita sobre os resultados do primeiro estágio e não do segundo estágio porque o primeiro fornece uma visão mais ampla dos preditores relevantes enquanto que o segundo mostra resultados mais restritivos.

Isso é importante de ser considerado em função das características dos preditores utilizados no trabalho. Trabalhamos com interações, relações inversas e decomposição em componentes com o objetivo de considerar várias dimensões de cada informação. Com isso, a visão do *SVS* sobre o grupo de variáveis signals é mais informativa e menos restritiva que a do *CVC*. Com o *SVS* conseguimos tirar conclusões sobre o conjunto de informação signal menos viesadas.

Referências

- Daron Acemoglu, Simon Johnson, and James A. Robinson. Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly Journal of Economics*, 117(4):1231–1294, November 2002.
- Mark Aguiar and Gita Gopinath. Emerging market business cycles: The cycle is the trend. *Journal of Political Economy*, 2007.
- Joshua Aizenman, Yothin Jinjarak, and Donghyun Park. Capital flows and economic growth in the era of financial integration and crisis, 1990–2010. *Open Economies Review*, 2013.
- Enrique Alberola, Aitor Erce, and Jose Maria Serena. International reserves and gross capital flows dynamics. *Journal of International Money and Finance*, 60(C):151–171, 2016.
- Patrick Bajari, Denis Nekipelov, Stephen P. Ryan, and Miaoyu Yang. Machine learning methods for demand estimation. *American Economic Review*, 105(5):481–85, May 2015a.
- Patrick Bajari, Denis Nekipelov, Stephen P. Ryan, and Miaoyu Yang. Demand estimation with machine learning and model combination. NBER Working Papers 20955, National Bureau of Economic Research, Inc, February 2015b.
- Robert J. Barro. Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 1991.
- Robert J. Barro. *Determinants of Economic Growth: A Cross-Country Empirical Study*, volume 1 of *MIT Press Books*. The MIT Press, June 1998.
- Tamim Bayoumi, Hamid Faruquee, Douglas Laxton, Philippe D Karam, Alessandro Rebucci, Jaewoo Lee, Benjamin Hunt, and Ivan Tchakarov. Gem: A new international macroeconomic model. Technical Report 239, International Monetary Fund, 2004.
- Andrew Berg, Jonathan D. Ostry, and Jeromin Zettelmeyer. What makes growth sustained? *Journal of Development Economics*, 98(2):149–166, 2012.
- Olivier Blanchard, Giovanni Dell’Ariccia, and Paolo Mauro. Rethinking macroeconomic policy. *Journal of Money, Credit and Banking*, 42(s1):199–215, 09 2010.
- Olivier J Blanchard, Giovanni Dell’Ariccia, and Paolo Mauro. Rethinking Macro Policy II; Getting Granular. IMF Staff Discussion Notes 13/003, International Monetary Fund, April 2013.
- Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. *Machine Learning Strategies for Time Series Forecasting*. Springer Berlin Heidelberg, 2013.
- Barry P. Bosworth and Susan M. Collins. The empirics of growth: An update. *Brookings Papers on Economic Activity*, 34(2):113–206, 2003.

- Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001. ISSN 0885-6125.
- Joseph P. Byrne and Norbert Fiess. International capital flows to emerging markets: National and global determinants. *Journal of International Money and Finance*, 61(C), 2016.
- Ricardo J. Caballero, Emmanuel Farhi, and Pierre-Olivier Gourinchas. An equilibrium model of “global imbalances” and low interest rates. *American Economic Review*, 98(1):358–93, March 2008.
- Ryan Chahrour, Stephanie Schmitt-Grohe, and Martin Uribe. A model-based evaluation of the debate on the size of the tax multiplier. *American Economic Journal: Economic Policy*, 4(2):28–45, May 2012.
- Roberto Chang and Andres Fernandez. On the sources of aggregate fluctuations in emerging economies. *International Economic Review*, 2013.
- Punam Chuhan, Stijn Claessens, and Nlandu Mamingi. Equity and bond flows to latin america and asia: the role of global and country factors. *Journal of Development Economics*, 55(2):439 – 463, 1998.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- William Easterly. National policies and economic growth: A reappraisal. In Philippe Aghion and Steven Durlauf, editors, *Handbook of Economic Growth*, volume 1 of *Handbook of Economic Growth*, chapter 15, pages 1015–1059. Elsevier, 2005.
- William Easterly, Michael Kremer, Lant Pritchett, and Lawrence H. Summers. Good policy or good luck?: Country growth performance and temporary shocks. *Journal of Monetary Economics*, 32(3): 459–483, December 1993.
- Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *Annals of Statistics*, 2004.
- Ayse Y. Evrensel. Corruption, growth, and growth volatility. *International Review of Economics & Finance*, 19(3):501–514, June 2010.
- Ana M. Fernandes. Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries. *Journal of International Economics*, 71(1):52–71, March 2007.
- Eduardo Fernandez-Arias. The new wave of private capital inflows: Push or pull? *Journal of Development Economics*, 48(2):389–418, March 1996.
- Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, 2014.
- Jerome Friedman, Trevor Hastie, Holger Häfling, and Robert Tibshirani. Pathwise coordinate optimization. Technical report, Annals of Applied Statistics, 2007.
- J. Friedmand and B. Popescu. Predictive learning via rule ensembles. *Annals of Statistics*, 2008.

- W. J. Fu. Penalized regressions: the bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7(3):397–416, 1998.
- Javier Garcia-Cicco, Roberto Pancrazi, and Martin Uribe. Real business cycles in emerging countries? *American Economic Review*, 100(5):2510–31, 2010.
- Isabelle Guyon, Amir Saffari, Gideon Dror, and Gavin Cawley. Model selection: Beyond the bayesian/frequentist divide. *J. Mach. Learn. Res.*, 11:61–87, March 2010. ISSN 1532-4435.
- Robert Hable and Andreas Christmann. On qualitative robustness of support vector machines. *Journal of Multivariate Analysis*, 102(6):993–1007, July 2011.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2009.
- Ricardo Hausmann, Lant Pritchett, and Dani Rodrik. Growth accelerations. *Journal of Economic Growth*, 10(4):303–329, December 2005.
- Donald A Hay. The Post-1990 Brazilian Trade Liberalisation and the Performance of Large Manufacturing Firms: Productivity, Market Share and Profits. *Economic Journal*, 111(473):620–41, July 2001.
- Torsten Hothorn, Kurt Hornik, and Achim Zeileis. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674, 2006.
- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics, 2013.
- Amir E. Khandani, Adlar J. Kim, and Andrew W. Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11):2767–2787, November 2010.
- Yoonbai Kim. Causes of capital flows in developing countries. *Journal of International Money and Finance*, 19(2):235–253, April 2000.
- Pravin Krishna and Devashish Mitra. Trade liberalization, market discipline and productivity growth: new evidence from India. *Journal of Development Economics*, 56(2):447–462, August 1998.
- Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2013.
- Douglas Laxton, Peter Isard, Hamid Faruqee, Eswar Prasad, and Bart Turtelboom. Multimod mark iii - the core dynamic and steady-state models. *IMF Occasional Paper 164*, 1998.
- Ross Levine and David Renelt. A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 1992.
- Shu-Chin Lin and Dong-Hyeon Kim. The link between economic growth and growth volatility. *Empirical Economics*, 46(1):43–63, February 2014.
- David L. Lindauer and Lant Pritchett. What s the big idea? the third generation of policies for economic growth. *Economia*, 2002.

- Jakob B. Madsen, Shishir Saxena, and James B. Ang. The Indian growth miracle and endogenous growth. *Journal of Development Economics*, 93(1):37–48, September 2010.
- Anna Maria Mayda and Dani Rodrik. Why are some people (and countries) more protectionist than others? *European Economic Review*, 49(6):1393–1430, August 2005.
- Nicolai Meinshausen. Lasso with relaxation. *Computational Statistics and Data Analysis*, 2007.
- Marcelo Mello and Roberto Perrelli. Growth equations: a quantile regression exploration. *The Quarterly Review of Economics and Finance*, 43(4):643–667, 2003.
- Pablo A. Neumeyer and Fabrizio Perri. Business cycles in emerging economies: the role of interest rates. *Journal of Monetary Economics*, 2005.
- NiGEM. National institute’s global econometric model. *National Institute of Economic and Social Research*, 2016.
- Tonatiuh Peña, Serafin Martínez, and Bolanle Abudu. *Bankruptcy Prediction: A Comparison of Some Statistical and Machine Learning Techniques*, pages 109–131. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- Vasilios Plakandaras, Periklis Gogas, and Theophilos Papadimitriou. Directional forecasting in financial time series using support vector machines: The usd/euro exchange rate. *Journal of Computational Optimisation on Economics and Finance*, 5(2):126–139, 2013.
- Elena Podrecca and Gaetano Carmeci. Fixed investment and economic growth: new results on causality. *Applied Economics*, 33(2):177–182, 2001.
- C. R. Rao, Y. Wu, Sadanori Konishi, and Rahul Mukerjee. On model selection. *Lecture Notes-Monograph Series*, 38:1–64, 2001. ISSN 07492170.
- Dennis Reinhardt, Luca Antonio Ricci, and Thierry Tressel. International capital flows and development: Financial openness matters. *Journal of International Economics*, 2013.
- Francisco Rodriguez and Dani Rodrik. Trade Policy and Economic Growth: A Skeptic’s Guide to the Cross-National Evidence. In *NBER Macroeconomics Annual 2000, Volume 15*, NBER Chapters, pages 261–338. National Bureau of Economic Research, Inc, May 2001.
- Dani Rodrik. Growth strategies. In Philippe Aghion and Steven Durlauf, editors, *Handbook of Economic Growth*, volume 1 of *Handbook of Economic Growth*, chapter 14, pages 967–1014. Elsevier, June 2005.
- Dani Rodrik. Making room for china in the world economy. *American Economic Review*, 100(2):89–93, May 2010.
- Dani Rodrik. The future of economic convergence. NBER Working Papers 17400, National Bureau of Economic Research, Inc, September 2011.
- Paul M Romer. Increasing returns and long-run growth. *Journal of Political Economy*, 94(5):1002–37, October 1986.

- Xavier Sala-I-Martin. I just run tow million regressions. *American Economic Review*, 1997.
- Xavier Sala-I-Martin. The disturbing rise of global income inequality. Working Paper 8904, National Bureau of Economic Research, April 2002.
- Xavier Sala-I-Martin, Gernot Doppelhofer, and Ronald I. Miller. Determinants of long-term growth: A bayesian averaging of classical estimates (bace) approach. *American Economic Review*, 94(4):813–835, September 2004.
- Stephanie Schmitt-Grohe and Martin Uribe. What’s news in business cycles. *Econometrica*, 80, 2012.
- Stephanie Schmitt-Grohe and Martin Uribe. How important are terms of trade shocks? Working Paper 21253, National Bureau of Economic Research, June 2015.
- Frank Smets and Rafael Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 2007.
- Holger Strulik, Klaus Prettnner, and Alexia Prskawetz. The past and future of knowledge-based growth. *Journal of Economic Growth*, 18(4):411–437, December 2013.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58, 1996.
- Cedric Tille and Eric van Wincoop. International capital flows. *Journal of International Economics*, 80 (2):157–175, March 2010.
- James Tybout, Jamie de Melo, and Vittorio Corbo. The effects of trade reforms on scale and technical efficiency : New evidence from Chile. *Journal of International Economics*, 31(3-4):231–250, November 1991.
- James R. Tybout and M. Daniel Westbrook. Trade liberalization and the dimensions of efficiency change in Mexican manufacturing industries. *Journal of International Economics*, 39(1-2):53–78, August 1995.
- Athanasios Vamvakidis. How Robust Is the Growth-Openness Connection? Historical Evidence. *Journal of Economic Growth*, 7(1):57–80, March 2002.
- Hal R. Varian. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28, Spring 2014.
- Jurgen von Hagen and Haiping Zhang. Financial development, international capital flows, and aggregate output. *Journal of Development Economics*, 2014.
- William. Life during growth. *Journal of Economic Growth*, 4(3):239–76, September 1999.
- Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*. The Morgan Kaufmann Series in Data Management Systems, 2011.
- Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.

5 Apêndice - Dados

Para montar os painéis de países para usar nas análises do capítulo 3 e capítulo 4, a tese usou dados do Banco Mundial, BIS, a OCDE, Bloomberg e a Chicago Board Options Exchange.

A tabela (20) descreve os dados base utilizados para construção da base de dados do capítulo 3.

As variáveis investimento (i_a) e poupança externa (nx_a) foram calculados de acordo com a equação (5.1) e equação (5.2) respectivamente.

$$i_a = dnb_a - c_a \quad (5.1)$$

$$nx_a = x_a - m_a \quad (5.2)$$

A variável de crédito privado ($cred_a$) foi calculado de acordo com a equação (5.3).

$$cred_a = (cred.c_a - cred.g_a) * y_a \quad (5.3)$$

Por fim as variáveis econômicas (y_a , i_a , nx_a e $cred_a$) foram calculadas per capita de acordo com a equação (5.4) onde $[v]$ representa pode ser $\{y_a, i_a, nx_a, cred_a\}$.

$$[v]_{capta} = (\ln [v] - \ln(pop_{14.64})) * 100 \quad (5.4)$$

No caso do capítulo 4, utilizamos bases diferentes tanto para o primeiro estágio como para o segundo estágio tal como descrito no tópico 4.3.4.

Para o primeiro estágio, tal como descrito na tabela (12), temos muitas variáveis e por isso não iremos discriminá-las individualmente. Aqui vamos indicar as bases de dados carregadas para formar a base de dados da tabela (12). Os arquivos carregados são as bases de dados trimestrais (i) Joint External Debt Hub (BIS), (ii) Public Sector Debt Data (Banco Mundial), (iii) Quarterly External Debt Statistics GDDS (Banco Mundial/FMI), (iv) Special Data Dissemination Standard SDDS (Banco Mundial/FMI), (v) Quarterly Sector Accounts QASA/QASA_TABLE810 (OCDE), (vi) Monthly Economic Indicator

Tabela 20 – Dados Capítulo - 3

Variável	Código da variável	Código na fonte de dados	Capítulo	Periodicidade	Fonte
produto	y_a	NY.GDP.MKTP.CD	3	anual	Banco Mundial
despesa nacional bruta	dnb_a	NE.DAB.TOTL.CD	3	anual	Banco Mundial
exportações	x_a	NE.EXP.GNFS.CD	3	anual	Banco Mundial
importações	m_a	NE.IMP.GNFS.CD	3	anual	Banco Mundial
consumo total (privado mais governo)	c_a	NE.CON.TETC.CD	3	anual	Banco Mundial
investimento direto estrangeiro	idi_a	BX.KLT.DINV.CD.WD	3	anual	Banco Mundial
crédito doméstico (% do pib)	$cred.c_a$	FS.AST.DOMS.GD.ZS	3	anual	Banco Mundial
crédito para o governo (% do pib)	$cred.g_a$	GFDD.EI.08	3	anual	Banco Mundial
taxa de câmbio	e_a	PA.NUS.FCRF	3	anual	Banco Mundial
taxa de juros	r_a	FR.INR.LEND	3	anual	Banco Mundial
inflação (deflator anual)	$defl_a$	NY.GDP.DEFL.KD.ZG	3	anual	Banco Mundial
população 14 - 64	$pop_{14.64}$	SP.POP.1564.TO	3	anual	Banco Mundial
VIX	VIX_a		3	diário	Chicago Board Exchange

NOTA: A tabela mostra os dados usados para construção da base do capítulo 3. A coluna “código da variável” mostra a abreviação da variável e a coluna “código na fonte de dados” mostra o código da variável na base de dados da instituição de origem.

Tabela 21 – Variáveis Base - nível e índices

Número	Nome	Código	Fonte	Código Original
Contas Nacionais - fluxos internacionais (List NX)				
1	exportações líquidas	<i>nx</i>	OCDE	B6BLTT01
2	importações de serviços	<i>m_s</i>	OCDE	B6DBSE01
3	importações de bens	<i>m_g</i>	OCDE	B6DBTD01
4	exportações de serviços	<i>x_s</i>	OCDE	B6CRSE01
5	exportações de bens	<i>x_g</i>	OCDE	B6CRTD01
6	produto	<i>pib</i>	OCDE	B1_GI
7	reservas	<i>res</i>	OCDE	B6FARA01
8	conta derivativos fluxo de capitais	<i>der</i>	OCDE	B6FAFD01
9	. curto prazo	<i>fcp</i>	OCDE	B6FAPI10
10	. longo prazo	<i>flp</i>	OCDE	B6FADI01
11	. saldo	<i>fsd</i>	OCDE	B6FATT01
Governo - política fiscal (List G)				
impostos				
12	. bens	<i>t_g</i>	OCDE	D21REC
13	. capital	<i>t_k</i>	OCDE	D91REC
14	. valor agregado	<i>t_i</i>	OCDE	D211REC
15	. produção	<i>t_y</i>	OCDE	D29REC
16	. renda	<i>t_w</i>	OCDE	D5REC
17	gastos do governo	<i>g_c</i>	OCDE	P3S13
Mercado de Trabalho e Consumidor (List L)				
emprego total				
		<i>emprego.total</i>	OCDE	LFEM64TT
custo da mão de obra				
		<i>wage</i>	OCDE	ULQEUL01
população total				
		<i>pop.total</i>	OCDE	LFPOTTTT
Outros (List_O)				
30	consumo privado	<i>c_c</i>	OCDE	P31S14_S15
31	formação bruta de capital fixo	<i>i_i</i>	OCDE	P51
Outras variáveis				
	Deflator	<i>defl</i>	OCDE	B1_GE
	População Economicamente Ativa	<i>pop_{14-64,t}</i>	OCDE	PEANC

NOTA: A tabela lista as variáveis de referência do trabalho que não fazem parte do mercado financeiro. Essa é uma base de referência porque essas variáveis serão usadas para gerar a base final de preditores através de transformações aplicadas nessas variáveis de referência. Estas variáveis são divididas em 4 grupos que controlam para fatores importantes para variações do produto no curto prazo para o período considerado. O campo **código** mostra o nome da variável nas estimações e campos **fonte** e **medida** são autoexplicativos. (*) A produtividade ($a_{e,t}$) é estimada usando dados trimestrais da OCDE na equação $\ln\left(\frac{pibt}{pop_{14-64,t}}\right) = \beta_i \ln\left(\frac{i_{i,t}}{pop_{14-64,t}}\right) + \beta_l \ln\left(\frac{emprego}{pop_{14-64,t}}\right) + a_{e,t}$.

MEI (OCDE), (vii) Quarterly National Accounts QNA (OCDE) e (viii) Unit labour costs and labour productivity ULC_EEQ (OCDE). O arquivo total tem 2.976.327 linhas.

No segundo estágio contruímos uma base de dados, a partir dos resultados do primeiro estágio, escolhendo algumas variáveis. Note que no primeiro estágio a mesma informação poderia ser medida de diferentes formas e os resultados desse “datamining” permitiu a escolha das melhores variáveis (identificadas pelos códigos na base/instituição fonte), assim com a disponibilidade dessas variáveis, para os painéis que pretendíamos montar. A tabela (21) mostra o código na base/instituição fonte das variáveis do conjunto de referência da tabela (13).

Todas as variáveis estão medidas em USD, sendo que, quando a variável original estava em moeda doméstica, convertemos pelo câmbio médio do trimestre.

As variáveis da tabela (13) e da tabela (14) que tiveram como fonte a Bloomberg não discriminamos os códigos da base/instituição fonte porque essas variáveis foram coletadas uma a uma e país a país. Com isso, para uma mesma variável, os códigos eram diferentes para cada país.