

RAE-CEA

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Juntos sobreviveremos, divididos pereceremos.”

Quais países mais entram em coalizões no GATT/OMC?

Profª Silvia L. P. Ferrari

Athos P. Damiani

Julio A. Z. Trecenti

São Paulo, maio de 2012

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP**RELATÓRIO DE ANÁLISE ESTATÍSTICA**

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO: “Juntos sobreviveremos, divididos pereceremos’. Quais países mais entram em coalizões no GATT/OMC?”.

PESQUISADORES: Dr. Gabriel Cepaluni.

INSTITUIÇÃO: Universidade Estadual Paulista - Franca

FINALIDADE DO PROJETO: Publicação de artigo.

RESPONSÁVEIS PELA ANÁLISE: Prof^a Silvia Lopes de Paula Ferrari

Athos Petri Damiani

Julio Adolfo Zucon Trecenti

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

- HOLE, A. R. (2006) Calculating Murphy-Topel variance estimates in Stata: A simplified procedure. **The Stata Journal**, number 4, pp. 521–529.
- BOLFARINE, H. e SANDOVAL, M. C. (2010) **Introdução à Inferência Estatística**. Sociedade Brasileira da Matemática, Rio de Janeiro.
- BUSSAB, W. O. e MORETTIN, P. A. (2010). **Estatística Básica**. 6ªed. São Paulo: Saraiva, 540p.
- CHIPMAN, A. H., GEORGE, E. I. e McCULLOCH, R. E., (2010). BART: Bayesian Additive Regression Trees. **The Annals of Applied Statistics**, Vol.4, Nº1, 266-298.
- DAWID, A. P. (2007). **Fundamentals of Statistical Causality**. Versão de 17 de setembro de 2007. Universidade de Sheffield. Disponível em <http://www.ucl.ac.uk/statistics/research/pdfs/rr279.pdf>.
- DEGROOT, M. H., (2004) **Optimal Statistical Decisions**. Wiley Classics Library.
- HASTIE, T. J. e TIBSHIRANI, R. J. (1990). **Generalized Additive Models**. Chapman & Hall/CRC, London.
- HILBE, J. M. e HARDIN, J. W. (2007). **Generalized Linear Models and Extensions**, 2a. Edição. Stata Press, College Station, Texas.
- KUTNER, M. H.; NETER, J.; NACHTSHEIM, C. J. e LI, W. (2005), **Applied Linear Statistical Models**, 5a. Edição. Editora McGraw-Hill Irwin, Illinois.
- MARRA, G. e RADICE, R. (2010) A Flexible Instrumental Approach. **Statistical Modelling**, 2011, 581-603
- McCULLAGH, P. e NELDER, J. A. (1989). **Generalized Linear Models**, 2a. Edição. Chapman & Hall, London.
- PAULA, G. A. (2010). **Modelos de Regressão com Apoio Computacional**. Disponível em <http://www.ime.usp.br/~giapaula>.
- PEARL, J. (2010). **An Introduction to Causal Inference**. Los Angeles.

- PREGIBON, D. (1981). Logistic regression diagnostics. **Annals of Statistics** 9, 705-724.
- RIGBY, R. A. e STASINOPOULOS, D. M. (2001). The GAMLSS project: a flexible approach to statistical modeling. **New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling**, ed: Klein, B. e Korsholm, L., 249-256, Odense, Denmark.
- RIGBY, R. A. e STASINOPOULOS, D. M. (2009). **A flexible regression approach using GAMLSS in R**. Disponível em <http://gamlss.org>.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Microsoft Excel *for Windows*®, versão 2007/2010;

Microsoft Word *for Windows*®, versão 2007/2010;

R *for Windows*®, versão 2.14.0/2.15.0.

TÉCNICAS ESTATÍSTICAS UTILIZADAS: (entre parênteses a classificação ISI)

Análise Descritiva Unidimensional (03:010)

Regressão Logística (07:090)

Outros – Modelos Aditivos Generalizados (07:990)

ÁREA DE APLICAÇÃO

Ciências Políticas (14:990)

Sumário

1	Introdução	8
2	Objetivos	8
3	Descrição do estudo	9
3.1	Banco de dados.....	9
3.2	Histórico de versões	10
3.3	Premissas.....	11
3.4	Restrições.....	12
3.5	Riscos.....	12
4	Descrição das variáveis	13
4.1	Características geográficas	13
4.2	Características demográficas	13
4.3	Características econômicas.....	14
4.4	Características de democracia	14
4.5	Variáveis resposta	14
5	Dificuldades e desafios	15
6	Análise Descritiva.....	16
6.1	Análise em painel	16
6.2	Medidas-resumo.....	24
7	Causalidade	31
7.1	Variáveis instrumentais	31

	6
7.2 Visualização da causalidade utilizando o GAM	33
8 Análise Inferencial.....	35
8.1 Modelo Linear Generalizado.....	36
8.2 Modelo BART	39
9 Conclusões	44
10 Comentários.....	45
A Figuras e tabelas da análise descritiva em painel.....	48
A.1 Figuras	48
A.2 Tabelas.....	67
B Figuras e tabelas da análise descritiva das medidas-resumo	71
B.1 Figuras.....	71
B.2 Tabelas.....	82
C Figuras e tabelas da análise inferencial	86
C.1 Modelo Linear Generalizado.....	86
C.2 Modelo BART	92
D Uma breve descrição do modelo BART	108
D.1 O modelo BART	108
D.2 Seleção de modelos	109

RESUMO

Este relatório propõe abordagens estatísticas para relacionar grandezas de cunho socioeconômico dos países membros da GATT/OMC com a sua propensão em estabelecer coalizões entre si. Entende-se por coalizão o ato consumado de comunicação de qualquer natureza entre dois países com o fim de obter vantagens em decisões pautadas pela OMC. Os dados nos quais o estudo se baseia possuem caráter longitudinal. Em contraste, traçar o perfil de países de mesma propensão em formar coalizões não envolve a modelagem da evolução temporal existente, o que motiva o esforço em transportar o problema longitudinal para o plano transversal. Isto foi efetuado utilizando-se medidas-resumo das quantidades envolvidas. Mais do que associação, o interesse da pesquisa é estabelecer relações de causalidade entre as quantidades, no sentido que variáveis socioeconômicas impliquem em alterações na probabilidade de coalizão. Neste contexto, técnicas de inferência causal, como a introdução de variáveis instrumentais e abordagens bayesianas para ajustes de modelos envolvendo árvores de decisão, foram aplicadas. Para construir as medidas-resumo, uma análise descritiva dos dados foi feita explorando, principalmente, o comportamento das variáveis no tempo. Na maioria dos casos, houve paralelismo dos perfis dos países, segundo diversas medidas. Variáveis instrumentais foram utilizadas para ajuste de modelos aditivos generalizados (GAM), buscando, mais do que para outras características, uma relação funcional entre o logaritmo do produto interno bruto (PIB) do país (ajustado) e a proporção de coalizões. Para a parte inferencial, modelos de soma de árvores de decisão foram ajustados utilizando-se o enfoque bayesiano. Como resultados, em ambas as abordagens obtiveram-se curvas crescentes, sugerindo que países com maior PIB têm maior propensão de entrar em coalizão. Tais resultados são concordantes entre as análises feitas para os dois critérios de coalizão distintos.

1 Introdução

O GATT (sigla inglesa para Acordo Geral sobre Tarifas e Comércio) e a WTO (sigla inglesa para Organização Mundial do Comércio) são instituições fundadas a partir da necessidade de combater a postura não cooperativa e protecionista das nações após a segunda guerra mundial, no que se refere às práticas comerciais. Inicialmente, o GATT foi proposto para ter um caráter provisório até que se fosse criada uma organização internacional devidamente formalizada. Porém, em virtude de diversas negociações fracassadas, passaram-se mais de 50 anos desde a sua criação para que o GATT se tornasse a OMC.

Muitas vezes, as principais divergências entre as nações participantes do GATT/OMC se dão pela polarização entre países em desenvolvimento e países desenvolvidos. Motivados por estes conflitos de interesses, países com objetivos comuns buscam dialogar entre si para aprimorar suas chances de resultados positivos em negociações da rodada.

Para esse agrupamento de países com objetivos comuns (que se propuseram a dialogar) se dá o nome “coalizão”. A ação de “dialogar”, neste contexto, está empregada propositalmente em seu sentido mais genérico devido ao caráter provisório previsto na própria proposta de criação do GATT (em outras palavras, o GATT nunca obteve uma personalidade jurídica própria, o que nunca possibilitou a formalização de alianças).

O estudo coloca em foco tais coalizões e o perfil dos países que tendem a fazê-las com maior frequência. Para tanto, técnicas estatísticas serão aplicadas em dados de natureza sócio-político-econômicas dos países participantes do GATT/OMC para mapear as características mais relevantes quem levam os países a entrarem em coalizões mais frequentemente.

2 Objetivos

As análises estatísticas têm por objetivos:

- Encontrar associações entre a frequência de entradas em coalizões com características de cunho econômico, político e social dos países participantes da GATT/OMC;
- Explorar a existência de causalidade entre características econômicas e políticas do país e a frequência de entrada em coalizões;
- Avaliar a robustez das associações entre dois diferentes critérios de coalizão¹.

3 Descrição do estudo

O estudo é de natureza observacional e baseia-se em informações de todos os 146 países participantes da GATT/OMC que ingressaram em 2006, ou antes, referentes ao período compreendido entre 1960 e 2006. Grande parte das informações foi retirada do compilado “The Quality of Government Dataset Codebook - 2006” (QoG), disponibilizado pela Universidade de Gothenburg.

Cada unidade amostral consiste de um país e a variável principal de interesse é o **número** de coalizões realizadas pelo país em um determinado ano.

Para estas contagens, dois critérios foram considerados, sendo eles idênticos não fosse o fato do primeiro incluir acordos regionais, como, por exemplo, o Mercosul. A verificação de robustez citada nos objetivos se dará por meio da comparação entre as análises dessas duas quantidades.

3.1 Banco de dados

¹ O informalismo da formação de uma coalizão faz com que a sua definição seja subjetiva, e sua caracterização dependa do critério utilizado para definir se ela efetivamente se consumou.

A coleta de dados foi realizada unificando-se diversas bases de dados oficiais, providas principalmente do QoG. Após a coleta e organização dos dados, foi feita uma limpeza nos anos do estudo, que passaram a ser de 1960 até 2006. Por fim, retiraram-se os anos em que os países não estavam presentes no GATT/OMC.

A variável resposta foi coletada pelo próprio pesquisador, a partir de pesquisas históricas dos países e das organizações de comércio. Apesar de essencialmente subjetiva, a variável resposta foi construída a partir de critérios bem definidos.

Vale ressaltar que, para cada ano, tem-se informação de todos os países que faziam parte da população (possivelmente com dados faltantes dentro de cada país). Logo, o banco de dados analisado não diz respeito somente a uma amostra, mas sim ao censo. Na Seção 10, uma discussão mais aprofundada sobre o assunto será realizada. No decorrer do presente relatório, eventualmente será atribuído a este censo o nome “amostra”, pelas razões apresentadas na Seção 10.

Como cada unidade amostral foi avaliada em diferentes tempos, o banco de dados apresenta uma estrutura longitudinal, e foi organizado de modo a acomodar esse esqueleto. Eventualmente, utilizar-se-á a terminologia “painel”, característica das áreas de ciências políticas e econometria, para se referir a esse tipo de estudo.

3.2 Histórico de versões

Desde o início do acompanhamento do Centro de Estatística Aplicada (CEA) pelos estatísticos responsáveis, diversas versões do banco de dados foram enviadas. Abaixo, apresenta-se uma breve descrição de cada versão.

07/03/2012 - Primeira versão (10.301 linhas, 73 colunas), chamada de “versão bruta”. Continham variáveis que não necessariamente seriam utilizadas, e continha países que não atendiam aos critérios de inclusão da amostra;

20/03/2012 - Segunda versão (4.457 linhas, 21 colunas), chamada de “versão limpa”. As observações que não faziam parte da população foram removidas, e a faixa de tempo em que seria realizado o estudo foi definida. Além disso, diversas variáveis que não eram de interesse da pesquisa foram retiradas;

23/03/2012 - Terceira versão (4.457 linhas, 27 colunas), chamada de “versão final”. Foram adicionadas no estudo variáveis regionais que ainda não estavam inclusas. Esta versão foi atribuída como final e não foi alterada posteriormente.

3.3 Premissas

Com o intuito de facilitar a compreensão e visualização do problema, algumas premissas a respeito do estudo foram elencadas previamente. Desse modo, é possível determinar o ponto de partida do projeto. Dentre as premissas, tem-se:

- As variáveis respostas consideradas de fato representam a entrada em coalizões dos países, de modo que o pesquisador possa interpretar os resultados da maneira correta²;
- Todos os países atendem aos critérios de inclusão e exclusão da amostra, a saber, países que fazem ou fizeram parte do GATT/OMC nos anos 1960 a 2006.
- Todos os dados a serem utilizados constam na versão final do banco de dados;
- Os dados obtidos na versão final do banco de dados estão corretos;

² Essa premissa é de extrema importância para a viabilidade do estudo, pois, a princípio, não existem dados a respeito do tempo de cada coalizão (que pode durar anos). Desse modo, é difícil comparar um país que entrou e se manteve em uma coalizão por muito tempo e um país que entra frequentemente em coalizões de curta duração. Essa premissa é uma forma de assumir que as perguntas do pesquisador serão respondidas coerentemente com as definições das variáveis que estão disponíveis.

3.4 Restrições

Algumas restrições capitais do estudo foram levadas em consideração para a viabilidade da análise. Dentre elas:

- As observações omissas (*missing data*) existentes não poderão ser obtidas;
- Conclusões tiradas a partir dos dados analisados dizem respeito somente ao período analisado (1960 a 2006). Os resultados só poderão ser extrapolados para períodos futuros por intermédio de uma suposição de homogeneidade dos tempos;
- O estudo não realizará nenhuma análise de previsão;
- O estudo será realizado no primeiro semestre de 2012, tendo como limites as datas de entrega do relatório parcial (04/05/2012) e final (18/06/2012);
- Os resultados que serão obtidos serão meramente estatísticos. Cabe aos estatísticos realizarem a contextualização do problema e trazer indicações, mas é papel do pesquisador tirar conclusões a respeito do que será apresentado e realizar a transição de resultados estatisticamente significantes para essencialmente significativos;

3.5 Riscos

É possível antecipar possíveis resultados insatisfatórios e manter a estabilidade na execução do projeto. Dentre os riscos, podem-se destacar:

- Eventuais erros de digitação no banco de dados;
- Tempo insuficiente para estudo dos conceitos teóricos necessários para a realização da análise;
- Resultados inconclusivos, necessitando de nova análise posterior, com inclusão de novas variáveis ou adoção de outras técnicas estatísticas;

Cada um dos riscos será monitorado no decorrer do estudo. A princípio, deseja-se evitá-los e, se isso não for possível, cada risco será mitigado para que os danos à pesquisa sejam mínimos.

4 Descrição das variáveis

A planilha de dados na qual o estudo se baseia é composta por 25 variáveis caracterizadoras dos países mais 4 variáveis respostas, medidas em cada um dos 25 anos de abrangência da pesquisa. As variáveis foram subdivididas em cinco categorias: geográficas, demográficas, econômicas, de democracia e respostas. São elas:

Nesta estrutura: Descrição da variável [código] (informações adicionais)

4.1 Características geográficas

- Latitude [latitude]
- Longitude [longitude]
- Hemisfério leste ou oeste [hem_we] (1 se oeste e 2 se leste)
- Hemisfério norte ou sul [hem_ns] (1 se norte e 2 se sul)
- Região [regiao] (1. Europa oriental e região pós União Soviética; 2. América Latina; 3. África do norte e Oriente Médio; 4. África subsaariana; 5. Europa ocidental e América do norte; 6. Ásia oriental; 7. Sudeste da Ásia; 8. Sul da Ásia; 9. Pacífico; 10. Caribe.)
- Proporção de terra dentro dos trópicos [tropicalar] (em porcentagem)
- Área do país [wdi_area] (em quilômetros quadrados)
- Ano [year] (de 1960 a 2006)

4.2 Características demográficas

- Nome do país [cname]

- População* [pwt_pop] (em milhares)
- Mortes relacionadas a batalhas [wdi_brd]
- População urbana [wdi_urban] (em porcentagem)
- Consumo de energia elétrica* [wdi_epc] (em KWh per capita)
- Linhas telefônicas* [wdi_pl] (por 100 pessoas)

4.3 Características econômicas

- Taxa de crescimento do PIB real per capita [pwt_grgdpch] (taxa a preços constantes)
- Abertura para o comércio [pwt_openk] (em porcentagem do PIB real)
- PIB per capita* [pwt_rgdpch] (a preços constantes)
- PIB real* [pwt_rgdp] (multiplicação do PIB per capita pela população do país)
- Exportação [wdi_exp] (em porcentagem do PIB real)
- Importação [wdi_imp] (em porcentagem do PIB real)
- Comércio total [wdi_ttr] (em porcentagem do PIB real; é o mesmo que a soma das importações com as exportações)
- Participação do governo no PIB [pwt_gsg] (em porcentagem)

4.4 Características de democracia

- Escore de democracia [p_polity] (varia de -10, ditadura completa a 10, democracia completa)
- Indicador de democracia [chgademo] (0 se ditadura e 1 se democracia)

4.5 Variáveis respostas

- Número de coalizões efetuadas incluindo acordos regionais [entar]
- Indicador de entrada em coalizões incluindo acordos regionais [entar01] (0 se não entrou em coalizão naquele ano e 1 caso contrário)
- Número de coalizões contando apenas aquelas que foram criadas para se negociar no GATT/OMC [ensar]

- Indicador de entrada em coalizões excluindo acordos regionais [ensar01] (0 se não entrou em coalizão naquele ano e 1 caso contrário)

* Para essas variáveis, possíveis transformações (e.g. logaritmo, escala de 100, etc.) foram utilizadas no decorrer da análise.

5 Dificuldades e desafios

Para a saúde da qualidade da análise, é importante destacar quais dificuldades foram encontradas, com o intuito de motivar possíveis soluções. Dentre elas, podem-se destacar:

- Muitos zeros nas variáveis respostas. Considerando todos os países e anos, a porcentagem de observações em que houve ao menos uma coalizão é inferior a 15%. Isso diminui a informação a respeito dos países que entram em coalizão com mais frequência;
- Visualização das respostas. Como a estrutura dos dados é em painel as variáveis respostas são contagens ou variáveis dicotômicas, a visualização dos perfis individuais de coalizões é mais nebulosa;
- Respostas binárias ou de contagem aumentam a complexidade da modelagem dos dados, sendo necessário adotar metodologias menos triviais. A estrutura longitudinal é mais um complicador para a análise;
- Variáveis explicativas dependentes do tempo. Algumas variáveis (e.g. PIB real) se modificam com o tempo. Isso dificulta a construção de medidas-resumo de fácil interpretação que auxiliariam na análise dos resultados;
- O desafio da causalidade. Por não ser uma metodologia tradicional nos estudos estatísticos mais frequentes, e por ser um conceito bastante profundo e até mesmo filosófico, avaliar a causalidade é uma questão assaz desafiadora. Uma discussão mais aprofundada deste tópico pode ser encontrada na Seção 7.

6 Análise Descritiva

A análise descritiva dos dados foi separada em dois enfoques principais: a análise em painel, considerando a estrutura longitudinal de grande parte das variáveis; e as medidas-resumo, que serão apresentadas em seguida.

6.1 Análise em painel

A descrição das variáveis com estrutura longitudinal por meio de gráficos de perfis é mais adequada (ver Singer, Nobre e Rocha, 2012). No apêndice A estão todas as figuras citadas nesta seção, assim como uma instrução de como analisá-las. Algumas características permanecem constantes no decorrer da história (e.g. região), enquanto as demais variáveis presentes do estudo - inclusive as respostas - alteram-se com o passar do tempo. Para essas variáveis que evoluem com o tempo, é relevante estudar o comportamento de sua variação temporal. Nessas variáveis em específico, optou-se por separar os perfis individuais por região, por se acreditar que existem diferentes padrões para cada subdivisão e que países de uma mesma região tenham características semelhantes (além de facilitar a visualização, também).

6.1.1 Características geográficas

Latitude, longitude e região

Latitude e longitude são valores constantes no tempo. Indicam a posição do país em relação à linha do Equador e ao meridiano de Greenwich. O par ordenado destes valores posicionam os países no plano cartesiano de modo a gerar o mapa-múndi.

As regiões foram classificadas segundo critérios político-geográficos, baseado na mistura de duas características: por proximidade geográfica (com uma parcial exceção da categoria 5) e por democratização (definidas com base nas contribuições teóricas dos especialistas da área). As 10 categorias são:

- 1) Europa oriental e região pós União Soviética
- 2) América Latina
- 3) África do norte e Oriente Médio
- 4) África subsaariana
- 5) Europa ocidental e América do norte
- 6) Ásia oriental
- 7) Sudeste da Ásia
- 8) Sul da Ásia
- 9) Pacífico
- 10) Caribe

Pela Figura A.1.1 e da Tabela A.2.1 observa-se que, por volta de 1960, observa-se que já estão presentes um grande número de países da Europa ocidental (17 dos 24 de 2006 já estão participando), enquanto que a África, o Oriente Médio e a Ásia ainda estão pouco representados. Em 1980 (Figura A.1.2) já se observa uma distribuição mais homogênea entre as regiões, principalmente com a substancial participação dos países africanos subsaarianos (a maioria dos países em 1980 pertencem a essa região e a proporção de países participantes em relação a 2006 é de $27/38 = 71\%$). Em 2006 (Figura A.1.3) estão presentes todos os países que fazem parte da GATT/OMC. Há uma aparente homogeneidade quanto à representatividade de cada região na organização.

Hemisfério leste ou oeste e Hemisfério norte ou sul

Variáveis constantes no tempo. Indicam, conjuntamente, em qual quadrante do mapa mundi (definidos pelo Equador e pelo meridiano de Greenwich) o país está localizado.

Nomeando os quadrantes por NE, NW, SE, SW (Figura A.1.4, Figura A.1.5 e Figura A.1.6), observamos das Tabelas A.2.2, A.2.3 e A.2.4 que a proporção de países

nestas regiões se mantém ao longo do tempo. Em 2006, o Norte representa 77% do total de países membros. No mesmo ano, o Leste representa 66% do número total de países participantes, o que era esperado, uma vez que no lado ocidental do mapa mundi encontra-se apenas o continente americano.

Proporção de terra dentro dos trópicos

Variável constante no tempo. Indica o quanto da área dos países encontram-se compreendidos entre os trópicos de Câncer e de Capricórnio, em porcentagem da área total. Em 1960, metade dos países possui área nos trópicos (Tabela A.2.5), e destes, 78% possuem proporção acima de 50% (Tabela A.2.6) Em 1980, a proporção de países com terra nesta área aumenta para 61% (Tabela A.2.5). Já em 2006, praticamente mantém-se a proporção de países com áreas no trópico (Tabela A.2.5) e 87% destes países têm mais da metade de sua área compreendida entre os trópicos (Tabela A.2.6).

Ano [year] (de 1960 a 2006)

Variável definidora do tempo e caracterizadora da estrutura em painel dos dados. Todas as demais variáveis foram coletadas condicionadas pelo ano (justificando que a suspeita de existência de uma correlação serial entre as respostas seja razoável).

6.1.2 Características demográficas

Nome do país [cname]

Nome do país em 2006, identificando a unidade amostral do estudo. Os países participantes estão listados na Tabela A.2.7.

População [pwt_pop] (em milhares)

Tamanho da população dos países. Na Figura A.1.7 tira-se que o crescimento do logaritmo do tamanho populacional cresce linearmente com o tempo. Além disso, pelo padrão não paralelo dos segmentos, as velocidades de crescimento dos países diferem

entre si. Tais velocidades podem ser representadas pelas inclinações de cada um dos segmentos.

Mortes relacionadas a batalhas [wdi_brd]

Esta quantidade apresenta valores positivos em raras ocasiões que são, em geral, justificadas por fatos históricos específicos. Com um par de exceções de países que apresentam valores positivos regularmente ao longo dos anos, a Figura A.1.8 ilustra o excesso de zeros para os demais países membros.

População urbana [wdi_urban] (em porcentagem)

Proporção da população total do país que vive em áreas urbanas. A Figura A.1.9 apresenta um perfil de paralelismo, onde - em geral - a relativa diferença entre os países é sustentada ao longo dos anos (observações dessa natureza serão frequentes para justificar a criação de medidas resumo para grande parte das variáveis envolvidas na pesquisa). Vide subseção 6.2.

Consumo de energia elétrica [wdi_epc] (em KWh per capita)

Com padrão paralelo (vide Figura A.1.10), o logaritmo do consumo de energia elétrica per capita cresce linearmente no tempo. Para países da região do Pacífico, note que não há informações disponíveis.

Linhas telefônicas [wdi_pl] (por 100 pessoas)

Denotando o número de linhas telefônicas por cada 100 habitantes, esta variável apresenta um comportamento interessante. Primeiro, um problema: por estar relacionada com uma tecnologia relativamente nova quando da criação do GATT, informações a seu respeito até 1975 eram escassas, aparentemente coletadas periodicamente, resultando em um excesso de observações faltantes até esta data (ver Figura A.1.11). Por último, uma evolução temporal peculiar: até meados dos anos 1990,

o crescimento de linhas era tímido; durante a década de 90 houve um aumento explosivo neste número; de 2000 até 2006, o número de linhas entrou em declínio, por razões que os dados apenas permitem especular. Mesmo assim, um perfil paralelo entre os perfis individuais é observado neste caso.

6.1.3 Características econômicas

Taxa de crescimento do PIB real per capita [pwt_grgdpch] (taxa a preços constantes)

A taxa de crescimento do PIB real per capita flutua em torno de uma média comum (próxima de 2) para todas as regiões, como pode ser visto na Figura A.1.12. Países da África subsaariana, da África do norte e do Oriente Médio possuem uma maior variabilidade, apontando por uma maior instabilidade econômica nessas regiões. Em contraste, a região da Europa ocidental e América do Norte (região 5) apresentam a menor volatilidade.

Abertura para o comércio [pwt_openk] (em porcentagem do PIB real)

Com perfis semelhantes entre as regiões, a abertura para o comércio veio sofrendo leve alteração (crescente) com o tempo, girando em torno de 126% do PIB real para o ano de 2006 (Figura A.1.13). A região 5 tem países que se destacam para mais, assim como alguns países da Ásia (Sul e Sudeste). Podem-se observar perfis paralelos novamente.

PIB per capita [pwt_rgdpch] (a preços constantes do PIB)

Trabalhando com o logaritmo do PIB per capita, a Figura A.1.14 deixa visível a diferença das economias entre as regiões. Enquanto a Europa ocidental e a América do Norte (vale apontar que Austrália e Nova Zelândia também fazem parte desta região) possuem um elevado PIB para todos os países correspondentes, a África subsaariana contém países estritamente mais pobres do que a primeira região e com maior

variabilidade de tamanho econômico, indo desde muito pobre até acima da média. América Latina, Caribe e Europa Oriental se mostram semelhantes em todos os aspectos. Já os países da Ásia oriental e do Sudeste da Ásia apresentam grande variedade de níveis de riqueza.

É importante destacar que aqui, como em todas as características que apresentam paralelismo entre os perfis individuais, a ordem relativa do tamanho do PIB dos países é mantida ao longo do tempo, ou seja, se o país A era mais rico do que o país B em 1960, ele continuará sendo, em regra, mais rico em 2006.

PIB real [pwt_rgdp] (multiplicação do PIB per capita pela população)

Novamente trabalhando com o log, o PIB real, em contraste com o log do PIB per capita, deixa transparecer uma maior variabilidade dentro das regiões (Figura A.1.15). Destaque para o Caribe (um arquipélago) que se assemelha à América Latina quanto ao PIB per capita, mas se mostra abaixo relativamente à mesma região para o PIB real. O paralelismo é observado novamente.

Exportação [wdi_exp] (em porcentagem do PIB real)

A porcentagem de exportação mantém-se aproximadamente constante ao longo do tempo, apresentando algumas exceções como no caso do Sudeste da Ásia (principalmente a China, que nos últimos anos passou por uma revolução econômica notável) e da região 5. Entre as regiões, o perfil praticamente não se altera (Figura A.1.16).

Importação [wdi_imp] (em porcentagem do PIB real)

A porcentagem de importação segue um padrão muito próximo do padrão da porcentagem de exportação, diferenciando-se apenas por uma ligeira variabilidade mais acentuada (Figura A.1.17).

Comércio total [wdi_ttr] (em porcentagem do PIB real; equivale a wdi_exp + wdi_imp)

Pela semelhança citada nos dois itens acima e pelo fato do comércio total ser calculado como a soma das porcentagens de exportação e importação, o perfil desta quantidade é igualmente semelhante às suas componentes, diferenciando-se apenas na magnitude das quantidades (Figura A.1.18).

Participação do governo no PIB [pwt_gsg] (em porcentagem)

Países da África subsaariana e o sul da Ásia são as regiões que se mostram ser mais heterogêneas em relação à participação do governo no PIB. As demais regiões comportam-se aproximadamente da mesma maneira, com uma participação média de 15%. A região da Europa ocidental e América do Norte é a mais estável em termos de variação da participação. Em geral, tais perfis são paralelos entre si e paralelos ao eixo do tempo (Figura A.1.19).

6.1.4 Características de democracia

Escore de democracia [p_polity] (varia de -10, ditadura completa a 10, democracia completa)

Pela própria natureza da quantidade que esta variável propõe-se a medir, muitos dados omissos ocorrem por dificuldade em mensurar ou até mesmo em acessar as informações necessárias para tal. O escore vai de +10 (fortemente democrático) até -10 (fortemente autocrático) e apresenta, como característica, forte variação dentro de um mesmo país, principalmente os países africanos subsaarianos.

Nota-se que houve uma “democratização” em nível mundial depois de 1990, além de haver menos dados faltantes nesse período. Uma região que chama a atenção é a região da Europa oriental e região pós União Soviética que, quando tinha

informação, eram valores baixos até meados de 1990 e depois disso, mais informações tornaram-se disponíveis e de magnitude elevada, como mostra a Figura A.1.20. Um fato interessante, neste caso, é que os dados faltantes podem não ser oriundos de mera casualidade, pois podem estar intimamente ligados, por exemplo, a valores negativos (ditaduras podem ter menos liberdade de expressão e de acesso a dados).

Indicador de democracia [chgademo] (0 se ditadura e 1 se democracia)

Este indicador proporciona uma medida binária para regimes políticos (democracias e ditaduras). Uma vantagem é que este indicador apresenta menos casos de *missings*. Neste caso, fica clara a intensa mudança de regime que veio ocorrendo na África subsaariana. Consegue-se observar o mesmo para a América Latina. Em contra partida, para as regiões com características políticas bem definidas, as regiões Europa ocidental e América do Norte contrastam com as regiões Sudeste da Ásia e Europa ocidental e região pós União Soviética, sendo o primeiro predominantemente democrático e os últimos predominantemente autocráticos, até meados de 1990. Veja a Figura A.1.21.

6.1.5 Variáveis respostas

Proporção de coalizões dentro da região por ano

A Figura A.1.22 tem o importante papel em justificar o uso da região como um fator de agrupamento de países com perfil semelhante. Além das evidências encontradas em todas as análises de perfis acima, nas quais se notam comportamentos específicos dentro de cada região, este gráfico mostra a alta correlação entre os países de uma mesma região quanto à entrada em coalizões em cada ano (pela própria construção da variável que inclui acordos regionais). O gráfico da Figura A.1.22 está construído da seguinte maneira: cada ponto do gráfico representa a proporção de países dentro da região que entrou em coalizão no respectivo ano. Por exemplo, em 1964, na região da América Latina, 100% dos países da região que faziam parte da GATT/OMC naquele ano entraram em coalizão. Enfatiza-se que a proporção

de países de cada ano é relativa ao total de países da região participantes naquele específico ano e não em relação ao total de países que são participantes nos dias de hoje (em 2006).

A área sombreada em segundo plano indica, por sua vez, a parcela de países da respectiva região já presentes na GATT/OMC em relação ao número total de presentes em 2006. O número n indicado no canto superior direito de cada gráfico denota a quantidade de elementos dentro daquela região em 2006. Por exemplo, em 1964 a América Latina já tinha 40% (pois a área sombreada está no nível de 0.4) dos $n = 20$ de seus membros que ainda virão a participar até 2006. Ou seja, na ocasião existiam apenas 8 países dos 20 que hoje estão dentro da GATT/OMC, e desses 8, 8 fizeram coalizão naquele ano.

O mais importante desta Figura é a presença preponderante de altos picos quando estes não são zero, indicando que, em geral, países de uma mesma região manifestam-se conjuntamente.

Nota-se, também, um aumento considerável na atividade dos países do mundo inteiro depois de 1982. Antes as manifestações eram escassas, com longos períodos de inatividade na maioria das regiões. Após meados de 1980, a ocorrência de coalizões torna-se mais frequente, ainda assim mantendo as altas proporções de países de uma mesma região entrando em coalizão.

Este fato, na verdade, tem explicação. Do ponto de vista histórico-teórico, pode-se dizer que as coalizões começaram a surgir a partir de uma série de mudanças que estavam ocorrendo no cenário internacional: ascensão do liberalismo de Thatcher e Reagan, e declínio da hegemonia norte-americana e confronto com novos países industrializados com práticas até então protecionistas.

6.2 Medidas-resumo

A utilização de medidas-resumo (Singer, Nobre e Rocha, 2012) é uma prática bastante comum em análise de dados longitudinais. Em geral, esta técnica é utilizada quando o pesquisador deseja controlar os indivíduos no tempo, mas o interesse não é analisar o efeito do tempo, e sim o comportamento da resposta dos indivíduos em relação a algumas outras covariáveis de interesse. Desse modo, existe a necessidade de redução da dimensionalidade do problema, transformando um estudo essencialmente longitudinal em uma análise mais simples e direta.

Mas por que controlar as variáveis no tempo, se o tempo não é o interesse principal da pesquisa? A pergunta pode ser respondida observando-se o fato de que controlar a correlação intrínseca dos indivíduos pode melhorar o poder dos testes realizados. Em suma, mesmo que o pesquisador, essencialmente, queira comparar a distribuição da variável resposta submetida a diferentes grupos (ou a uma covariável), se o estudo permite a avaliação dos indivíduos em diferentes situações, a análise poderá ser mais consistente.

Mais do que isso, a variável resposta do problema pode ser um valor agregado das entradas em coalizões avaliadas ano a ano. Nesse caso, é interessante a utilização de medidas-resumo, que atingem diretamente o objetivo da pesquisa e cuja interpretação é mais simples do que a fornecida pela análise em painel.

Para esta mudança de paradigmas, levantam-se as seguintes vantagens e desvantagens:

- **Vantagens:**

- Simplificação da análise;
- Facilidade de interpretação;
- Todas as informações relevantes ao objetivo da pesquisa contidas no painel são contempladas pelas medidas-resumo, desde que todos os tempos sejam considerados nos cálculos.

- **Desvantagens:**

- A interpretação dos resultados indexada no tempo se perde;

- Muitas vezes pode ser difícil construir medidas-resumo que tenham relação direta com o interesse do pesquisador. Nesse contexto, existem várias técnicas disponíveis na literatura;
- Quando as variáveis explicativas também se alteram no tempo, necessitam ser resumidas assim como a resposta. Desse modo, o esforço necessário para construir medidas-resumo que tenham significado interpretativo aumenta significativamente. Isto poderia ser um ponto limitador caso houvesse um número elevado de variáveis de interesse.

Voltando-se para o objetivo primário do pesquisador, que é traçar o perfil dos países que mais entram em coalizões, nota-se que este estudo se encaixa nas premissas necessárias para a utilização de medidas-resumo. Contudo, pelo fato de grande parte das variáveis explicativas - inclusive as econômicas, que são de maior interesse - também variarem no tempo, cada uma delas foi analisada individualmente, com o objetivo de construir medidas adequadas.

6.2.1 Construção das medidas-resumo

A partir deste ponto, apresentar-se-ão as medidas-resumo que foram adotadas na análise exploratória dos dados.

Variáveis respostas: Três foram as propostas de se construir um perfil de entradas em coalizão de cada país:

Taxa de entrada em coalizões. É dada pela razão do número total de entradas em coalizões e o número de anos que o país esteve presente no estudo. Assim, quanto maior o número observado, maior a propensão de o país entrar em coalizões;

Proporção de entrada em coalizões. É dada pela razão do número total de anos que o país entrou em coalizões e o número de anos que o país esteve presente no estudo. Proporções altas indicam que a frequência de entradas em coalizões é alta. Um

fato interessante dessa medida-resumo é que, pelo fato de estar compreendida entre zero e um, é de fácil interpretação e tem ligação direta com o interesse do pesquisador;

Taxa ponderada de entrada em coalizões. Para a construção desta medida, em cada ano foi calculada a razão do número de entradas em coalizões e o total de entradas para todos os países naquele ano. Estas medidas foram então reponderadas pelo número de países naquele ano e, por fim, todos os anos foram somados. Apesar de ser mais difícil a interpretação, essa variável leva algumas características dos períodos em consideração, dando maior peso a entradas em coalizão em períodos em que houve poucas coalizões no mundo todo.

Variáveis explicativas: Para as variáveis explicativas, quatro técnicas distintas de resumos foram empregadas: média, análise de desfecho e área sob a curva. A Tabela B.2.1 apresenta uma relação de quais medidas resumo foram aplicadas em quais variáveis.

Média ou proporção: O valor médio da variável é a medida resumo mais direta e intuitiva possível, mas só é adequado empregá-la em alguns casos;

Análise de desfecho: Esta medida resumo considera somente o último período avaliado (2006). As variáveis para as quais foi empregada a análise de desfecho em geral foram as que apresentaram paralelismo nos gráficos de perfis. O paralelismo indica que a relação de ordem dos países não se altera com o tempo, ou seja, se uma reta for ajustada para cada um dos perfis individuais, todos terão o mesmo coeficiente angular. Desse modo, em uma análise comparativa, seria equivalente obter como medida-resumo as medidas em qualquer um dos anos. O ano de 2006 foi escolhido justamente por possuir mais dados que os demais, e por ter uma interpretação bastante simples;

Área sob a curva: Para variáveis que oscilam no tempo, utilizar a análise de desfecho ou simplesmente a média não apresentam o verdadeiro perfil do país. Tome-se, por exemplo, a variável que indica o escore de democracia. Nota-se claramente que

as modificações bruscas dos países não são levadas em conta. Por esse motivo, uma proposta é calcular a área do polígono gerado por cada perfil individual, ponderado pelo número de anos em que o país esteve presente na amostra. Este valor é interpretado como uma espécie de “democracia acumulada”, e leva em conta todas as oscilações dos valores observados;

Coeficiente angular da reta: Alguns gráficos de perfis apresentam clara tendência linear. Contudo, nem sempre os crescimentos ou decrescimentos acontecem da mesma forma. É o caso, por exemplo, do logaritmo da população. Isso motiva criar uma medida-resumo que é o coeficiente angular da reta ajustada. Valores mais altos indicam que a taxa de crescimento da população é acentuada. Valores negativos indicam que a população está decrescendo.

Nas variáveis em que foram encontrados dados omissos, houve imputação pela média da região que a população estava presente. Esta metodologia é explicada pela relação entre a região e as variáveis explicativas encontrada na análise de perfis individuais.

6.2.2 Análise descritiva das medidas-resumo

Variáveis respostas

Em primeiro lugar apresentam-se, na Tabela B.2.2, as medidas de posição e variabilidade das variáveis respostas construídas utilizando-se o primeiro critério de coalizão (ver Seção 4). Observa-se que o coeficiente de variação da proporção de coalizões é inferior às demais. A partir dos gráficos na Figura B.1.1 pode-se observar a presença de *outliers* em todas as variáveis.

É possível também estabelecer relações entre as variáveis respostas construídas, com o fim de avaliar se elas possuem distribuições semelhantes, isto é, se as metodologias empregadas para seu cálculo são equivalentes. Observa-se pelos gráficos de dispersão da Figura B.1.2 que os métodos são semelhantes principalmente

para valores mais baixos da variável resposta. Além disso, apresentam-se as correlações das medidas-resumo das respostas. É interessante notar que as correlações são bastante altas (sendo a maior correlação de 86% entre taxa e proporção de coalizões), mas não o suficiente para ser possível afirmar que as metodologias são equivalentes.

Pelos gráficos e pelas retas ajustadas, pode-se perceber que os valores aberrantes influenciam bastante na correlação. O país que apresenta valor mais aberrante, segundo os três métodos, é a China. Por ser um dos países que entrou mais tarde na OMC, e por ter feito várias coalizões desde que entrou, a China apresenta um perfil bastante distante dos demais.

A mesma análise também foi realizada utilizando-se o segundo critério de coalizão (ver Seção 4). Nota-se nos gráficos B.1.3 e B.1.4 que os resultados encontrados foram bastante semelhantes. Neste caso, naturalmente, os valores da resposta ficaram abaixo do primeiro critério, já que o segundo critério é construído a partir do primeiro, retirando-se algumas coalizões. Observa-se também que as correlações entre os métodos são mais acentuadas do que as anteriores, chegando a 93% entre a taxa e a proporção de coalizões.

No restante do presente relatório, na maioria das vezes as análises serão voltadas somente para a variável dada pela proporção de coalizões, pela sua simplicidade de interpretação e ligação direta com o interesse da pesquisa. Contudo, como na análise exploratória não ficou claro qual seria o melhor método a ser adotado, na análise inferencial doravante apresentada, as outras possibilidades também serão levadas em conta.

Variáveis explicativas

Inicialmente, apresenta-se na tabela B.2.3 as medidas de posição e variabilidade das medidas-resumo obtidas como descrito na subseção 6.1.1. Além disso, apresentam-se nos gráficos B.1.5 ao B.1.16 os histogramas. Observa-se que a grande

maioria das variáveis parecem bem comportadas, com poucos casos de observações aberrantes.

Também é interessante analisar as correlações entre as variáveis explicativas, como é apresentado na tabela B.2.4. Observa-se que as variáveis econômicas são bastante correlacionadas, mas as de democracia, nem tanto. Um fato interessante é a alta correlação do logaritmo do PIB real e do logaritmo do número total de ligações telefônicas (93%). Este resultado será muito importante para a análise proposta na Seção 7.

Relação entre as respostas com as variáveis explicativas

Uma importante análise preliminar antes de ajustar qualquer modelo é construir os gráficos de dispersão da variável resposta (no caso, a proporção de coalizões) e as variáveis explicativas, com o intuito de obter uma relação marginal entre as respostas. Da Figura B.1.17 à Figura B.1.26 esses gráficos são apresentados.

Nota-se que, em todos os casos, a relação marginal entre a proporção de coalizões e a variável explicativa é bastante fraca, mostrando que essa associação pode não existir ou então pode não ser linear. No entanto, observa-se que a curva de suavização (método Loess, ver Hastie e Tibshirani, 1990) para a variável dada pela taxa média do crescimento do PIB real é inclinada, de modo a indicar, marginalmente, que variáveis econômicas podem estar associadas com a proporção de coalizões. Isto é um indicativo a favor da hipótese do pesquisador.

Além disso, na tabela B.2.5, são apresentadas as correlações entre as variáveis respostas (proporção, taxa e taxa ponderada) segundo o primeiro critério de coalizões e as medidas-resumo das variáveis explicativas. Fica evidente que nenhuma correlação é alta (todas, em valores absolutos, são menores que 50%).

7 Causalidade

De acordo com o segundo objetivo da Seção 2, o pesquisador tem interesse em atribuir relação de causalidade entre variáveis econômicas e a frequência de coalizões dos países, isto é, se certa característica econômica de um país implica em uma maior ou menor propensão daquele país entrar em coalizão.

Este tipo de análise, contudo, não é muito usual em análises estatísticas. Na maioria dos casos, o interesse do pesquisador está em medir associações. Para realizar a análise de causalidade dois enfoques podem ser seguidos: a utilização de variáveis instrumentais (Marra e Radice, 2010, Dawid, 2007 e Pearl, 2010) e árvore de decisão (Chipman, George e McCulloch, 2010). No presente relatório, ambas as abordagens serão utilizadas com o intuito de estudar a causalidade das características político-econômicas dos países na frequência de entradas em coalizões.

7.1 Variáveis instrumentais

Em geral, numa análise de dados, procura-se explicar uma variável endógena ou resposta, dependente do sistema, por variáveis exógenas, independentes do sistema (ver Barra e Radice, 2010). Para flutuações aleatórias e demais variáveis não controladas e pouco significativas, incluem-se os termos de erros aleatórios, que estruturam o modelo estatístico. Contudo, nem sempre é possível isolar as variáveis exógenas do sistema e, fatalmente, uma ou várias delas podem ser dependentes do termo de erro, fazendo com que as estimativas daquilo que se deseja estudar tornem-se enviesadas e inconsistentes.

Para resolver esse problema, a abordagem mais natural seria incluir mais variáveis exógenas que sejam relacionadas com a já existente no modelo. Desse modo, seria possível controlar a endogeneidade dessa variável explicativa. Contudo, nem sempre essas variáveis são observáveis. É o caso da relação do logaritmo do PIB real com as coalizões. Espera-se que o logaritmo do PIB real possa explicar a variabilidade das coalizões, mas existem diversos fatores, como o custo de transporte dos

diplomatas, crises econômicas, trocas de regimes políticos, entre outros, que não gozam de estruturas de dados confiáveis ou nem mesmo disponíveis.

Existem, no entanto, variáveis mais fáceis de serem obtidas, que não se relacionam com aquelas que não são observáveis, mas têm uma correlação muito alta com a variável da qual se deseja retirar a endogeneidade. A essas variáveis atribui-se o conceito de variáveis instrumentais. Segundo o pesquisador, esse é o caso do logaritmo do número total de linhas telefônicas de cada país. Aparentemente, esse valor não tem relação alguma com os fatores que influenciam internamente as coalizões através do logaritmo do PIB real. Contudo, estudos empíricos revelam que é alta a correlação desse instrumento com o logaritmo do PIB real (isso será verificado mais adiante).

A partir das variáveis instrumentais é possível obter uma estimativa do erro que é produzido pelas variáveis não observáveis. Corrigindo esse erro de modo aproximado a partir de uma técnica apropriada, o ajuste do modelo revelará estimativas mais consistentes do verdadeiro efeito da variável exógena de interesse.

A necessidade de inclusão de variáveis instrumentais nos modelos para análise de dados consiste, portanto, no fato de nem sempre ser possível controlar todas as variáveis que influem tanto na variável resposta quanto na variável explicativa de interesse, seja por motivos práticos, seja por motivos éticos. Sua utilização é uma tentativa de minimizar o efeito de variáveis explicativas que são importantes para controle, mas não podem ser observadas, corrigindo a variável explicativa de interesse por outra que tem ligação mais direta com a resposta.

Mas como relacionar causalidade com o conceito de variáveis instrumentais? Para que essa relação fique clara, vale mencionar o exemplo de um estudo realizado por Leigh e Schembri (2004) (ver Marra e Radice, 2010): é possível realizar uma pesquisa que relaciona o consumo de cigarros com o status físico funcional. Claramente, existe uma associação entre o consumo de cigarros e riscos à saúde do indivíduo, que não podem ser controlados facilmente. Uma análise de regressão

relacionando o status físico funcional e o consumo de cigarros poderia fatalmente ter resultados incompatíveis com a realidade, em que um leigo poderia afirmar, por exemplo, que não há efeito do consumo de cigarro na variável resposta.

Por outro lado, pode-se utilizar a variável instrumental dada pelo preço do cigarro, que está ligado ao consumo, mas não às consequências na saúde do indivíduo. Realizando a análise corretamente, é possível corrigir de modo aproximado o erro causado pelas variáveis que não são observáveis, deixando somente o efeito do consumo de cigarro propriamente dito no problema. Só neste momento, ou seja, supondo que todas as variáveis de confusão estão sendo controladas, é adequado estabelecer uma relação causal entre a variável explicativa e a resposta.

7.2 Visualização da causalidade utilizando o GAM

Mais do que estabelecer uma relação causal, é de interesse do pesquisador a visualização dessa relação. Em outras palavras, deseja-se estabelecer uma função que liga as variáveis econômicas, em particular, ao logaritmo do PIB real, com a probabilidade do país entrar em coalizão, dadas as variáveis de confusão.

Para tanto, utilizar-se-ão as variáveis explicativas descritas na Seção 6, bem como a variável instrumental dada pelo logaritmo do número total de linhas telefônicas de cada país. A análise será realizada em dois estágios:

Em primeiro lugar, será feita uma regressão linear simples múltipla com o logaritmo do PIB como variável dependente (ver Kutner et al., 2005), sendo explicado pelo número total de linhas telefônicas e pelas demais variáveis de controle. Com isso, deseja-se obter uma estimativa do logaritmo do PIB real, este corrigido, a partir da variável instrumental, pelas variáveis não observáveis;

Utilizando-se as variáveis de controle e a estimativa corrigida do logaritmo do PIB real, ajustar-se-á no segundo estágio um modelo aditivo generalizado (GAM, Hastie e Tibshirani, 1990), que é semi-paramétrico. A resposta, dada pelas proporções de

coalizões de cada país, será modelada pela distribuição binomial (Bolfarine e Sandoval, 2010). O intuito da utilização desse modelo é estabelecer uma função de suavização entre a proporção de coalizões e o logaritmo do PIB real, dadas as demais variáveis.

É importante observar que, para ser adequada a passagem do estágio 1 para o estágio 2, foi necessário verificar a existência de uma alta correlação entre o logaritmo do PIB real e a variável instrumental. De fato, a relação linear é significativa (valor- $p < 0.0001$). O coeficiente de determinação da regressão múltipla ajustado (R^2) (ver Kutner et.al, 2005) é de 86%, e a correlação obtida do logaritmo do número total de ligações telefônicas e o logaritmo do PIB real ajustado é de 93%.

No segundo estágio, inicialmente, ajustou-se o modelo com todas as variáveis disponíveis como efeitos fixos, menos o logaritmo do PIB real ajustado, que foi adicionado utilizando-se *splines* cúbicas. O modelo foi reduzido utilizando-se o critério de informação de Akaike (AIC) (Kutner et al., 2005), resultando no modelo final que incluía somente os indicadores dos hemisférios (norte/sul, oeste/leste) e a função de suavização do logaritmo do PIB real ajustado.

A Figura B.1.26a representa a relação funcional entre a proporção de coalizões e o logaritmo do PIB para países localizados na região noroeste do planeta. Para países da região nordeste obtém-se o gráfico da Figura B.1.26b. Para países da região sudoeste pode ser visualizada na Figura B.1.26c. Finalmente, para países da região sudeste, a função é dada na Figura B.1.26d.

Observa-se que a relação funcional não é simples (número de graus de liberdade maior que seis), mas é possível notar que, em linhas gerais:

- Países com menor logaritmo do PIB real entram em coalizão com bastante frequência;
- À medida que o logaritmo do PIB real aumenta, a proporção de coalizões diminui e, em seguida, aumenta novamente;

- Dos países intermediários em diante, a proporção de coalizões cresce com o logaritmo do PIB real;
- A relação funcional não muda de região para região, apenas torna-se mais ou menos acentuada.

Um efeito interessante ocorre quando se retira a China do estudo, na Figura B.1.27 é possível ver que a subida dos países com maior logaritmo do PIB per capita é bem menor do que aparentava no outro ajuste. Isso pode ter acontecido, pois, além de ser um valor aberrante na proporção de coalizões, é o segundo PIB mais alto de 2006, perdendo somente para os Estados Unidos. Isso faz com que este país seja bastante influente no ajuste do modelo aditivo generalizado.

Para verificar o terceiro objetivo do pesquisador, ou seja, a robustez dos resultados quanto ao critério de identificação de coalizões, a mesma análise foi repetida para a variável que considera apenas coalizões que negociam somente no GATT/OMC. De fato, as conclusões aparentemente não mudaram. Na Figura B.1.28, observa-se que a forma funcional do efeito do logaritmo do PIB real torna-se mais atenuada, mas continua com o mesmo perfil. A atenuação pode dever-se ao fato de, na segunda variável, o número de coalizões consideradas é menor, sendo assim mais difícil de detectar a associação existente. O mesmo resultado anterior pode ser reproduzido retirando-se a China do estudo, como é possível observar na Figura B.1.29.

8 Análise Inferencial

Apesar dos modelos aditivos generalizados serem considerados modelos para análise inferencial, estes foram usados somente como técnica para visualização da possível relação funcional entre o logaritmo do PIB e a proporção de coalizões. Para a análise inferencial, que tem como objetivo avaliar se de fato esta tendência existe, uma metodologia mais flexível se faz necessária.

Nas próximas subseções serão introduzidos o modelo GLM (Generalized Linear Model) e um segundo enfoque, chamado BART (Bayesian Additive Regression Trees), que será tratado com maiores detalhes.

Para o ajuste dos modelos, foram utilizadas as medidas-resumo para o período a partir de 1982. A decisão de trabalhar somente neste período é motivada pelo fato de que, como visto na análise descritiva, longos períodos de inatividade antes de 1982 poderiam distorcer as análises.

Além disso, serão apresentados somente os resultados para a resposta ‘proporção de coalizões’, segundo os dois critérios de coalizão definidos pelo pesquisador. Optou-se por essa variável pela sua simplicidade, facilitando tanto a modelagem quanto a interpretação dos resultados. Na Seção 10, considerações serão feitas a respeito das modelagens das outras respostas.

8.1 Modelo Linear Generalizado

Um possível enfoque para modelar as medidas-resumo, estabelecendo uma relação entre as variáveis explicativas de interesse e as respostas é o GLM (sigla inglesa para modelo linear generalizado). Este modelo consiste, como o próprio nome diz, numa generalização do modelo linear normal, em dois sentidos: primeiro, a resposta não precisa necessariamente possuir distribuição gaussiana; segundo, a relação funcional entre o valor esperado da resposta e as variáveis explicativas não precisa ser a identidade. Os modelos lineares generalizados foram introduzidos por Nelder e Weddenburn (1972) e popularizados com o livro de Nelder e McCullagh (1983).

Um fato notável é que, na realidade, o modelo aditivo generalizado (GAM) é uma generalização do GLM. Contudo, para fins inferenciais, o GLM oferece um ambiente mais simplificado, resumindo relações funcionais não paramétricas em simples coeficientes lineares. Além disso, o GLM também possui alternativas para utilização de variáveis instrumentais. O objetivo de ajustar um modelo linear generalizado, neste

contexto, é de corroborar com os resultados que foram encontrados até o momento com a utilização do GAM, e também com o BART (que está apresentado na Subseção 8.3).

8.1.1 Especificação e ajuste do modelo

O ajuste do GLM foi realizado em dois estágios. No primeiro, foi construída a variável ajustada do logaritmo do PIB, a partir da variável instrumental, utilizando-se a mesma metodologia descrita na Seção 7.2 (regressão linear simples). No segundo estágio, foi considerada a resposta com distribuição Binomial (ou seja, o número de entradas em coalizões foi suposto possuir distribuição binomial), com função de ligação logística. Esta metodologia de atribuição da distribuição da resposta e da função de ligação é também conhecida como regressão logística. Para mais detalhes sobre a especificação do modelo, ver McCullagh e Nelder (1983). Para a seleção de modelos, foi utilizado o critério de informação de Akaike (AIC) (Kutner et al., 2005). O modelo final obtido (apresentado a seguir) contempla somente as variáveis de controle indicando o quadrante hemisférico (L/N, L/S, O/N, O/S) e logaritmo do PIB ajustado (**LPIBa**). O **LPIBa** foi centrado para facilitar a interpretação dos parâmetros. Um sumário com as estimativas dos parâmetros, erros padrão e valores- p estão apresentados na Tabela C.1.1. Os erros padrão foram ajustados pelo método de Murphy-Topel (Murphy e Topel, 1985), para ficar mais adequado com a utilização de variáveis instrumentais.

Para interpretação dos resultados obtidos, é necessário introduzir o conceito de **chances** e **razão de chances**. A chance nada mais é do que a razão da probabilidade de um evento sobre a probabilidade do complementar desse evento. Existe uma relação biunívoca entre a probabilidade de um evento e a chance, de modo que uma probabilidade alta corresponde a uma chance alta, e vice-versa. A razão de chances é, como o próprio nome apresenta, a razão da chance de um evento pela chance de outro evento. Se esta razão assumir um valor α , isso indica que a chance do evento que está no numerador é α vezes a chance do evento que está no denominador. Os resultados da regressão logística construída foram apresentados utilizando-se essa terminologia.

Assume-se que

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \mu_j + \beta(\text{LPIBa}_{ij} - \overline{\text{LPIBa}})$$

com $j = 1, 2, 3, 4$ denotando os quadrantes (L/N, L/S, O/N, O/S, respectivamente), $i = 1, \dots, n_j$, e

n_j : número de países situados no quadrante j ;

π_{ij} : probabilidade do i -ésimo país situado no quadrante j formar coalizão;

μ_j : Para países da j -ésima região do globo, este parâmetro denota o *logaritmo da chance de um país entrar em coalizão dado que seu **LPIBa** é igual ao logaritmo do PIB médio dos países participantes do GATT/OMC ($\overline{\text{LPIBa}}$)* (exceto a China, que foi retirada do modelo). Assim, e^{μ_j} representa a chance acima definida.

β : Este coeficiente é o *logaritmo da razão das chances ao incrementar o **LPIBa** em uma unidade a partir de um valor do **LPIBa** fixado, quando comparados países situados no mesmo quadrante*. Desse modo, e^{β} representa a razão de chances que acabou de ser definida.

Os modelos foram ajustados para ambos os critérios de coalizão. Pela grande semelhança dos resultados, apenas o primeiro foi considerado nas interpretações. No Apêndice C, foram apresentados todos os resultados referentes a estes modelos.

8.1.2 Resultados

No modelo utilizado, o evento de interesse é a entrada em coalizão. Ou seja, a chance de entrada em coalizão $\frac{\pi_{ij}}{1 - \pi_{ij}}$ corresponde à razão da probabilidade de um país entrar em coalizão pela probabilidade de ele não entrar em coalizão. Para o primeiro critério de coalizões, os resultados obtidos foram (Tabela C.1.1):

- A região com maior chance de coalizão ($e^{\hat{\mu}}$), dado que o **LPIBa** é igual ao ($\overline{\text{LPIBa}}$), é a região O/S (chance $0,41 \pm 0,07$), enquanto que a região com menor chance é a L/N (chance $0,19 \pm 0,02$);
- Como a razão de chances é maior que um ($1,17 \pm 0,03$), pode-se afirmar que a chance de coalizão cresce com o aumento do logaritmo do PIB ajustado.

Pelos gráficos de diagnóstico apresentados na Figura C.1.5 e na Figura C.1.7, o modelo parece bem ajustado, considerando-se que o estudo em questão é observacional. Quatro possíveis pontos de influência foram detectados, sendo um deles o Uruguai (observação 143 na Figura C.1.1), o Djibouti (observação 47), a Índia (62) e o Reino Unido (139). Análises sem estas duas observações não sofreram alterações importantes.

Os coeficientes estimados para o modelo considerando o segundo critério de coalizão foram apresentados a Tabela C.1.2. Além disso, algumas comparações com a utilização ou não de variáveis instrumentais e ajuste utilizando a regressão por dois estágios foram apresentadas nas Tabelas C.1.3 e C.1.4.

8.2 Modelo BART

A especificação do modelo BART segue a estrutura geral dos modelos de regressão (ver Kutner et. al, 2008), em que se deseja ajustar a resposta Y por um conjunto de p variáveis explicativas $x = (x_1, x_2, \dots, x_p)$, por intermédio de uma função f e um erro aleatório ε , de modo que

$$Y = f(x) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

Observa-se que o BART assume os erros com distribuição normal. Além disso, a função f é aproximada por uma estrutura de **soma de m árvores de decisão**, ou seja,

$$f(x) \approx h(x) = \sum_{j=1}^m g_j(x).$$

Mais detalhes sobre a especificação do modelo, como a parametrização, metodologia utilizada para ajuste e outras características constam no Apêndice D. Ver também Chipman, George e McCulloch (2010).

As principais vantagens do modelo BART são:

- Ajuste flexível (não paramétrico) para quaisquer relações funcionais, não necessariamente lineares, nem mesmo contínuas;
- Possibilidade de construção de intervalos de confiança para cada país, obtidos através de simulações;
- Detecção de possíveis interações automaticamente, sem necessidade de especificação;
- Tem a capacidade de lidar com a possível heterogeneidade das variáveis explicativas, sem a necessidade de ajuste em dois estágios e utilização de variáveis instrumentais;
- Permite uma abordagem imediata para estimação do “efeito causal”.

O modelo possui diversas características desejáveis. Em especial, a capacidade de gerar relações não necessariamente contínuas, que permitiu a substituição das variáveis de hemisfério (utilizadas no GAM) pelas variáveis de latitude e longitude, contribuindo para um melhor controle nos modelos utilizados.

Assim, o BART será tomado como modelo principal da análise, de tal forma que as inferências em questão terão este modelo como base e os demais como auxiliares.

Uma possível dificuldade para ajuste do BART é a origem da variável resposta, que é uma proporção, ou seja, está no intervalo entre zero e um, o que é um problema, considerando-se que é suposta a distribuição normal dos erros. Por isso, foi considerada na variável resposta a transformação probito, que possui uma

probabilidade no intervalo dos números reais. Para contornar problemas relacionados a países que não apresentaram coalizões durante todo o período, foi realizado um *shift* (ou incremento) dos dados em 0,5%. Esta técnica foi utilizada unicamente para o ajuste do modelo, sem complicações nas inferências que possam ser feitas.

8.2.1 Seleção de variáveis

A Figura C.2.5 mostra as proporções de vezes que cada variável participa das árvores de decisão (critério usado para selecionar variáveis, ver apêndice D) em ajustes do modelo com 10, 20, 50 e 200 árvores, para ambos os critérios de coalizão. Observa-se, com base no critério dado, que as variáveis de controle mais importantes são as regionais (latitude e longitude), e as variáveis explicativas mais importantes são o logaritmo do PIB e o escore de democracia acumulado. Estas variáveis, além de terem interesse direto na pesquisa, também aparecem nos ajustes dos modelos GAM e GLM.

Portanto, o modelo final construído baseou-se somente nas variáveis regionais de latitude e longitude, no logaritmo do PIB e no escore de democracia acumulado. As análises a seguir serão baseadas nessas variáveis.

8.2.2 Utilização do BART para visualização de efeitos marginais

Uma possibilidade bastante flexível do modelo utilizado é a avaliação dos efeitos marginais das variáveis explicativas. A construção destes efeitos se dá pelas simulações no processo de criação das árvores (McCulloch, Chipman e George, 2010). Uma vantagem do BART é que também é possível estimar efeitos conjuntos das variáveis, fornecendo uma forma de visualizar como duas variáveis explicativas interagem nas suas relações com a resposta.

Figura C.2.6: efeito marginal do logaritmo do PIB na função probito das probabilidades de coalizão estimadas. Observa-se neste gráfico a mesma tendência crescente observada nas análises, de modo que países mais ricos possuem maior

probabilidade de coalizão, para ambos os critérios de coalizão. Além disso, é possível identificar, graficamente, pontos de corte ou “saltos” no comportamento da relação entre o logaritmo do PIB e as probabilidades estimadas. A primeira faixa apresenta uma tendência de subida, a segunda permanece constante e a terceira volta a apresentar inclinação positiva. No Apêndice C.2 a listagem dos países em cada faixa está apresentada.

Figura C.2.7: efeito marginal do escore de democracia acumulado nas probabilidades de coalizão estimadas. Para esta variável, fica clara a mudança de comportamento para certo ponto de corte, em ambos os critérios de coalizão. Na primeira faixa, não há uma tendência clara e, na segunda faixa, a tendência é crescente. Ou seja, existe um ponto p_0 (12,5 para o primeiro critério de coalizões, considerando acordos regionais e 12 para o segundo critério) tal que país com escore de democracia acumulada menor que p_0 , possuem probabilidades de coalizão estimada iguais, enquanto que, se o país possuir escore maior que p_0 , a probabilidade de coalizão estimada aumenta com o crescimento do escore.

Figura C.2.3: efeito marginal conjunto do logaritmo do PIB e do escore de democracia acumulado nas probabilidades de coalizão estimadas. Este gráfico apresenta três variáveis cruzadas: no eixo das abscissas, encontra-se o escore de democracia acumulado. No eixo das ordenadas, o logaritmo do PIB. Por fim, nas cores e nas curvas de nível, apresentam-se as proporções de coalizões estimadas (azul escuro indica baixos valores, verde indica valores intermediários e amarelo indica altos valores). Os pontos representam os países. Observam-se no gráfico resultados semelhantes aos encontrados na análise marginal, ou seja, existe aumento na probabilidade estimada de coalizões com o aumento do logaritmo do PIB e do escore de democracia. É interessante notar que a probabilidade de coalizão estimada, para os dois critérios de coalizão, só assume valores altos em relação aos demais quando ambas as variáveis explicativas tomam seus maiores valores.

Figura C.2.4: efeito marginal conjunto de latitude e longitude. O objetivo principal ao construir este gráfico foi obter uma visualização dos valores estimados do

modelo no mapa-múndi. Seguindo a mesma ideia da Figura C.2.3, o gráfico apresenta a longitude e a latitude nas abscissas e ordenadas, respectivamente. A cor azul escura indica valores estimados baixos; a cor verde, valores intermediários; a cor amarelada valores altos. Pode-se observar uma “nuvem” de baixa probabilidade estimada de coalizão concentrada no lado leste da Europa. Nas Américas, na África, na Oceania e no sudeste da Ásia, os valores aumentam. É importante ressaltar que a China foi retirada do estudo e por esse motivo, na região onde ela se encontra originalmente, as probabilidades estimadas foram relativamente baixas.

8.2.3 Visualização do “efeito causal” do PIB segundo o BART

Outra facilidade do modelo BART é a possibilidade de realizar inferências causais de modo direto (Hill, 2011). A metodologia consiste, basicamente, em submeter os países em questão a diferentes condições financeiras, democráticas, etc., mantendo as demais características fixas, e observar o que ocorre com as estimativas das probabilidades de coalizão. O BART, através da utilização de um algoritmo de estimação baseado em simulações, retorna valores preditos *caso o país fosse submetido a uma outra situação, mantendo as outras características constantes*. O cálculo da diferença entre o que foi estimado no universo real e o que foi previsto no universo de teste representa o “efeito causal” daquela mudança.

Uma análise desse tipo, contudo, possui algumas suposições. A mais importante delas assume que, dadas as informações de um país, a probabilidade deste país assumir valores distintos de alguma das variáveis em que se deseja avaliar o “efeito causal” é positiva e menor que 100%. Assim, deve-se assumir que, para um país que possui logaritmo do PIB alto, a probabilidade deste assumir um logaritmo do PIB igual a um certo valor de referência, mantendo-se as outras características intactas, deve ser diferente de zero e diferente de um.

Para realizar esta análise, o valor de referência utilizado foi o logaritmo do PIB médio. Na Figura C.2.10, apresenta-se, em vermelho, os valores ajustados da probabilidade de coalizão do primeiro critério de coalizão, ordenados pelo logaritmo do

PIB, enquanto que, em preto, são apresentados os mesmos valores preditos, caso o país tivesse o logaritmo do PIB igual o logaritmo do PIB médio, mantendo-se as outras características constantes. Na Figura C.2.11, é apresentada a diferença dessas quantidades para cada país, com os respectivos intervalos de credibilidade (com probabilidades de 90% e 95%). Nas Figuras C.2.12 e C.2.13, os mesmos gráficos foram obtidos para o segundo critério de coalizão considerado.

Observando os gráficos de efeitos causais e os intervalos de confiança, pode-se interpretar que os países mais pobres (em relação ao logaritmo do PIB real), se *possuíssem* o logaritmo do PIB igual ao mediano, mantendo suas características regionais e de democracia, entrariam mais em coalizões do que de fato o fazem. Nas mesmas condições, países mais ricos (também sob o mesmo critério) entrariam menos em coalizão se *tivessem* o logaritmo do PIB igual ao logaritmo do PIB médio.

9 Conclusões

A análise descritiva e exploratória descrita ao longo deste texto, juntamente com a pergunta a ser respondida, norteou o estudo de forma a resumir o problema longitudinal em uma abordagem pontual, baseada no uso de medidas-resumo. Após a realização de diversos enfoques para atacar os mesmos objetivos, pode-se dizer que os resultados, de maneira geral, são coincidentes.

Sobre estabelecer relações de causalidade no sentido de que características de um país afetem o seu envolvimento em coalizões, técnicas um pouco mais sofisticadas de inferência causal em sintonia com embasamentos teóricos do pesquisador permitiram a construção de argumentos a favor desta relação de causa e efeito desejada. Apesar disso, as conclusões tomadas são apenas parciais, mostrando que ainda existe muita discussão a respeito deste tópico.

Em resumo, os aspectos mais relevantes obtidos após a análise estatística de dados foram:

- Quanto maior o PIB de um país, maior é a probabilidade de ele entrar em coalizão;
- Dentre países com índice de democracia acumulado maior que 12,5, a probabilidade estimada de entrar em coalizão cresce com o aumento do índice de democracia. Para os demais países, esta probabilidade estimada não se altera com a variação do índice.
- Os dois critérios para definir coalizão utilizados apontam para resultados semelhantes;
- As 4 regiões do globo (hemisférios Nordeste, Noroeste, Sudoeste e Sudeste) se manifestam de maneira distinta quanto às entradas em coalizões. Países do sudoeste têm a maior probabilidade estimada de entrarem em coalizão enquanto a região nordeste possui a menor. O efeito do logaritmo do PIB, contudo, permanece o mesmo em todas estas regiões.

10 Comentários

- *População.* O fato de a amostra representar completamente a população envolvida no estudo pode ser um problema para diversas técnicas estatísticas empregadas que, geralmente, admitem que as populações têm tamanhos infinitos. Isso faz com que possa ser necessária a utilização de técnicas de estatística de populações finitas. Entretanto, tais fatores tornam-se irrelevantes para as análises descritivas, o que, com abuso de linguagem, possibilitou o tratamento do censo como uma amostra.
- *Causalidade.* A utilização de variáveis instrumentais para possibilitar o estabelecimento de relações causais depende de diversas suposições. Primeiro, deve-se supor que as variáveis de controle que não são ajustadas utilizando-se variáveis instrumentais não possuem endogeneidade. Além disso, deve-se supor que, num sistema completamente controlado é possível atribuir causalidade às associações porventura obtidas. Só assim faz-se válida a ideia da utilização da variável instrumental como tentativa de retirar possíveis influências externas, tornando o estudo aproximadamente controlado. Por esses motivos, as

afirmações sobre causalidade do presente estudo, no estado atual, ainda são bastante inconclusivas e depende, necessariamente, de suposições do pesquisador.

- *Gráficos de efeitos marginais conjuntos.* É importante ressaltar que as cores dos gráficos apresentados são baseadas em percentis das probabilidades estimadas, o que pode mascarar algumas informações dos países individualmente. Além disso, a construção destes gráficos para todos os países é inviável, pois requer um custo computacional muito acima do poder das máquinas que foram utilizadas para ajuste dos modelos.
- Um modelo BART considerando a estrutura em painel foi ajustado e os resultados, como os efeitos marginais, foram concordantes com aqueles observados no modelo principal. Isso indica que a metodologia utilizada para agregar os dados é justificada.

APÊNDICE A

Figuras e tabelas da análise descritiva em painel

A Figuras e tabelas da análise descritiva em painel

A.1 Figuras

Instrução para interpretação das Figuras de **A.1.7** até **A.1.22**:

Os gráficos apresentam todos os países, representados por um único perfil individual que descreve uma curva no tempo (eixo x). Os perfis estão agrupados por região e cada um dos 10 gráficos mostra o comportamento de uma mesma característica.

No canto superior direito de cada um desses 10 gráficos encontra-se um valor, identificado pela letra n, que indica o número total de países participantes da GATT/OMC daquela região no ano de 2006 (o último ano em estudo).

Em segundo plano nota-se a presença de uma área sombreada. Esta área indica a proporção de países presentes até a data correspondente em relação ao total de países presentes em 2006 (sendo esta última o valor n).

O perfil de cor preta representa o perfil médio dos países da respectiva região e a linha vertical em vermelho é uma referência para o ano 1986.

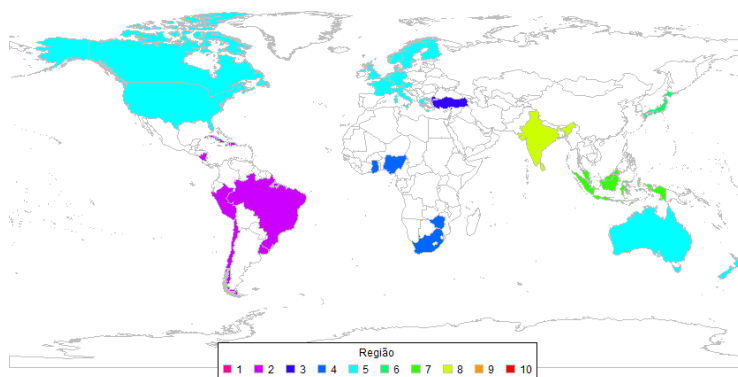


Figura A.1.1 Mapa dos países participantes no ano de 1960.

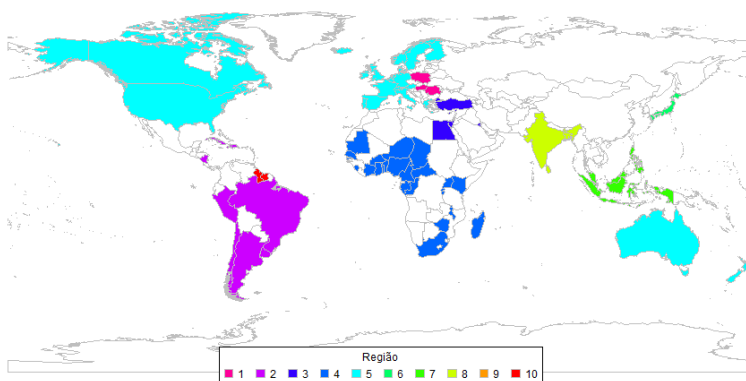


Figura A.1.2 Mapa dos países participantes no ano de 1980.

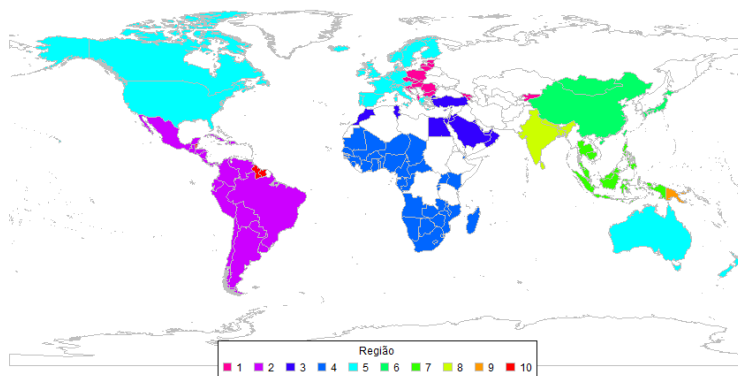


Figura A.1.3 Mapa dos países participantes no ano de 2006.

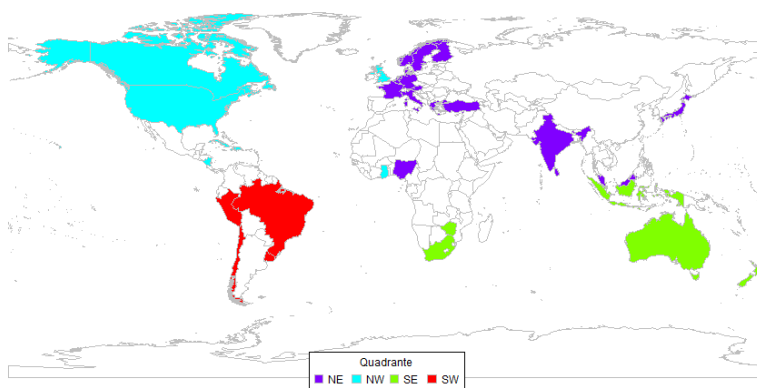


Figura A.1.4 Mapa dos países participantes no ano de 1960.

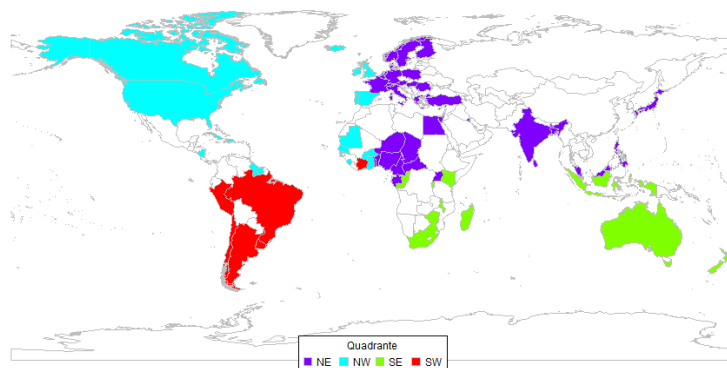


Figura A.1.5 Mapa dos países participantes no ano de 1980.

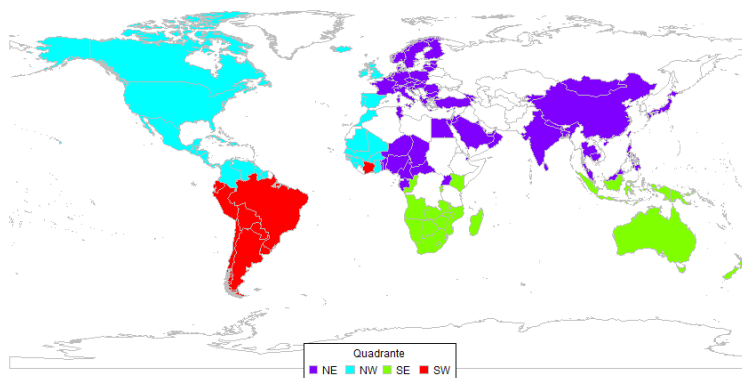


Figura A.1.6 Mapa dos países participantes no ano de 2006.

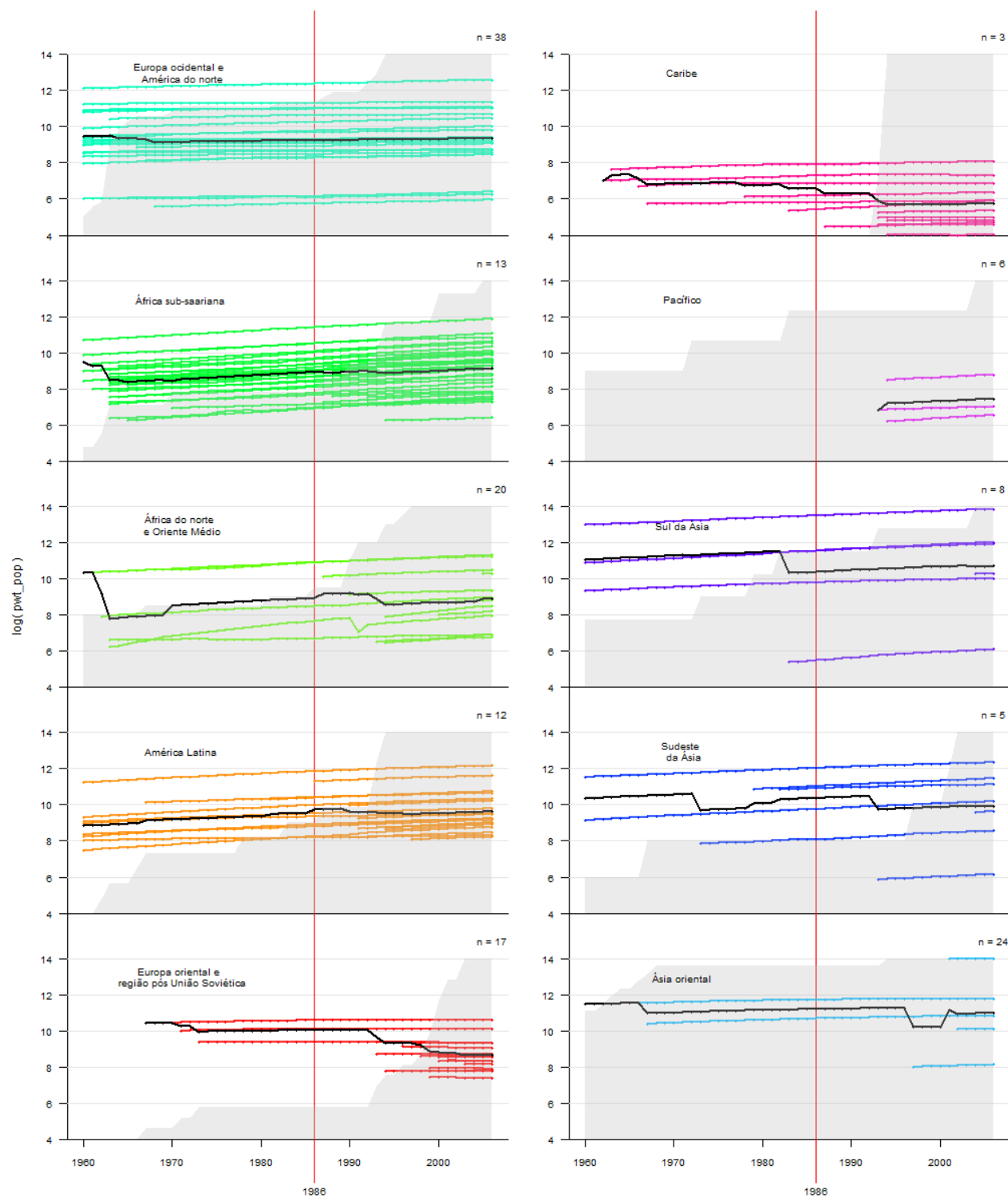


Figura A.1.7 Perfis individuais do logaritmo do tamanho da população (em milhares).

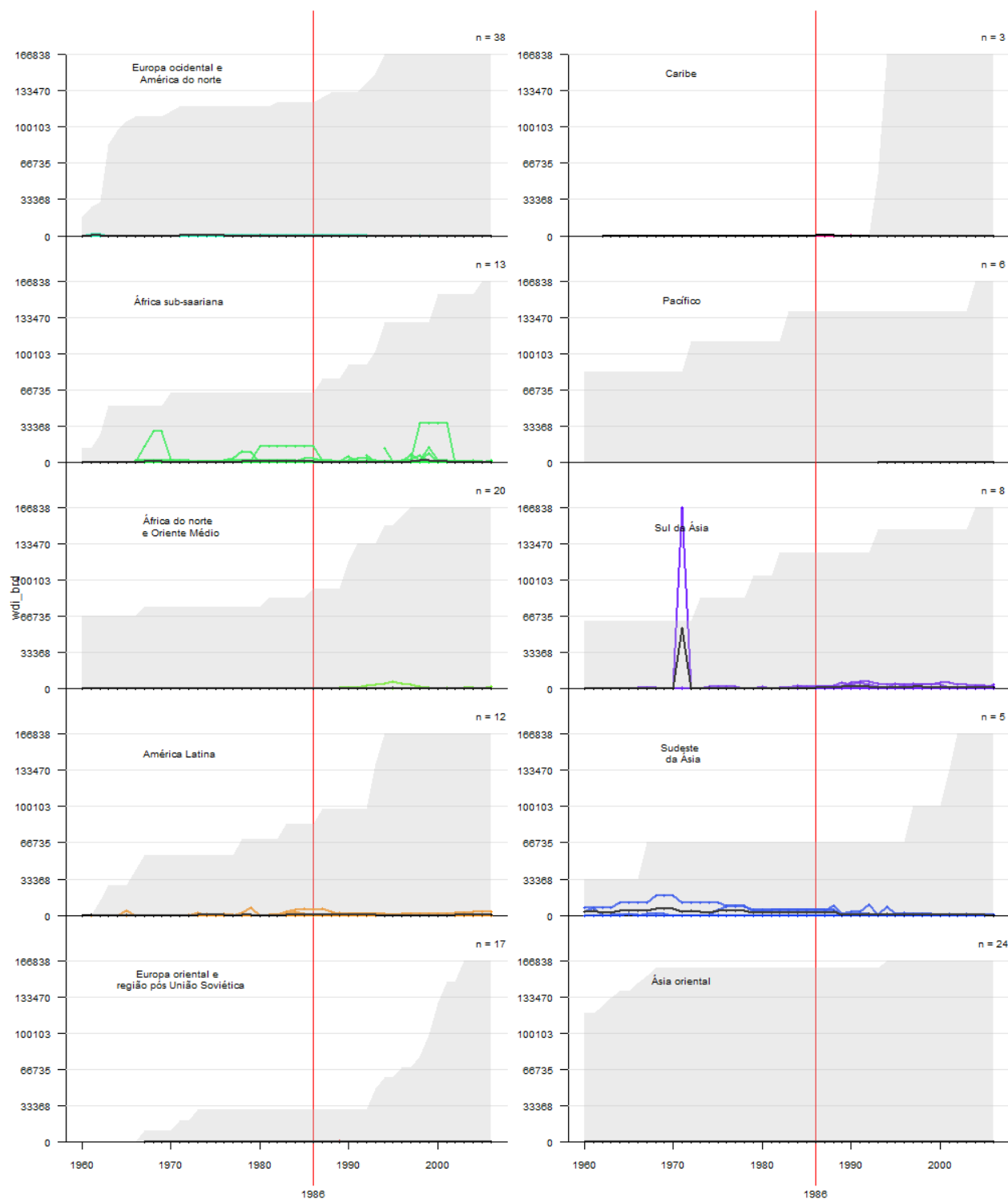


Figura A.1.8 Perfis individuais do número de mortes relacionadas a batalhas.

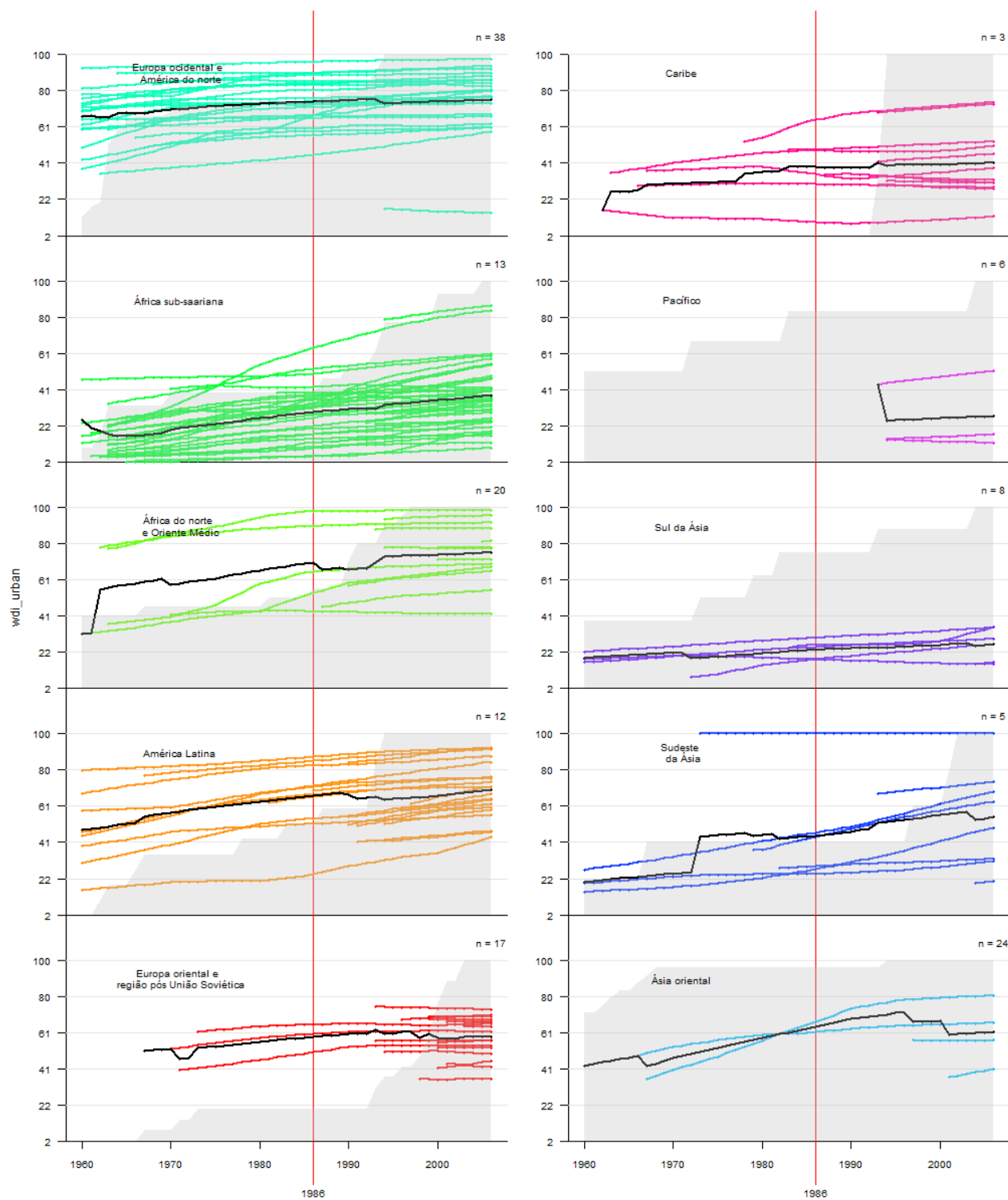


Figura A.1.9 Perfis individuais da proporção da população total vivendo em áreas urbanas (em porcentagem).

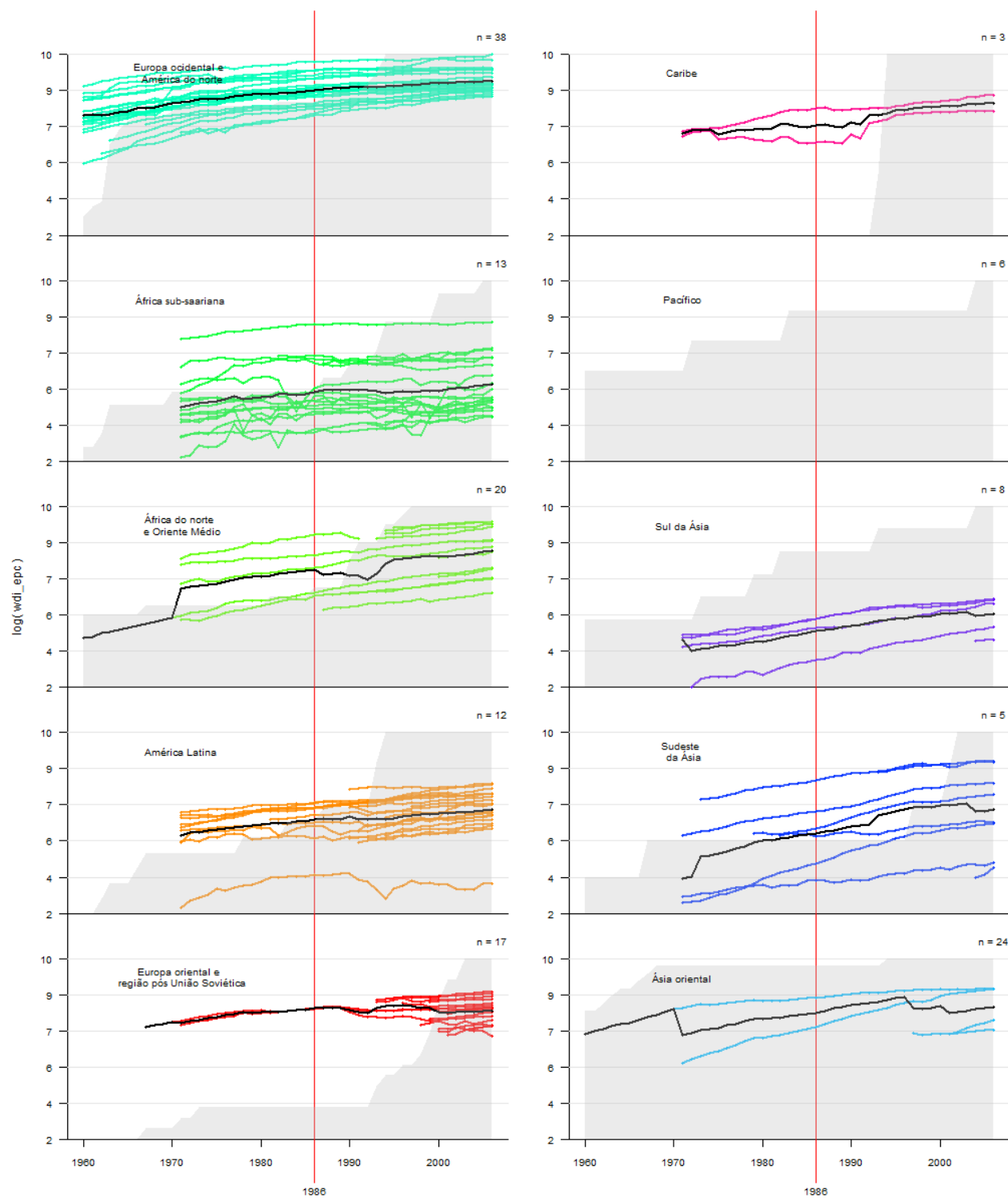


Figura A.1.10 Perfis individuais do logaritmo do consumo de energia elétrica per (em KWh per capita).

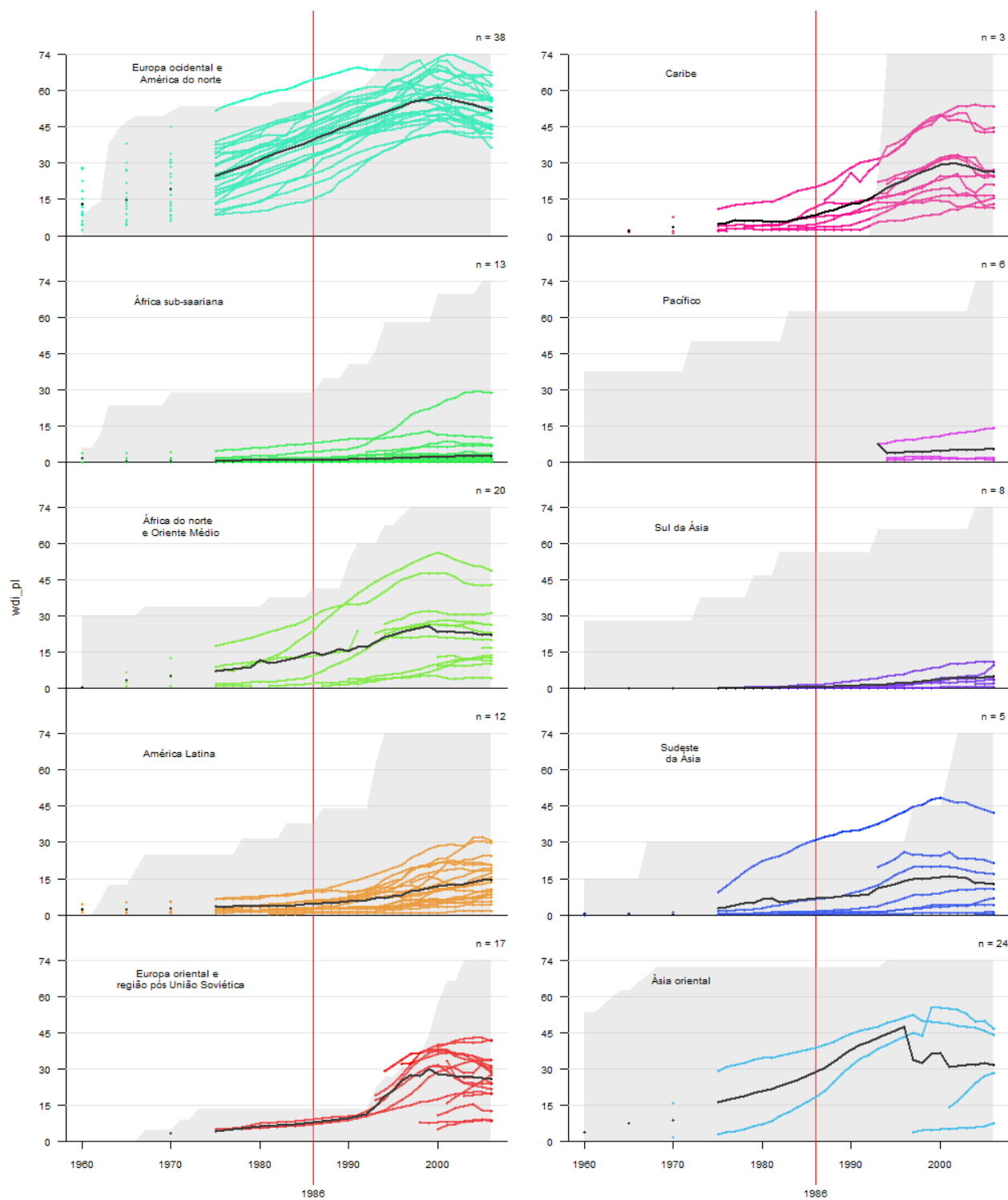


Figura A.1.11 Perfis individuais do número de ligações telefônicas em cada 100 habitantes.

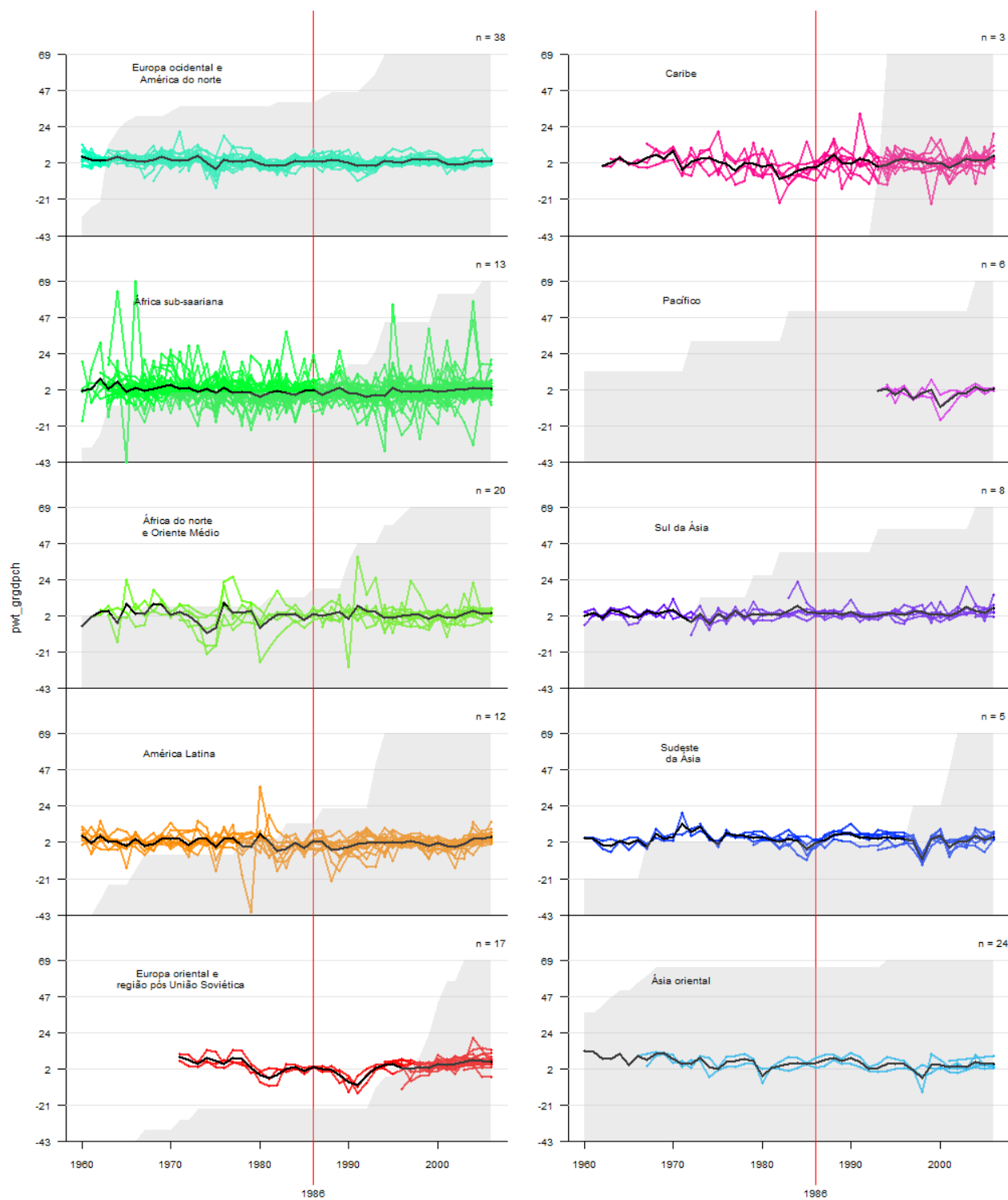


Figura A.1.12 Perfis individuais da taxa de crescimento do PIB real (taxa a preços constantes).

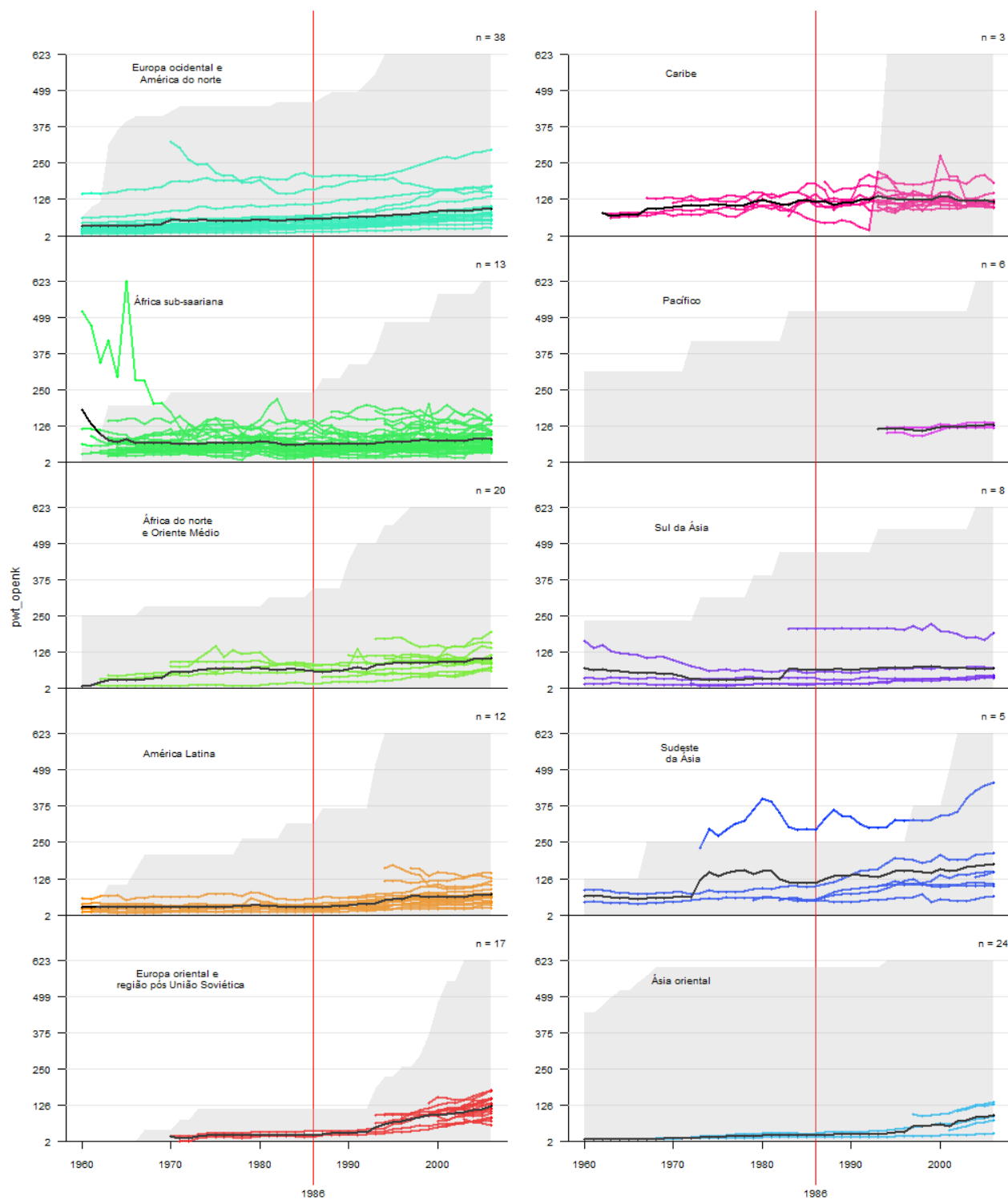


Figura A.1.13 Perfis individuais da abertura para o comércio (em porcentagem do PIB real).

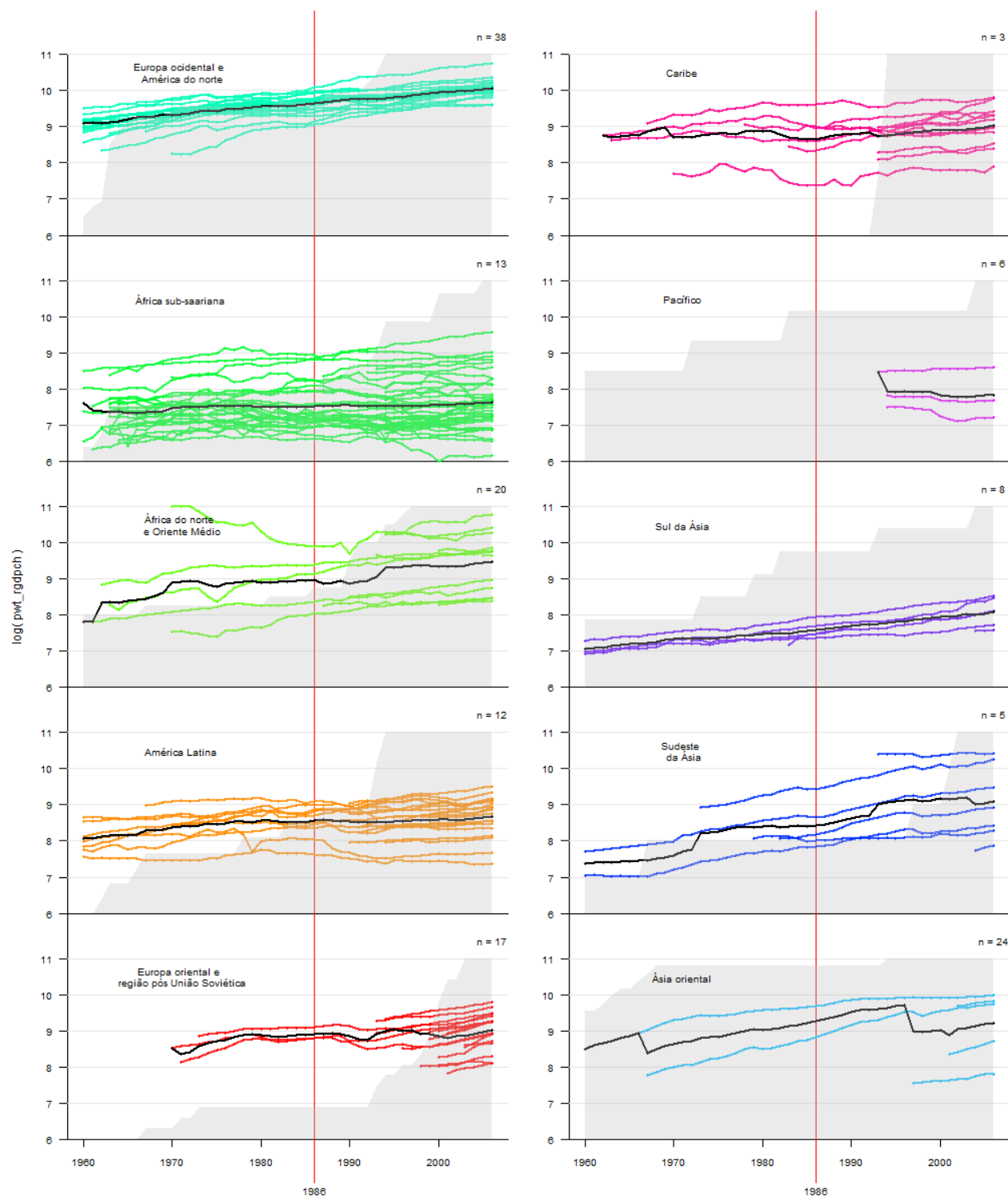


Figura A.1.14 Perfis individuais do logaritmo do PIB per capita.

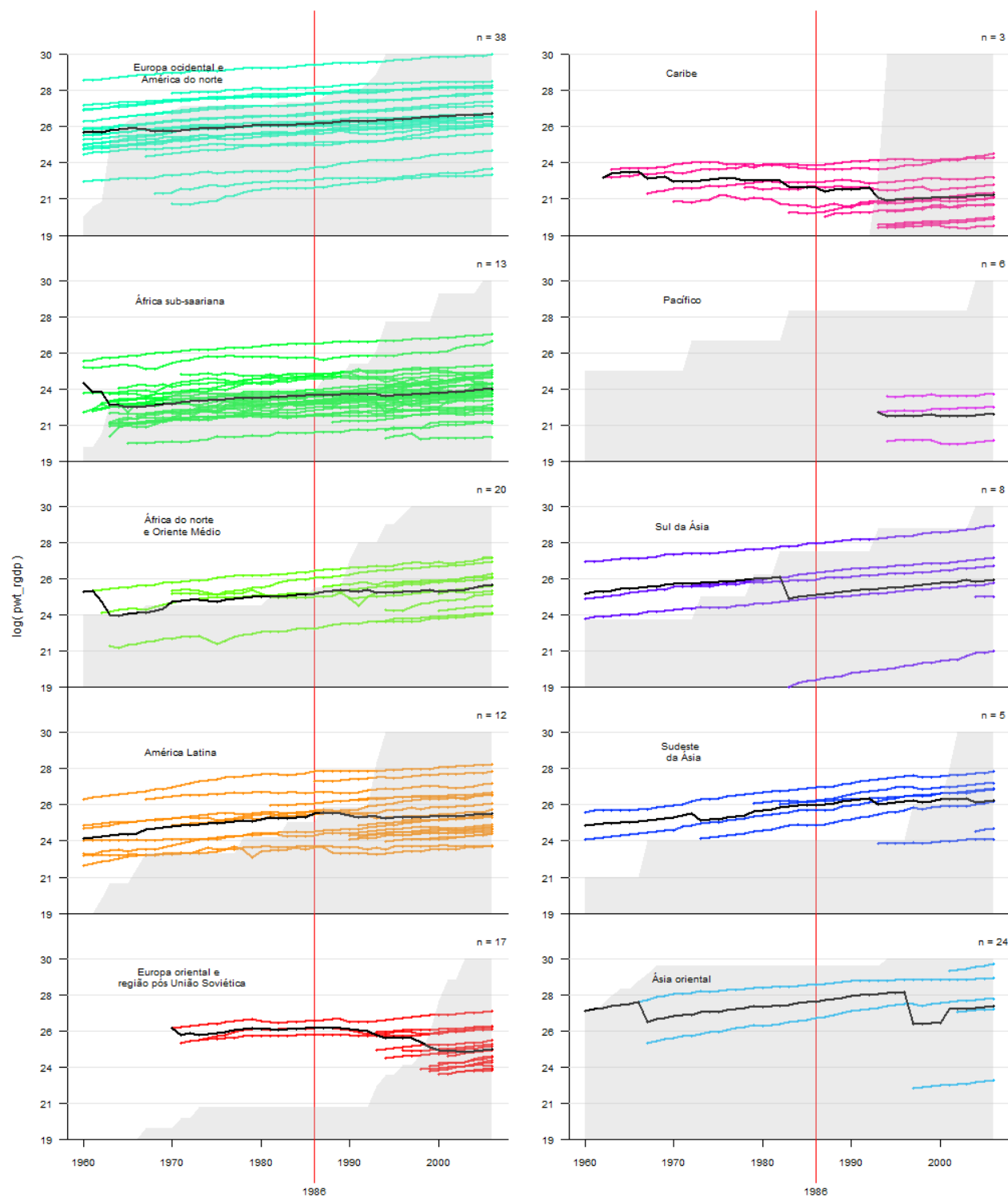


Figura A.1.15 Perfis individuais do logaritmo do PIB real.

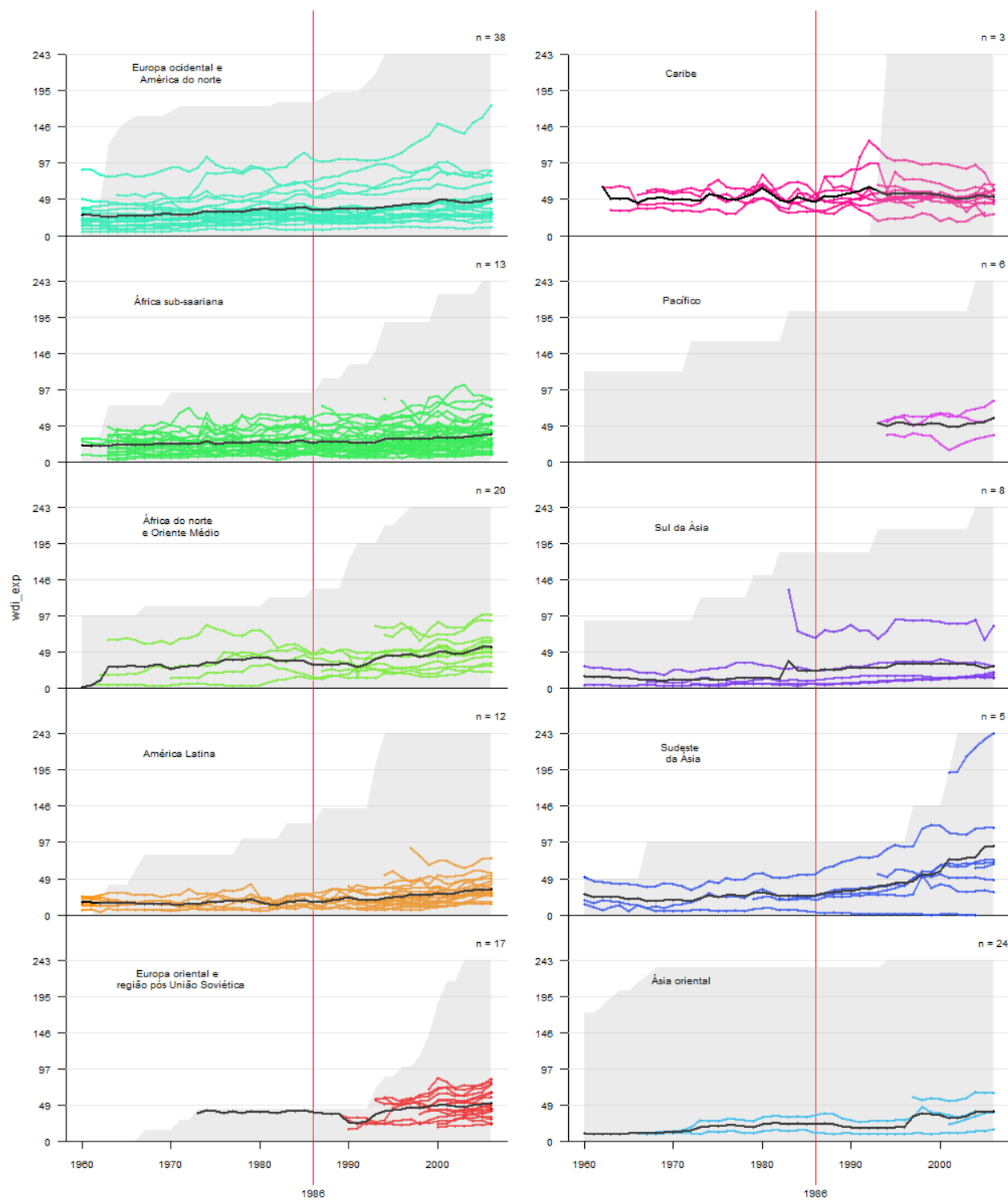


Figura A.1.16 Perfis individuais da exportação (em porcentagem do PIB).

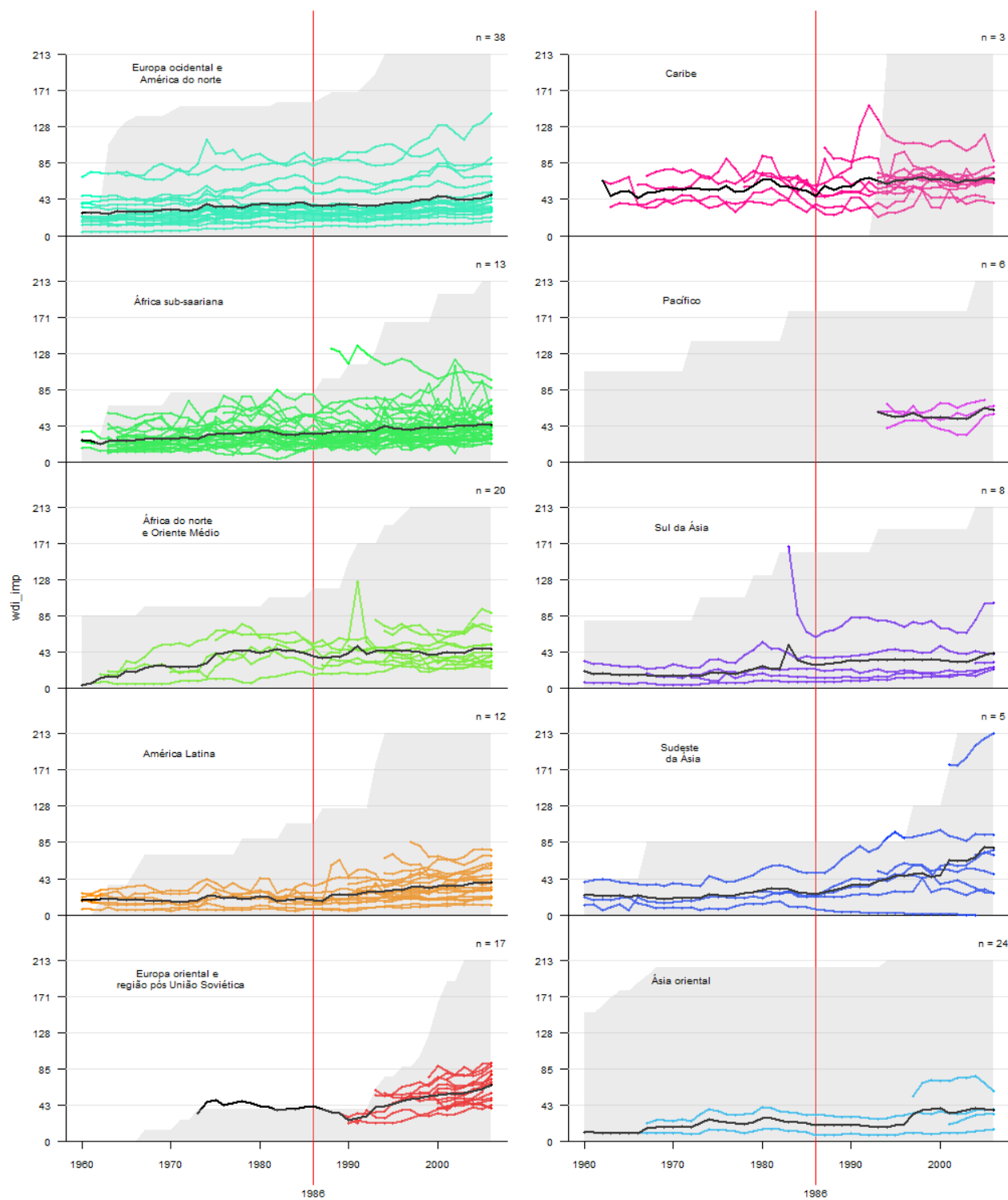


Figura A.1.17 Perfis individuais da importação (em porcentagem do PIB).

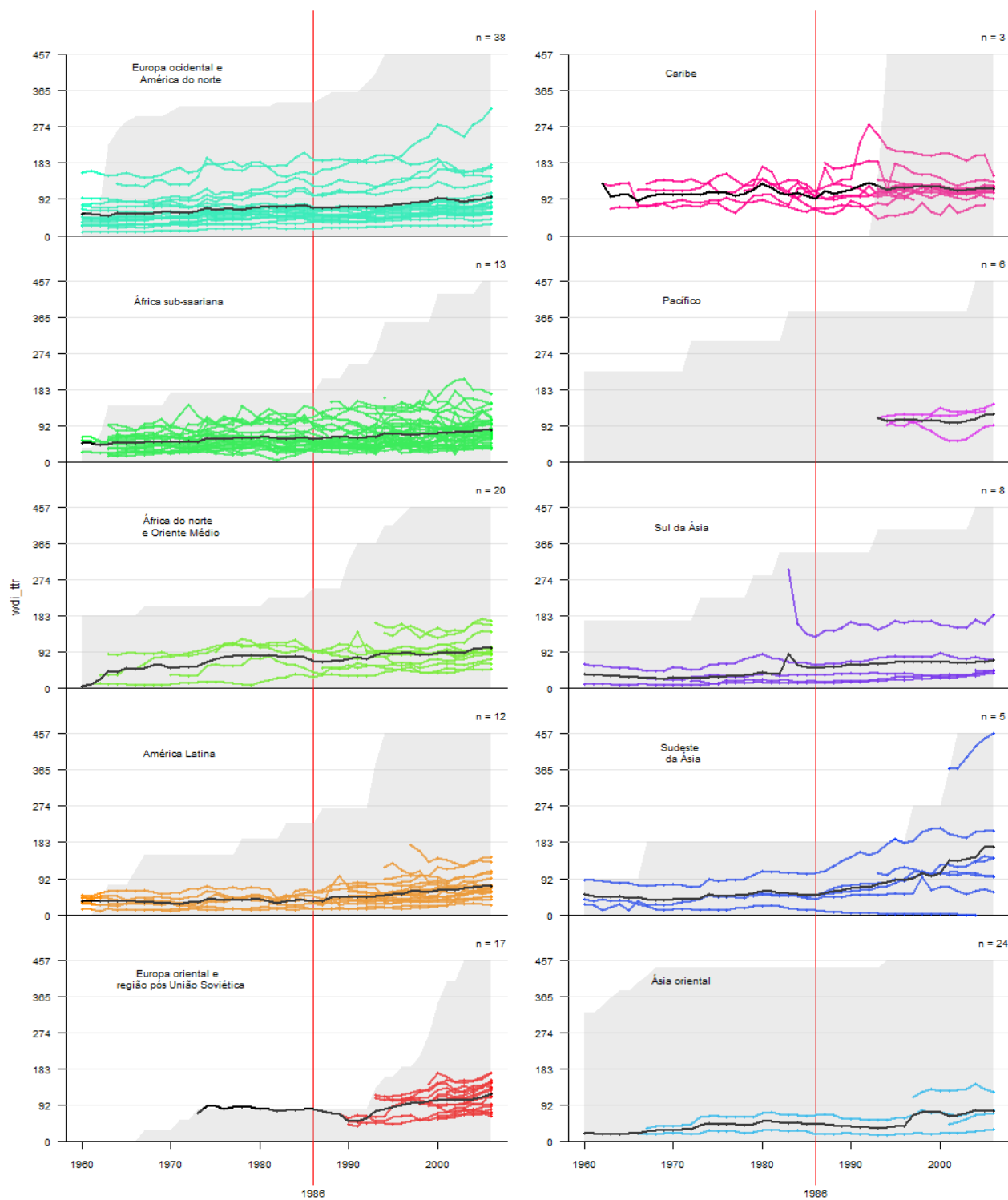


Figura A.1.18 Perfis individuais do comércio total (em porcentagem do PIB).

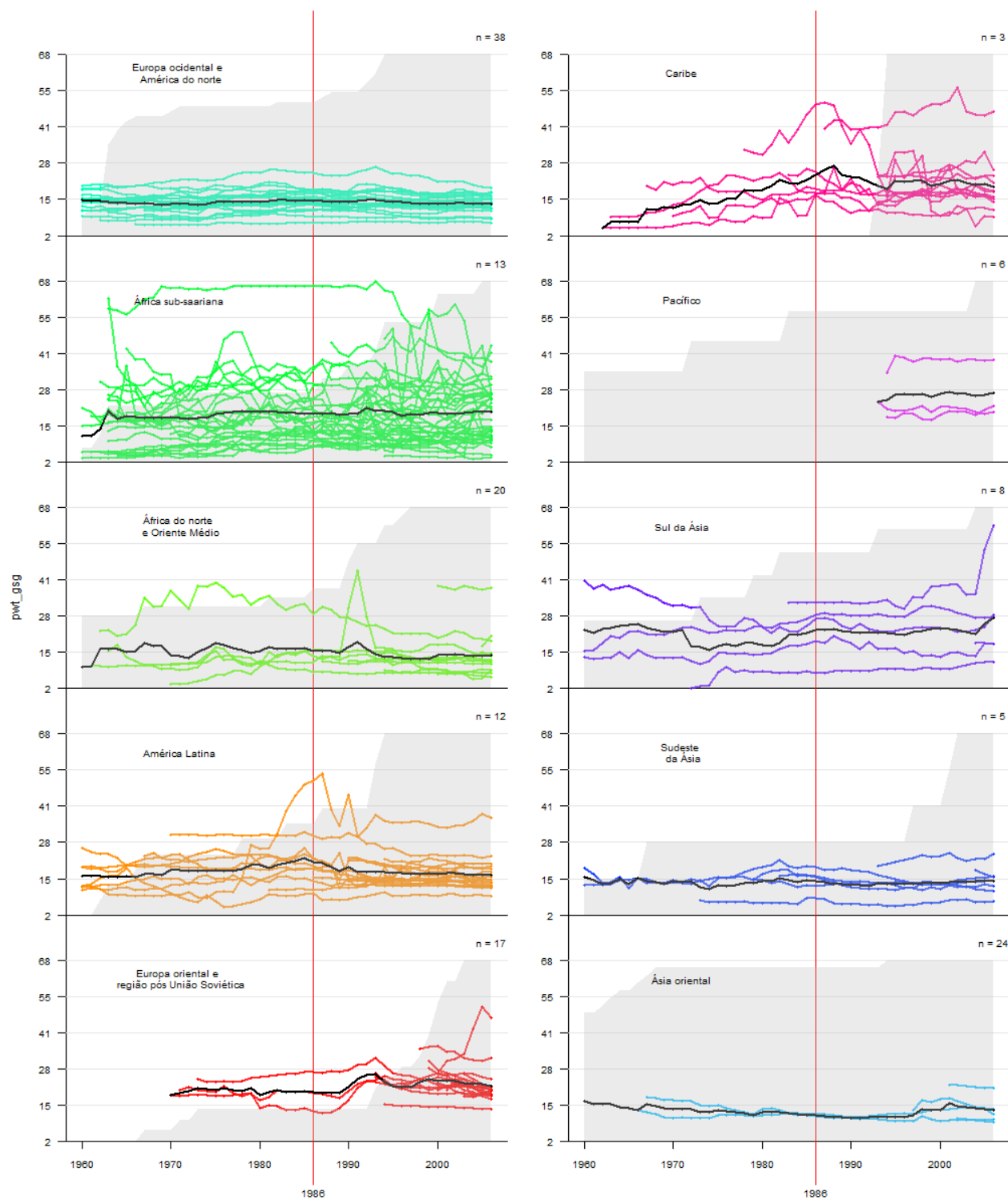


Figura A.1.19 Perfis individuais da participação do governo no PIB (em porcentagem).

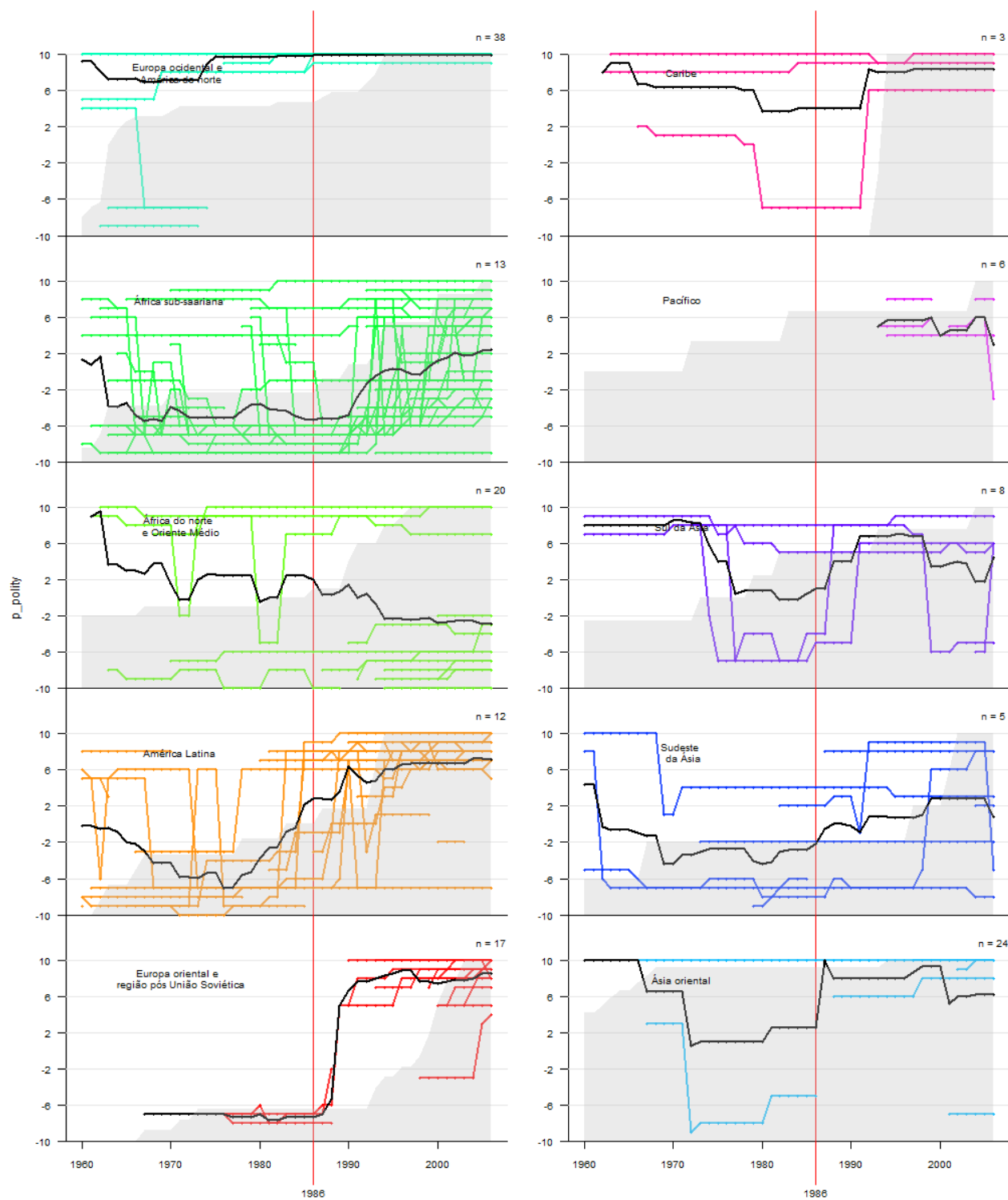


Figura A.1.20 Perfis individuais do escore de democracia.

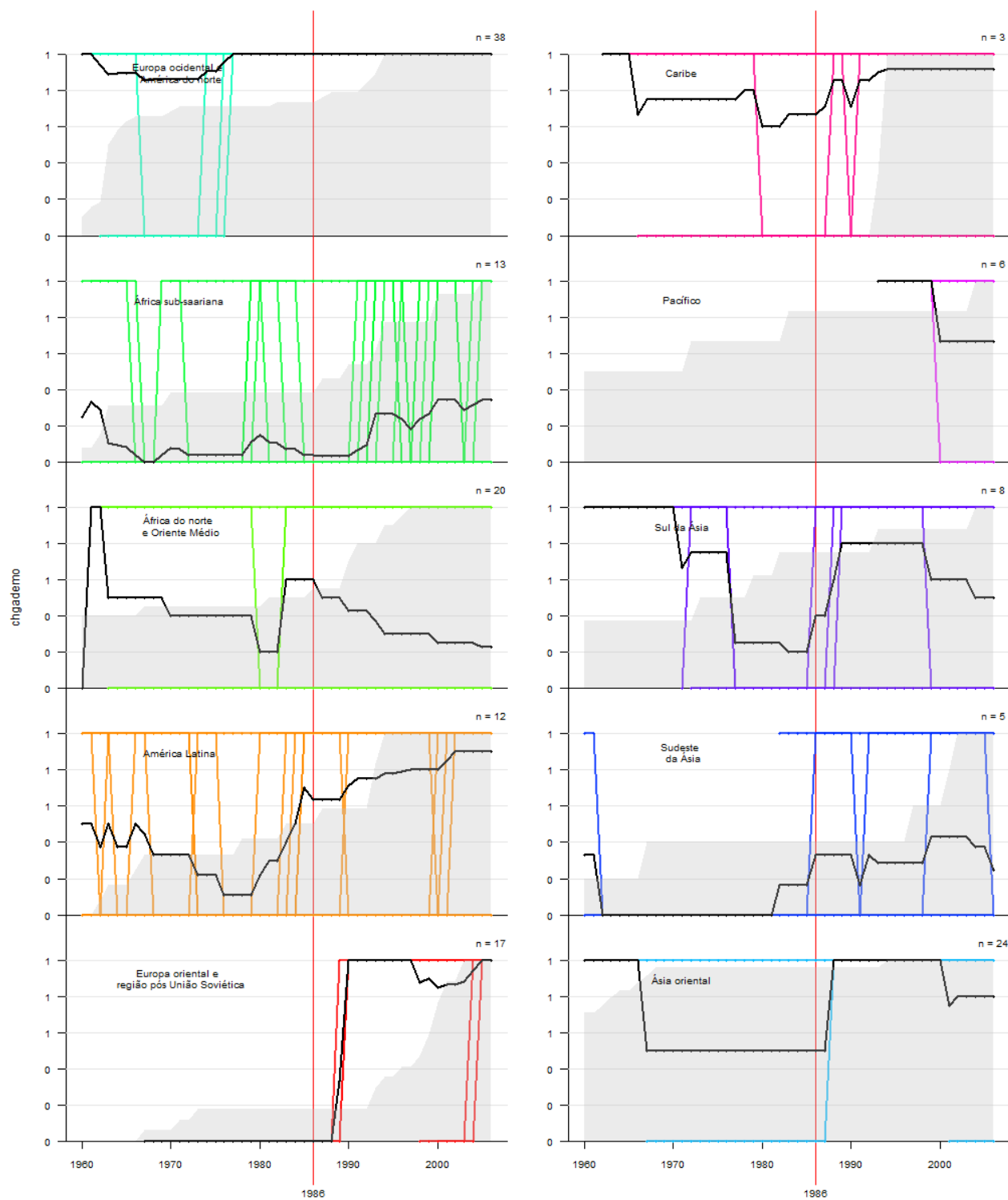


Figura A.1.21 Perfis individuais do indicador de democracia.

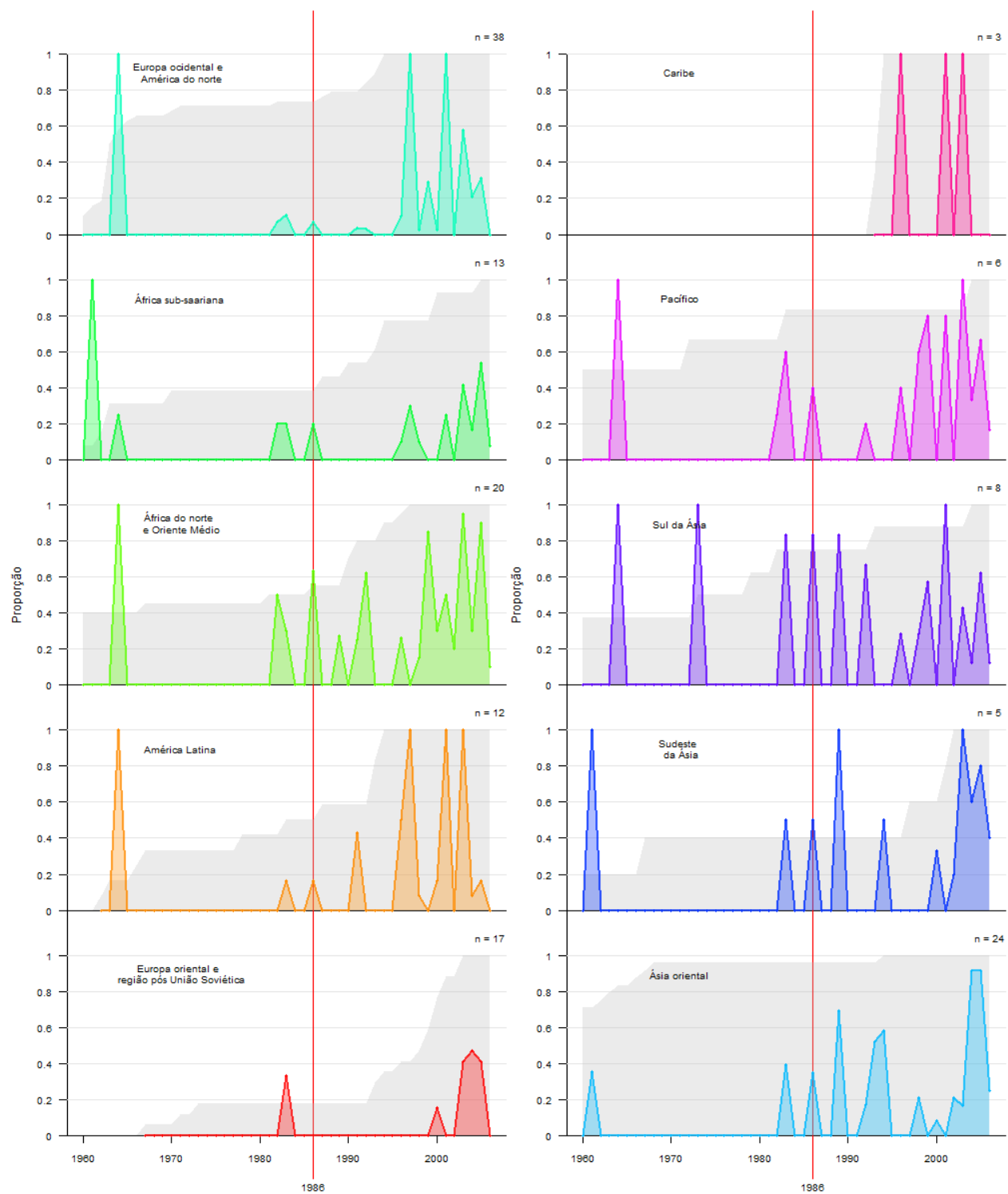


Figura A.1.22 Perfis individuais da proporção de coalizões dentro da região.

A.2 Tabelas

Tabela A.2.1 Frequência absoluta e relativa de países nas regiões por ano.

Ano	Região										Total
	1	2	3	4	5	6	7	8	9	10	
1960	0	8	1	4	17	1	3	3	0	0	37
	0%	22%	3%	11%	46%	3%	8%	8%	0%	0%	100%
1980	3	9	5	27	23	2	5	4	0	5	83
	4%	11%	6%	33%	28%	2%	6%	5%	0%	6%	100%
2006	17	20	13	38	24	5	8	6	3	12	146
	12%	14%	9%	26%	16%	3%	5%	4%	2%	8%	100%

Tabela A.2.2 Frequência absoluta e relativa de países nos quadrantes em 1960.

	W		E		Total	
N	8	22%	20	54%	28	76%
S	4	11%	5	14%	9	24%
Total	12	32%	25	68%	37	100%

Tabela A.2.3 Frequência absoluta e relativa de países nos quadrantes em 1980.

	W		E		Total	
N	21	25%	41	49%	62	75%
S	6	7%	15	18%	21	25%
Total	27	33%	56	67%	83	100%

Tabela A.2.4 Frequência absoluta e relativa de países nos quadrantes em 2006.

	W		E		Total	
N	40	27%	72	49%	112	77%
S	9	6%	25	17%	34	23%
Total	49	34%	97	66%	146	100%

Tabela A.2.5 Frequência absoluta e relativa de países quanto à posse de terras entre os trópicos.

Ano	Tem terra entre os trópicos		
	Não	Sim	Total
1960	19	18	37
	51%	49%	100%
1980	32	51	83
	39%	61%	100%
2006	56	90	146
	38%	62%	100%

Tabela A.2.6 Frequência absoluta e relativa de países quanto à posse de terras entre os trópicos.

Ano	Proporção de terras entre os trópicos		Total
	< 50%	> 50%	
1960	4	14	18
	22%	78%	100%
1980	7	44	51
	14%	86%	100%
2006	12	78	90
	13%	87%	100%

Tabela A.2.7 Nomes dos países participantes do GATT/OMC em 2006 (em inglês)

Albania	Djibouti	Latvia	Qatar
Angola	Dominica	Lesotho	Romania
Antigua and Barbuda	Dominican Republic	Liechtenstein	Rwanda
Argentina	Ecuador	Lithuania	Saudi Arabia
Armenia	Egypt	Luxembourg	Senegal
Australia	El Salvador	Macedonia	Sierra Leone
Austria	Estonia	Madagascar	Singapore
Bahrain	Fiji	Malawi	Slovakia
Bangladesh	Finland	Malaysia	Slovenia
Barbados	France	Maldives	Solomon Islands
Belgium	Gabon	Mali	South Africa
Belize	Gambia	Malta	Spain
Benin	Georgia	Mauritania	Sri Lanka
Bolivia	Germany	Mauritius	St Kitts and Nevis
Botswana	Ghana	Mexico	St Lucia
Brazil	Greece	Moldova	St Vincent and the Grenadines
Brunei	Grenada	Mongolia	Suriname
Bulgaria	Guatemala	Morocco	Swaziland
Burkina Faso	Guinea	Mozambique	Sweden
Burundi	Guinea-Bissau	Myanmar	Switzerland
Cambodia	Guyana	Namibia	Taiwan
Cameroon	Haiti	Nepal	Tanzania
Canada	Honduras	Netherlands	Thailand
Central African Republic	Hungary	New Zealand	Togo
Chad	Iceland	Nicaragua	Trinidad and Tobago
Chile	India	Niger	Tunisia
China	Indonesia	Nigeria	Turkey
Colombia	Ireland	Norway	Uganda
Congo	Israel	Oman	United Arab Emirates
Congo, Democratic Republic	Italy	Pakistan (1972-)	United Kingdom
Costa Rica	Jamaica	Panama	United States
Cote d'Ivoire	Japan	Papua New Guinea	Uruguay
Croatia	Jordan	Paraguay	Venezuela
Cuba	Kenya	Peru	Zambia
Cyprus	Korea, South	Philippines	Zimbabwe
Czech Republic	Kuwait	Poland	
Denmark	Kyrgyzstan	Portugal	

APÊNDICE B

Figuras e tabelas da análise descritiva das medidas-resumo

B Figuras e tabelas da análise descritiva das medidas-resumo

B.1 Figuras

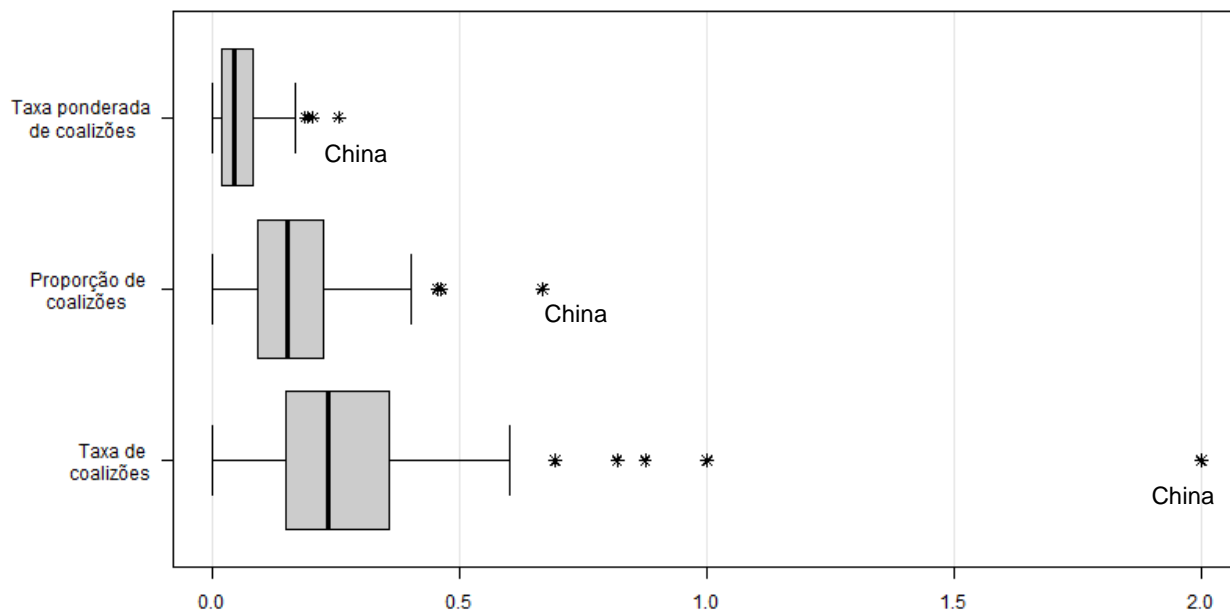


Figura B.1.1 Box-plots para as medidas-resumo das variáveis resposta baseados no primeiro critério de coalizão (entar).

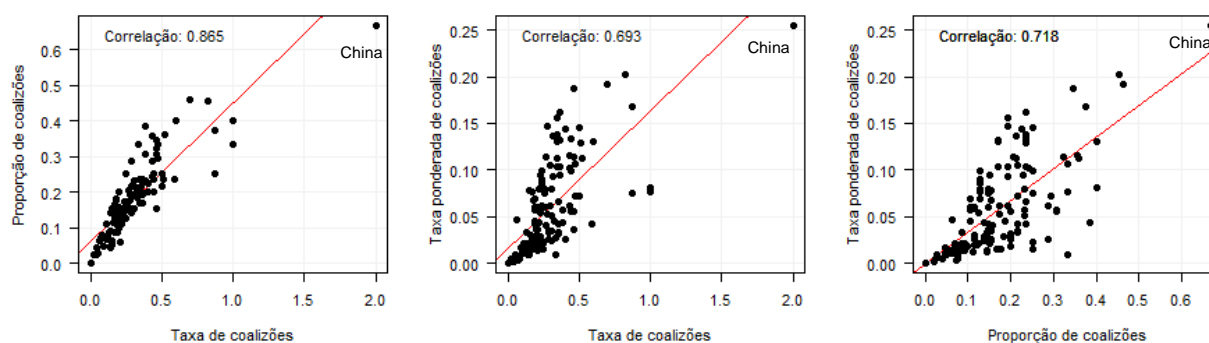


Figura B.1.2 Gráficos de dispersão entre as medidas-resumo das variáveis resposta baseados no primeiro critério de coalizão (entar).

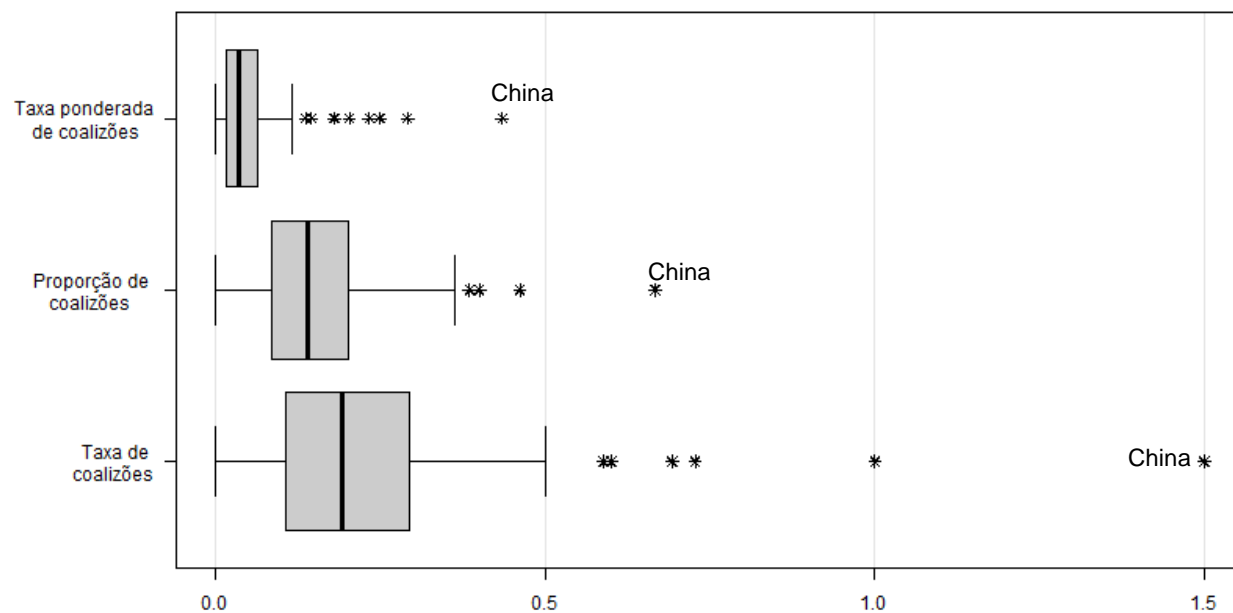


Figura B.1.3 *Box-plots* para as medidas-resumo das variáveis resposta baseados no segundo critério de coalizão (ensar).

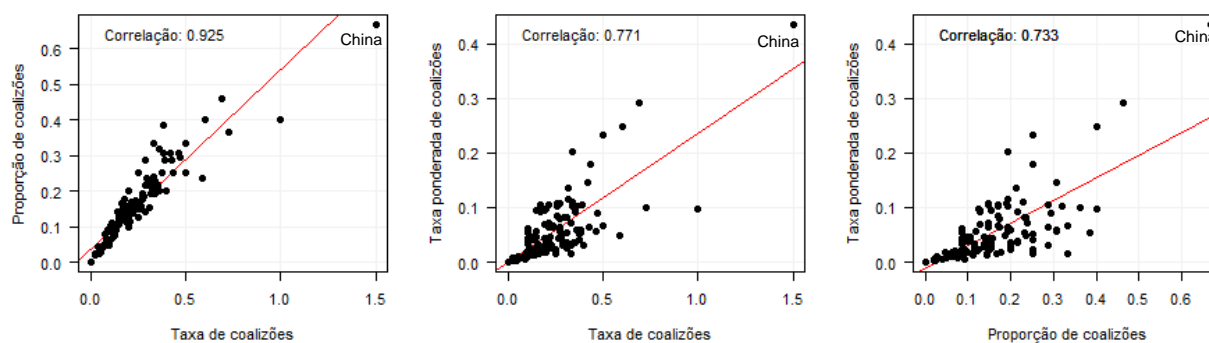


Figura B.1.4 Gráficos de dispersão entre as medidas-resumo das variáveis resposta baseados no segundo critério de coalizão (ensar).

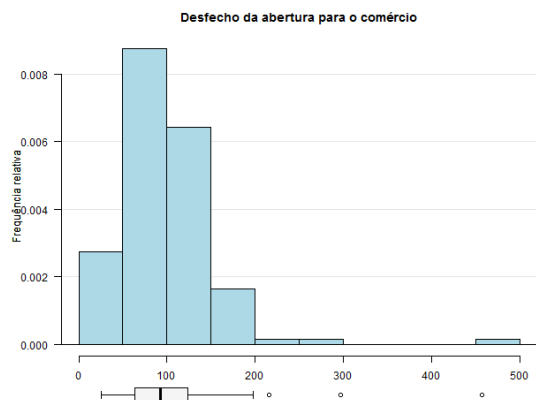


Figura B.1.5 Histograma do desfecho da abertura para o comércio.

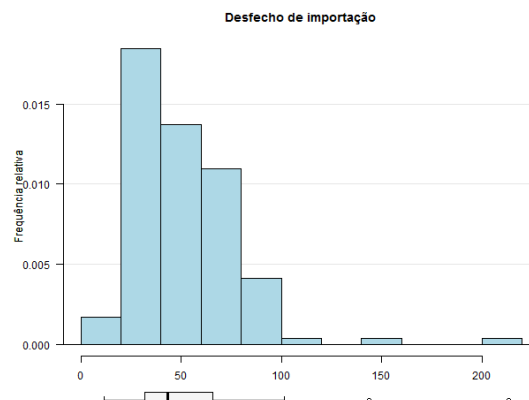


Figura B.1.8 Histograma do desfecho de importação.

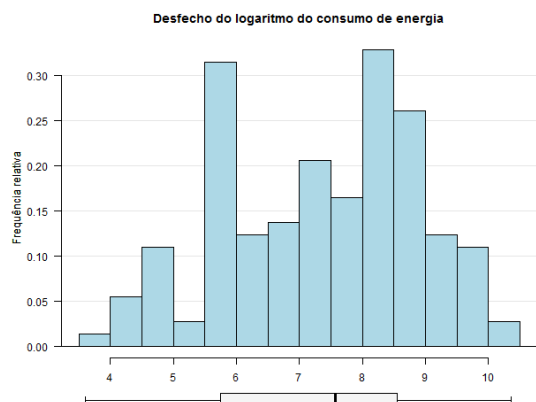


Figura B.1.6 Histograma do desfecho do logaritmo do consumo de energia.

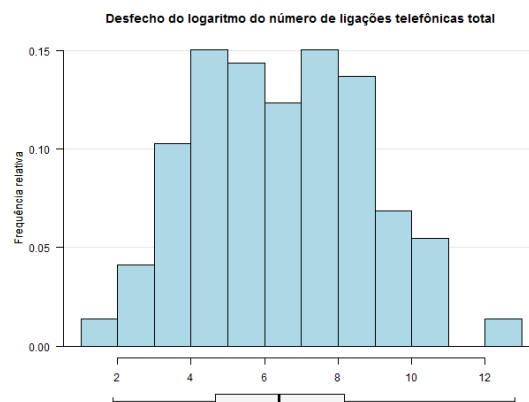


Figura B.1.9 Histograma do desfecho do logaritmo do número de ligações telefônicas total.

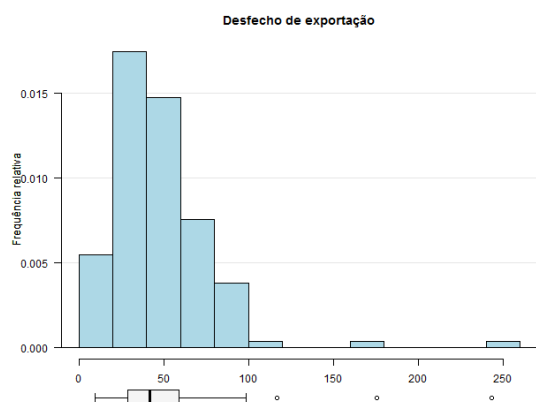


Figura B.1.7 Histograma do desfecho de exportação.

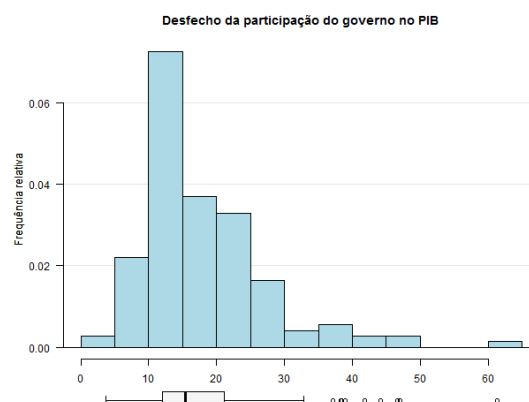


Figura B.1.10 Histograma do desfecho da participação do governo no PIB.

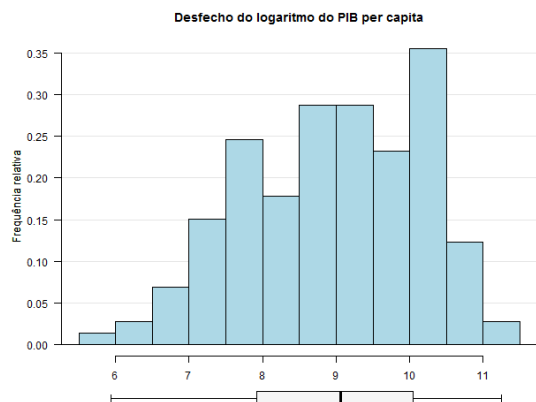


Figura B.1.11 Histograma do desfecho do logaritmo do PIB per capita.

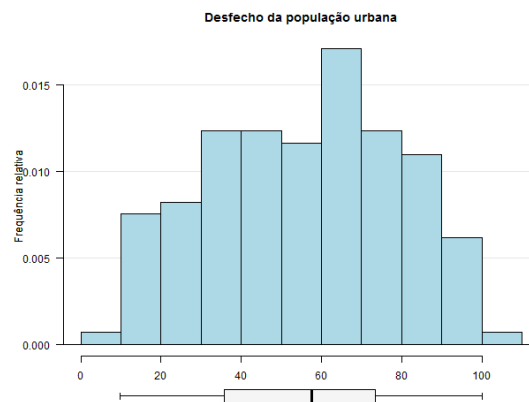


Figura B.1.14 Histograma do desfecho da população urbana.

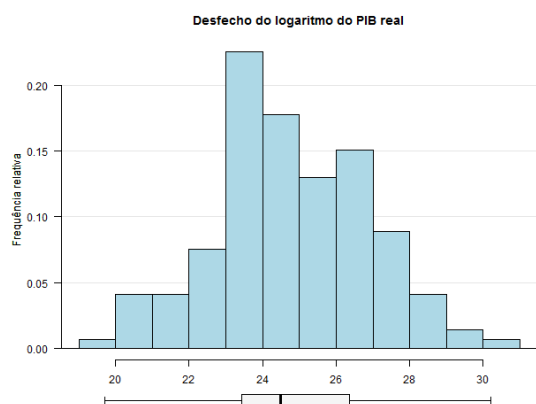


Figura B.1.12 Histograma do desfecho do logaritmo do PIB real.

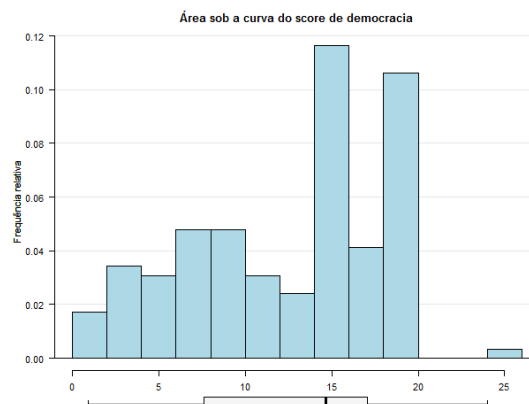


Figura B.1.15 Histograma da área sob a curva do escore de democracia.

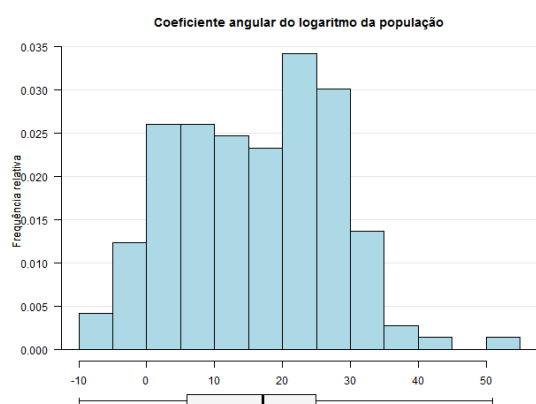


Figura B.1.13 Histograma do coeficiente angular do logaritmo da população.

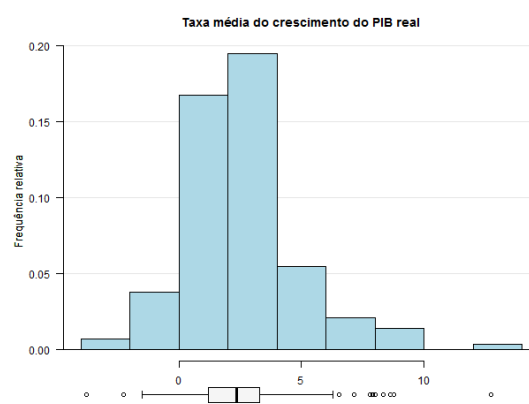


Figura B.1.16 Histograma da taxa média de crescimento do PIB real.

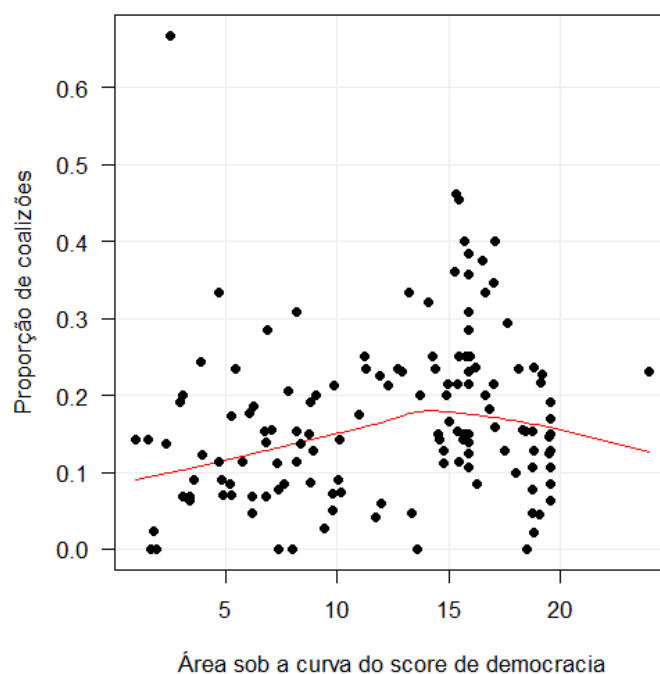


Figura B.1.17 Gráficos de dispersão entre a área sob a curva do escore de democracia e proporção de coalizões.

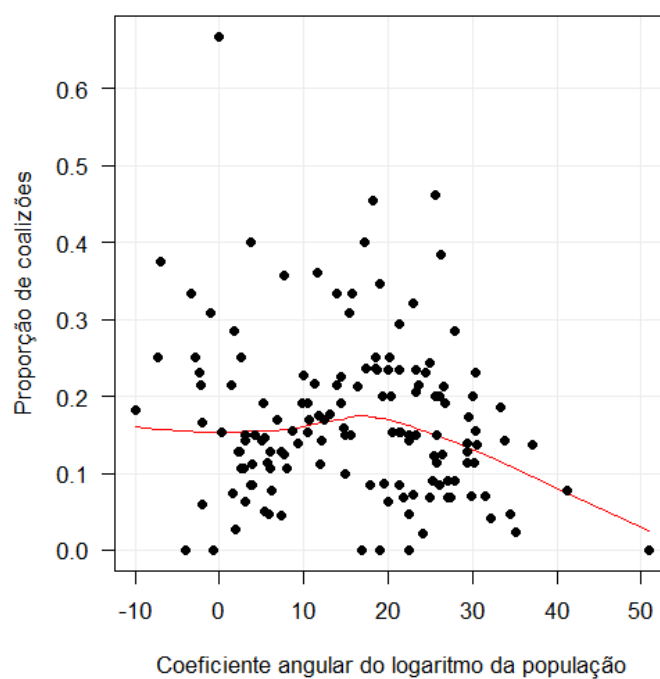


Figura B.1.18 Gráficos de dispersão entre o coeficiente angular do logaritmo da população e proporção de coalizões.

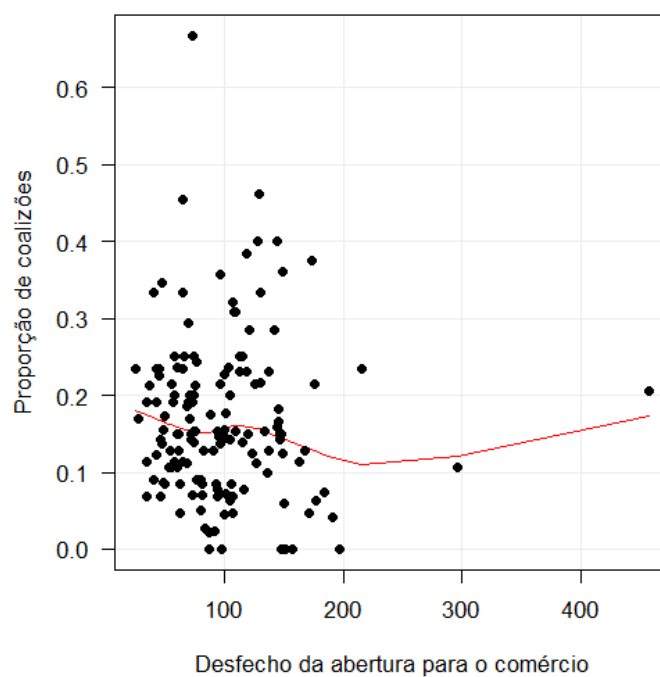


Figura B.1.19 Gráficos de dispersão entre o desfecho da abertura para o comércio e proporção de coalizões.

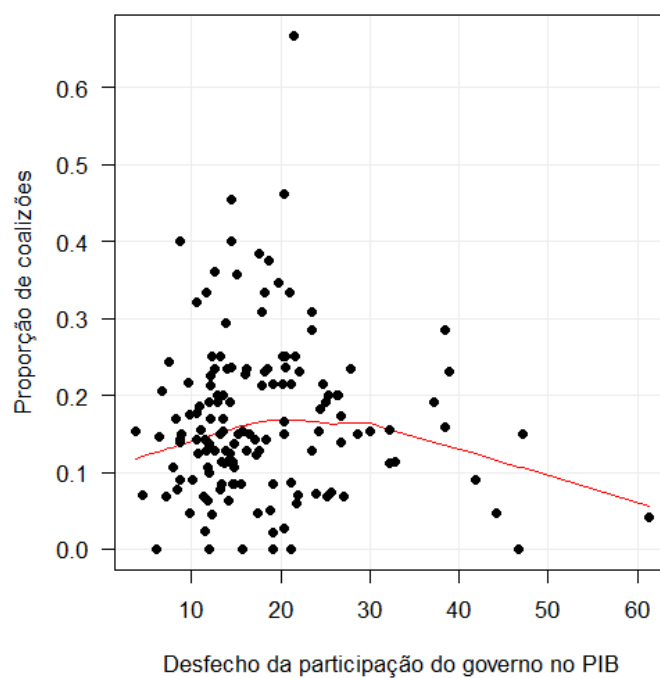


Figura B.1.20 Gráficos de dispersão entre o desfecho da participação do governo no PIB e proporção de coalizões.

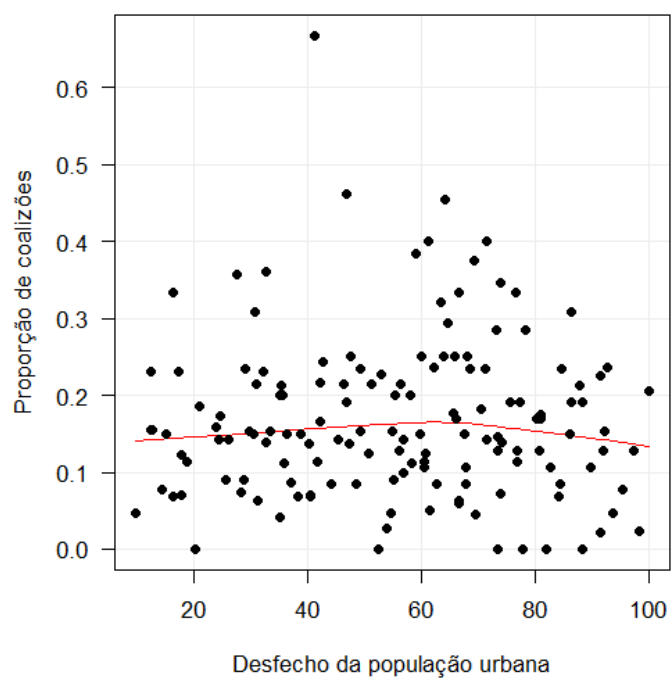


Figura B.1.21 Gráficos de dispersão entre o desfecho da população urbana e proporção de coalizões.

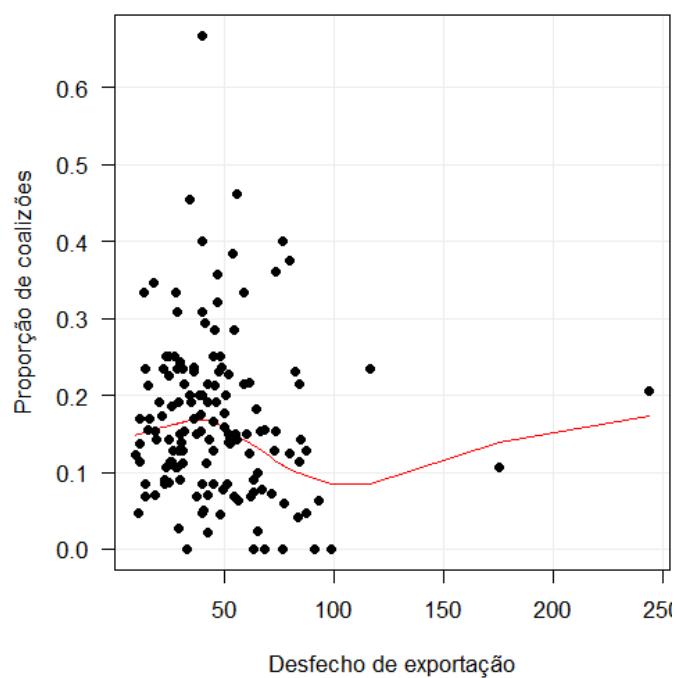


Figura B.1.22 Gráficos de dispersão entre o desfecho de exportação e proporção de coalizões.

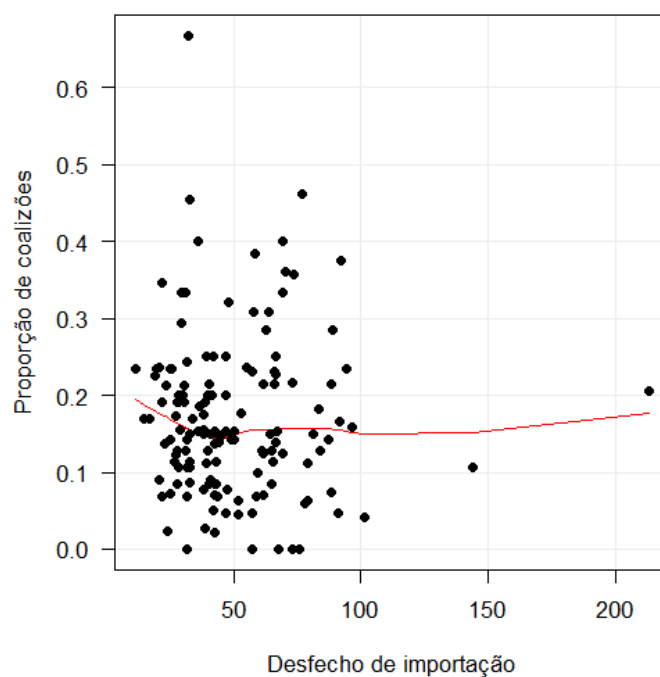


Figura B.1.23 Gráficos de dispersão entre o desfecho de importação e proporção de coalizões.

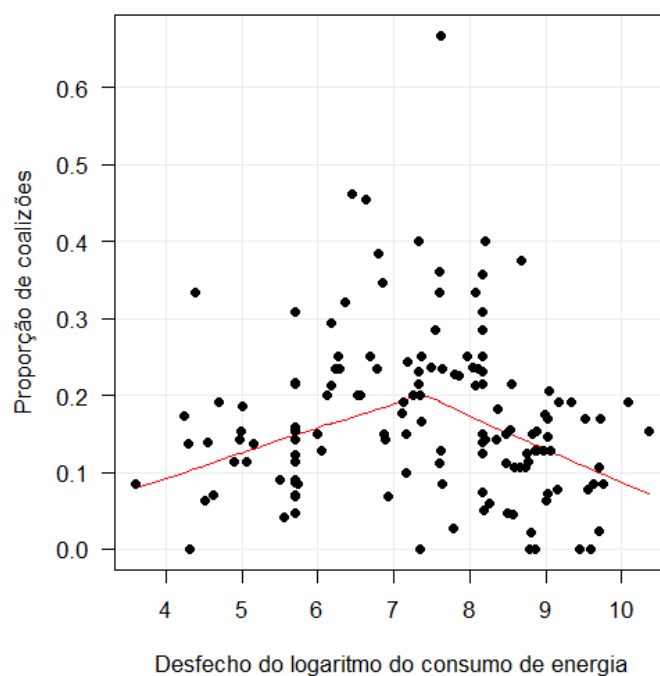


Figura B.1.24 Gráficos de dispersão entre o desfecho do logaritmo do consumo de energia e proporção de coalizões.

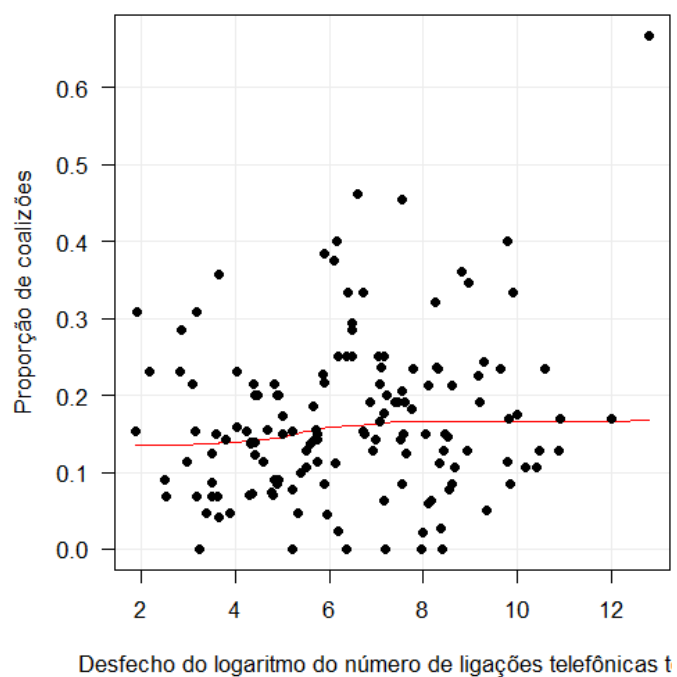


Figura B.1.25 Gráficos de dispersão entre o desfecho do logaritmo do número de ligações telefônicas por 100 habitantes e proporção de coalizões.

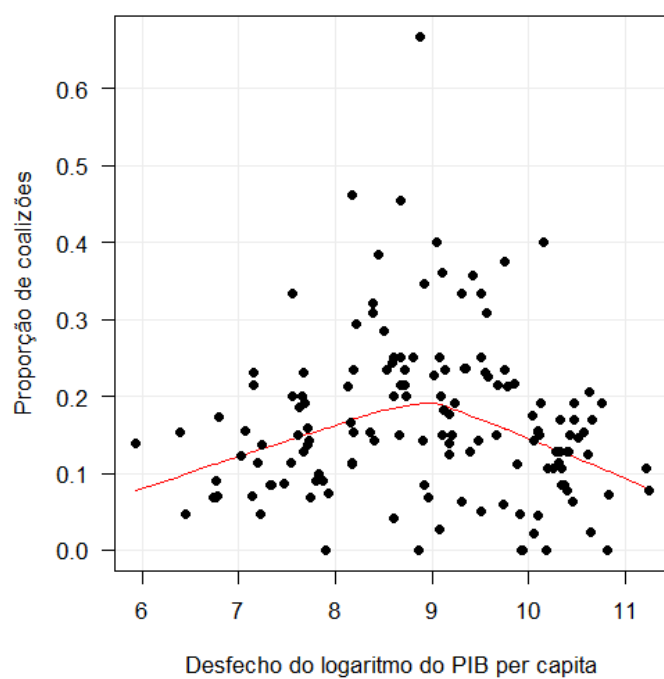


Figura B.1.26 Gráficos de dispersão entre o desfecho do logaritmo do PIB per capita e proporção de coalizões.

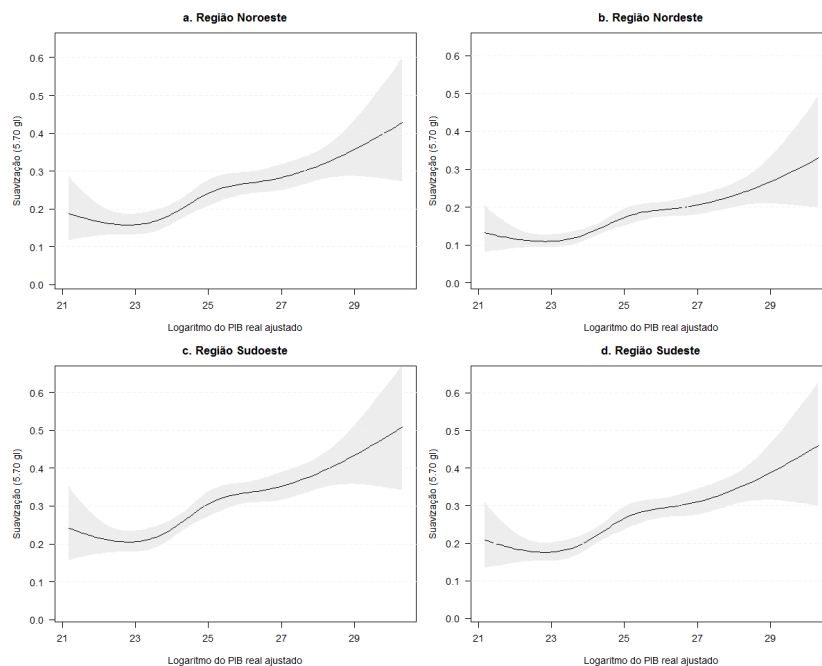
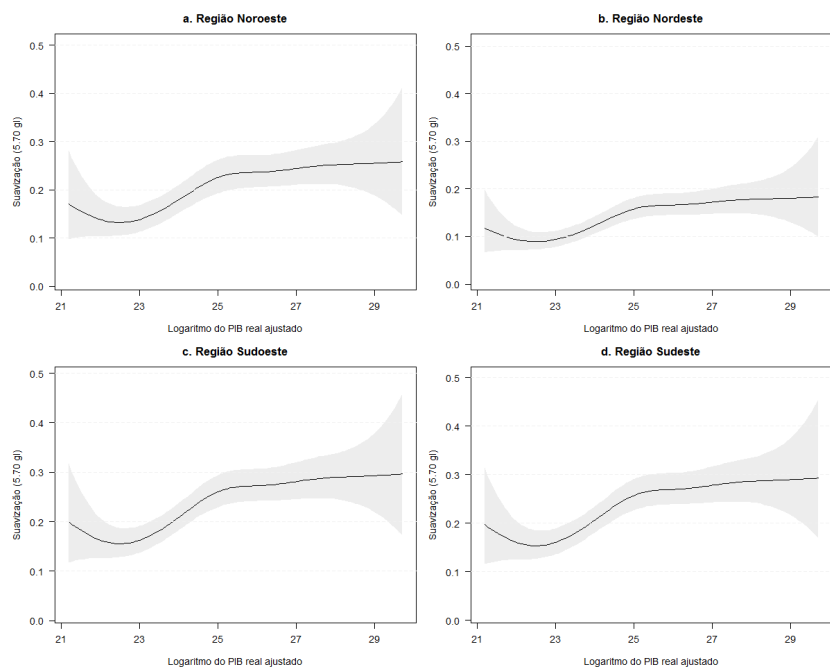


Figura B.1.27 Relações funcionais entre variáveis explicativas e proporção de coalizões, sob o primeiro critério de coalizões (ENTAR).



F

Figura B.1.28 Relações funcionais entre variáveis explicativas e proporção de coalizões, sob o primeiro critério de coalizões (ENTAR) (sem a China).

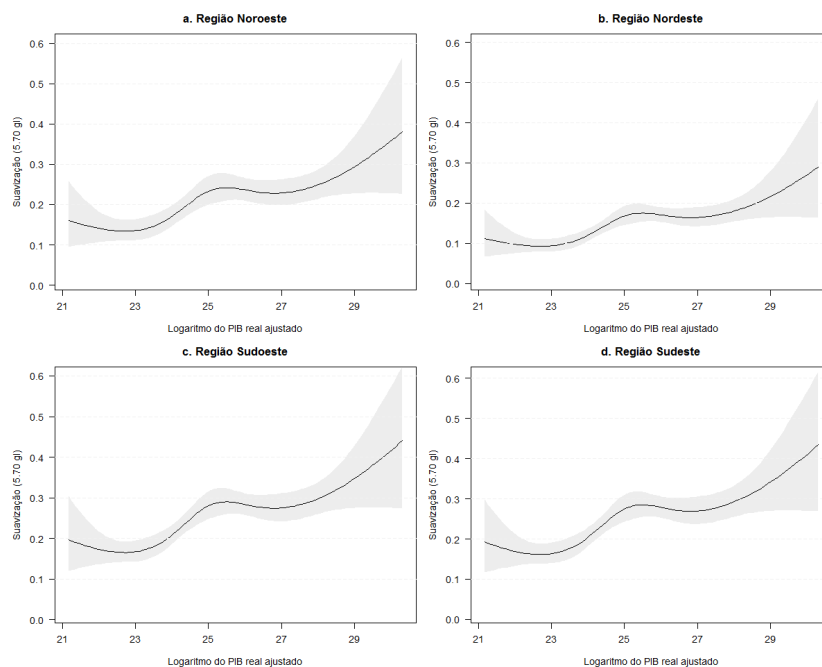


Figura B.1.29 Relações funcionais entre variáveis explicativas e proporção de coalizões, sob o segundo critério de coalizões (ENSAR).

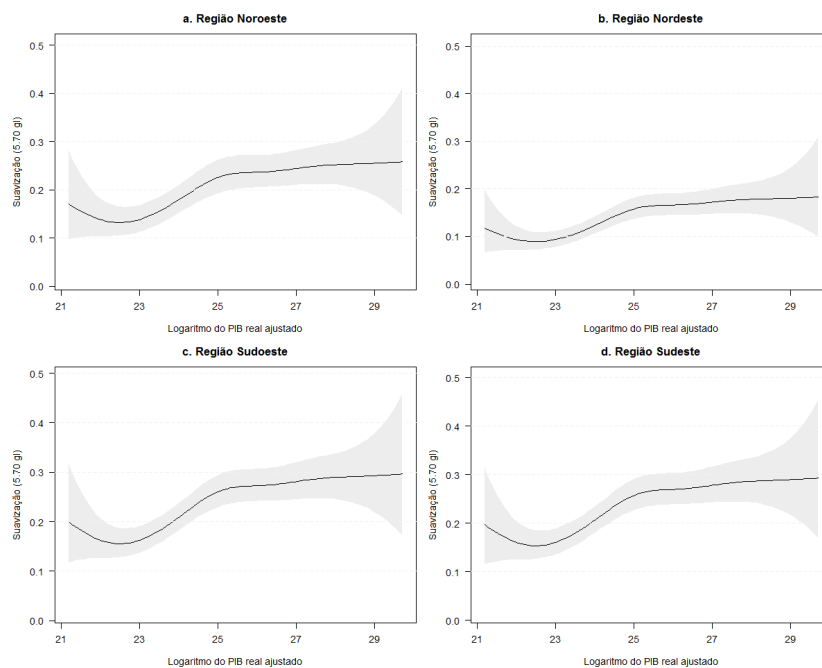


Figura B.1.30 Relações funcionais entre variáveis explicativas e proporção de coalizões, sob o segundo critério de coalizões (ENSAR) (sem a China).

B.2 Tabelas

Tabela B.2.1 Relação das variáveis e suas respectivas técnicas de medida-resumo.

Variável	Medida-resumo
Abertura para o comércio	Desfecho
Consumo de energia elétrica	Logaritmo do desfecho
Exportação	Desfecho
Importação	Desfecho
Indicador de democracia	Média (proporção)
Linhas telefônicas totais*	Logaritmo do desfecho
Participação do Governo no PIB	Desfecho
PIB per capita	Logaritmo do desfecho
PIB real	Logaritmo do desfecho
População	Coeficiente angular da reta
População urbana	Desfecho
Score de democracia	Área sob a curva
Taxa de crescimento do PIB	Média

Tabela B.2.2 Medidas de posição e variabilidade das medidas-resumo das variáveis resposta.

Critério de coalizão	Variável	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo	Média	Desv.Pad.	Coef. Var.
1	Taxa de coalizões	0	0,15	0,23	0,36	2,00	0,28	0,23	0,82
1	Proporção de coalizões	0	0,09	0,15	0,22	0,67	0,17	0,10	0,62
1	Taxa de coalizões ponderada	0	0,02	0,04	0,08	0,25	0,06	0,05	0,88
2	Taxa de coalizões	0	0,11	0,19	0,29	1,50	0,22	0,19	0,85
2	Proporção de coalizões	0	0,09	0,14	0,20	0,67	0,15	0,10	0,69
2	Taxa de coalizões ponderada	0	0,02	0,04	0,06	0,43	0,05	0,06	1,15

Tabela B.2.5 Correlação entre as medidas-resumo das variáveis explicativas com as medidas-resumo das variáveis resposta (em relação ao ENTAR).

	Taxa de coalizões	Proporção de coalizões	Taxa de coalizões ponderada
Abertura para o comércio	0,17	0,15	0,27
Coeficiente angular - População	0,25	0,21	0,31
Exportação	0,1	0,21	0,42
Importação	-0,04	0,04	0,21
Indicador de democracia médio	0,19	0,28	0,48
Logaritmo consumo de energia elétrica	-0,04	0,04	0,16
Logaritmo linhas telefônicas totais	-0,02	0,02	0,21
Logaritmo PIB per capita	0,13	0,27	0,11
Logaritmo PIB real	-0,04	-0,04	-0,17
Participação do Governo no PIB	-0,04	-0,04	-0,07
População urbana	-0,15	-0,24	-0,16
Score de democracia acumulado	-0,08	-0,06	-0,02
Taxa de crescimento do PIB médio	-0,01	-0,05	-0,07

APÊNDICE C

Figuras e tabelas da análise inferencial

C Figuras e tabelas da análise inferencial

C.1 Modelo Linear Generalizado

Tabela C.1.1 Estimativas dos coeficientes do modelo em 8.1.1 para a variável ENTAR

Coeficientes		Estimativa	Erro Padrão	Valor-P
β	LPIBa	0,153	0,026	<0,0001
μ_1	Nordeste	-1,652	0,078	<0,0001
μ_2	Noroeste	-1,237	0,084	<0,0001
μ_3	Sudoeste	-1,167	0,108	<0,0001
μ_4	Sudeste	-0,884	0,158	<0,0001

Tabela C.1.2 Estimativas dos coeficientes do modelo em 8.1.1 para a variável ENSAR

Coeficientes		Estimativa	Erro Padrão	Valor-P
β	LPIBa	0,124	0,027	<0,0001
μ_1	Nordeste	-1,817	0,083	<0,0001
μ_2	Noroeste	-1,404	0,088	<0,0001
μ_3	Sudoeste	-1,269	0,110	<0,0001
μ_4	Sudeste	-1,099	0,165	<0,0001

Tabela C.1.3 Comparações entre os coeficientes com e sem a utilização da variável instrumental (para variável ENTAR)

Coeficientes	Estimativa com GLM simples		Estimativa com GLM em dois estágios	
	Estimativa	Erro Padrão	Estimativa	Erro padrão
β	0,135	0,024	0,153	0,026
μ_1	-1,657	0,077	-1,652	0,076
μ_2	-1,212	0,084	-1,237	0,084
μ_3	-1,218	0,106	-1,167	0,108
μ_4	-0,853	0,157	-0,884	0,158

Tabela C.1.3 Comparações entre os coeficientes com e sem a utilização da variável instrumental (para variável ENSAR)

Coeficientes	Estimativa com GLM simples		Estimativa com GLM em dois estágios	
	Estimativa	Erro Padrão	Estimativa	Erro padrão
β	0,118	0,025	0,124	0,027
μ_1	-1,831	0,081	-1,817	0,083
μ_2	-1,383	0,088	-1,404	0,088
μ_3	-1,307	0,109	-1,269	0,110
μ_4	-1,083	0,164	-1,099	0,165

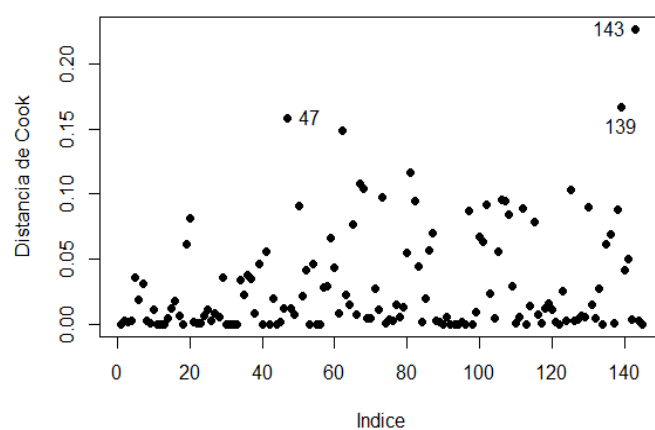


Figura C.1.1 Distância de Cook para os resíduos do modelo em 8.1.1 aplicado à variável ENTAR

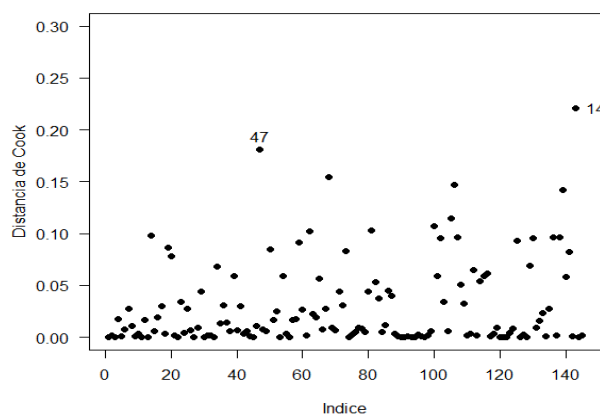


Figura C.1.2 Distância de Cook para os resíduos do modelo em 8.1.1 aplicado à variável ENSAR

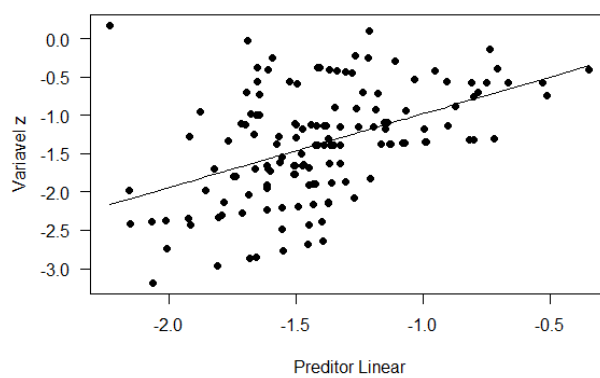


Figura C.1.3 Valores observados versus valores ajustados pelo modelo em 8.1.1 aplicado à variável ENTAR

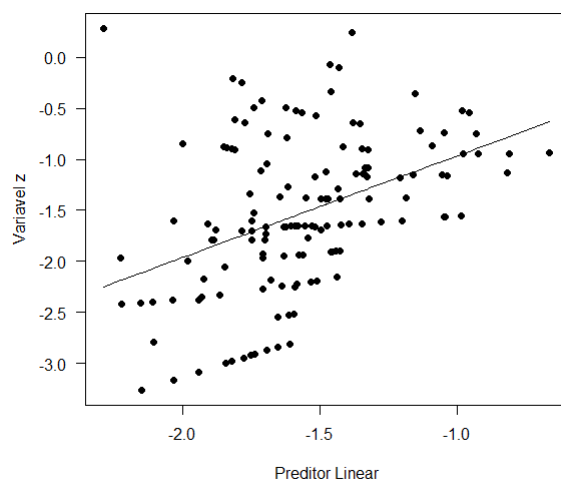


Figura C.1.4 Valores observados versus valores ajustados pelo modelo em 8.1.1 aplicado à variável ENSAR

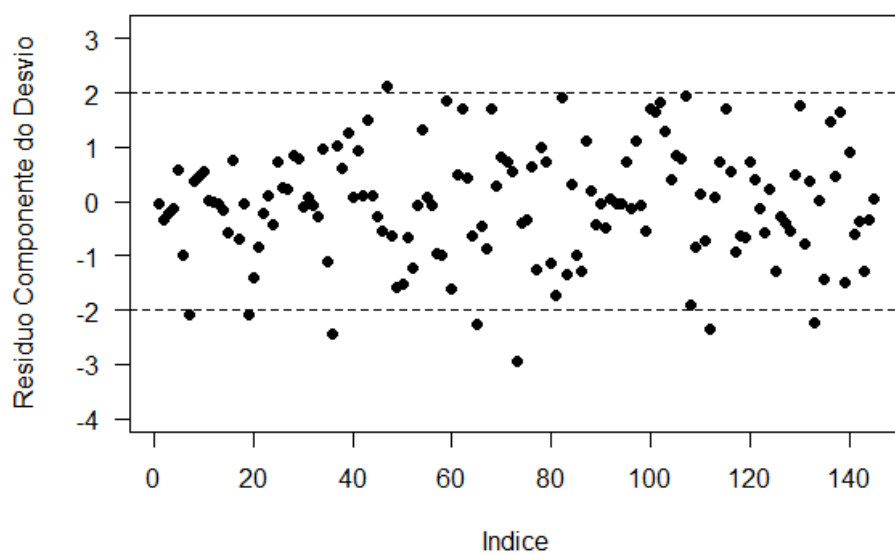


Figura C.1.5 Gráfico de dispersão dos resíduos do modelo em 8.1.1 aplicado à variável ENTAR

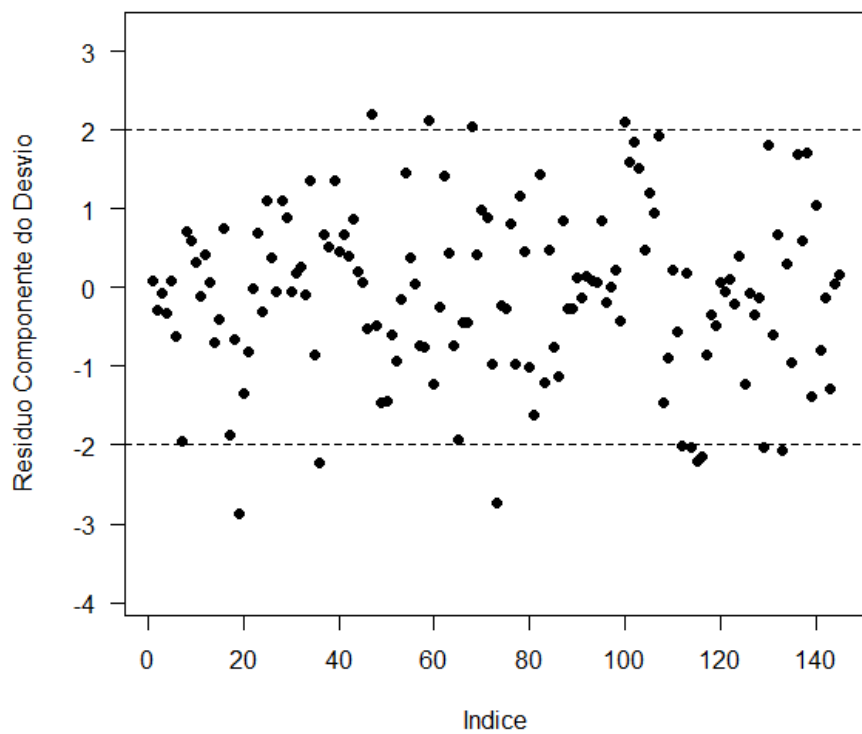


Figura C.1.6 Gráfico de dispersão dos resíduos do modelo em 8.1.1 aplicado à variável ENSAR

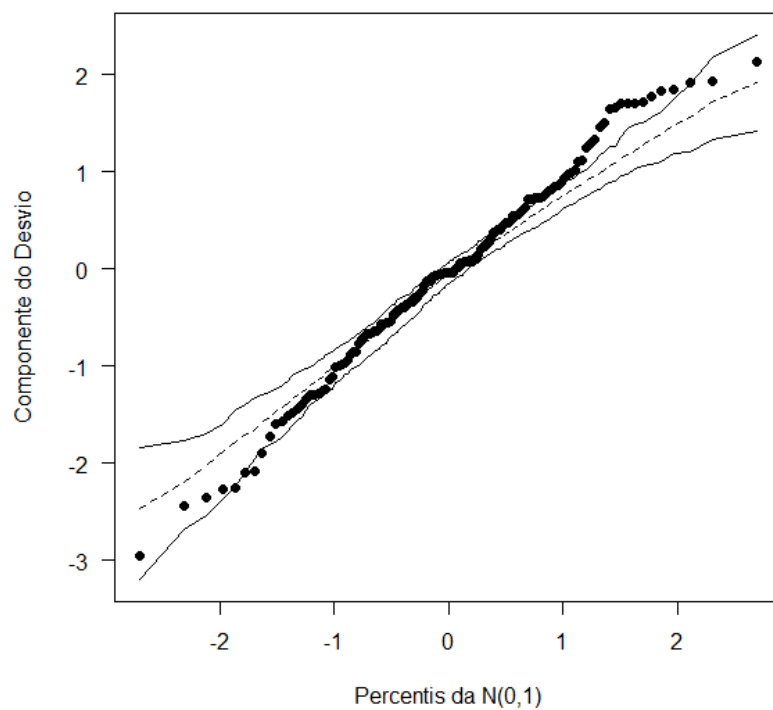


Figura C.1.7 Gráfico Q-Q normal para os resíduos do modelo em 8.1.1 aplicado à variável ENSAR

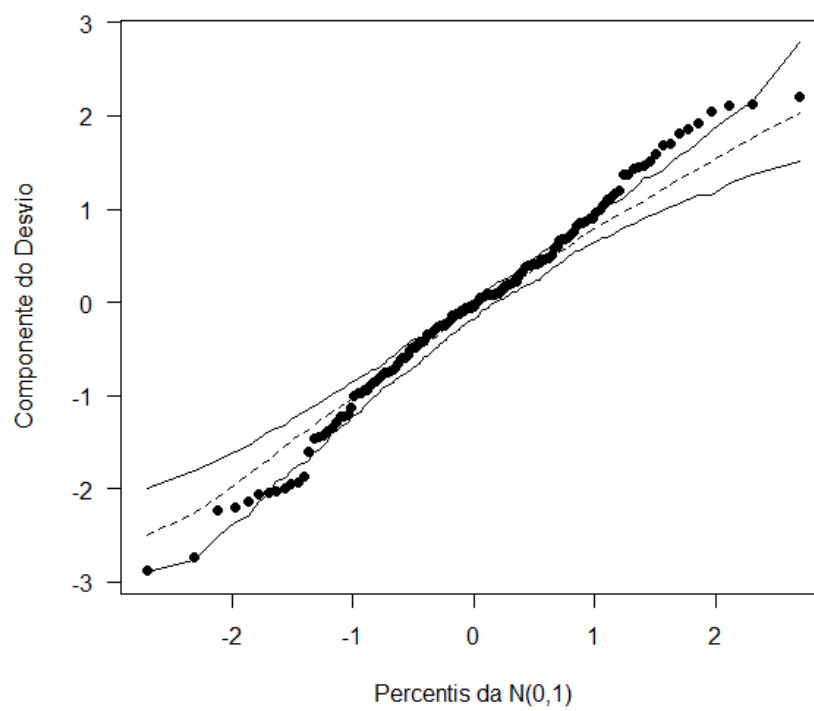


Figura C.1.8 Gráfico Q-Q normal para os resíduos do modelo em 8.1.1 aplicado à variável ENSAR

C.2 Modelo BART

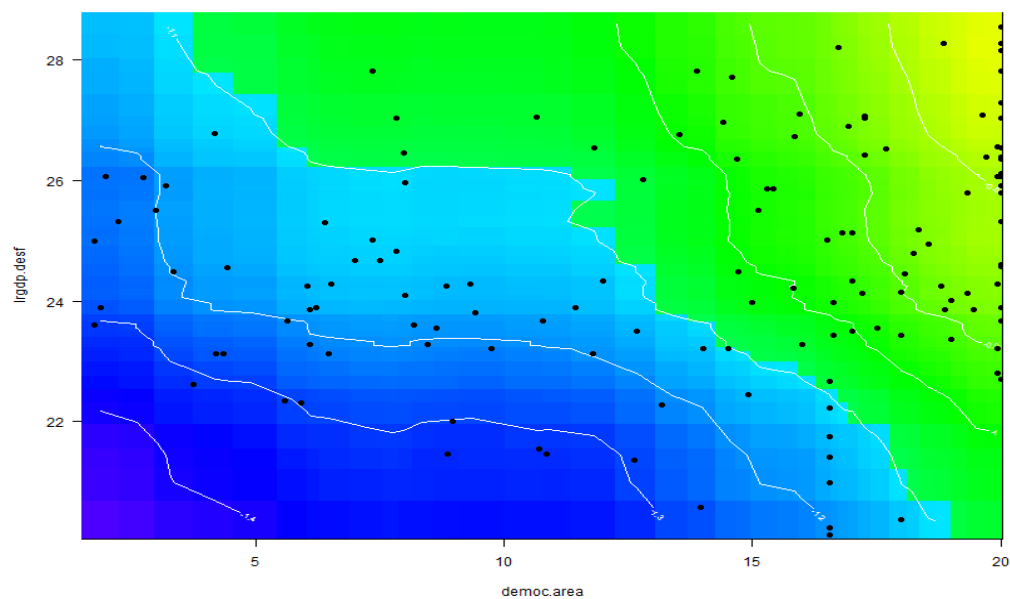


Figura 0.1 probabilidade estimada de entrar em coalizão pelo logaritmo do PIB e pelo índice de democracia acumulado avaliada segundo o critério ENSAR

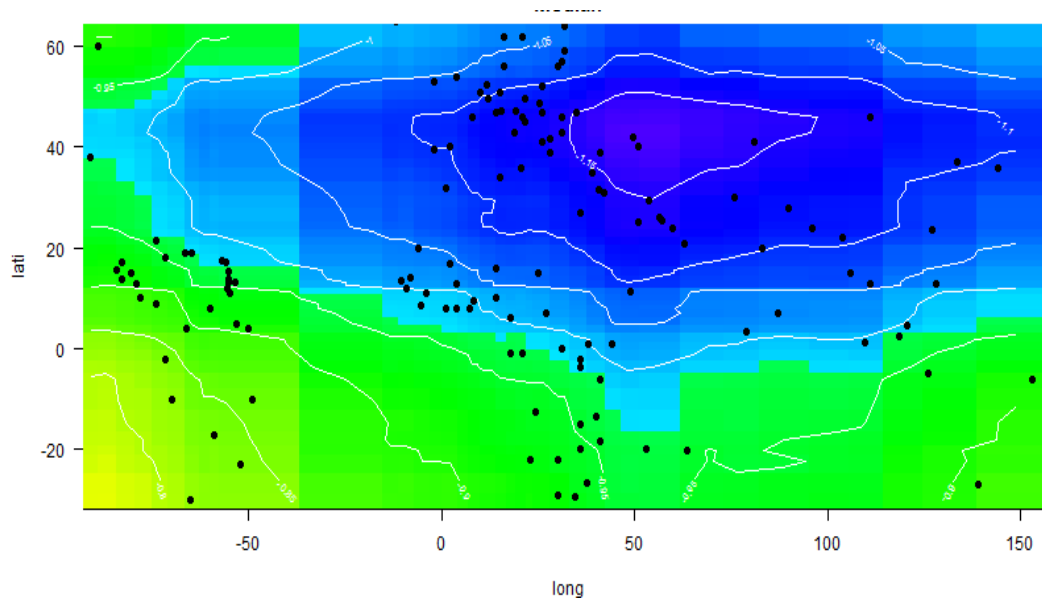


Figura 0.2 probabilidade estimada de entrar em coalizão pela posição geográfica avaliada segundo o critério ENSAR

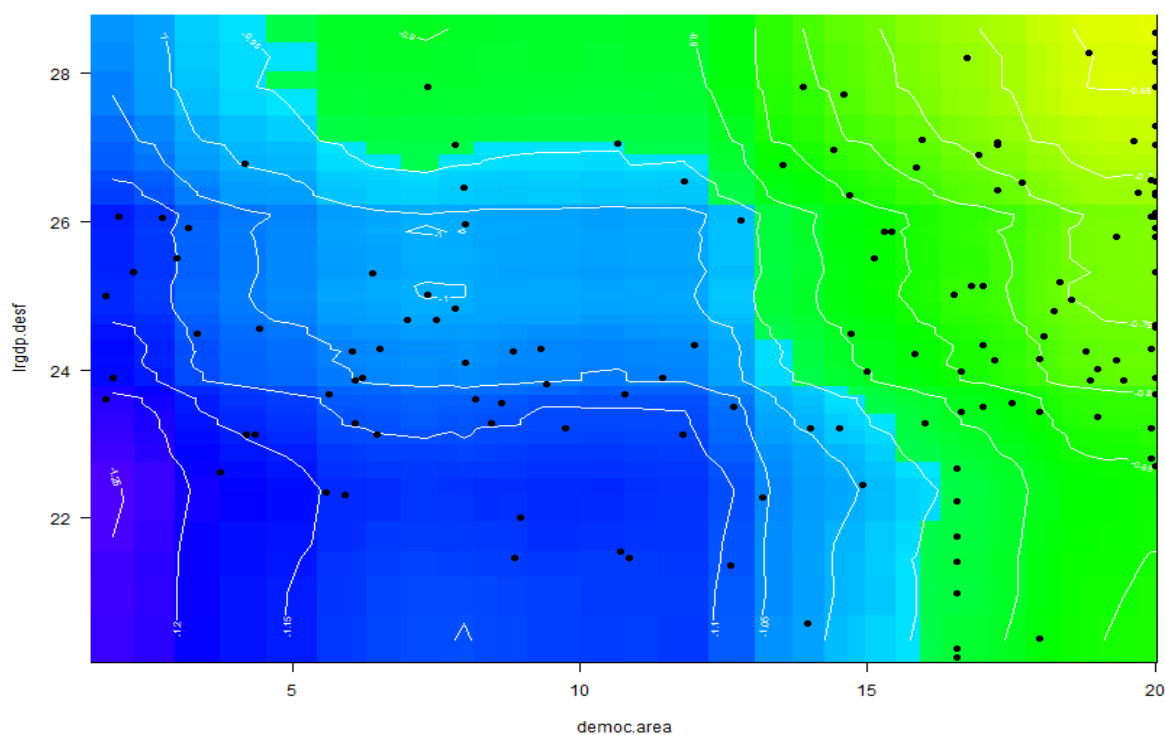


Figura 0.3 probabilidade estimada de entrar em coalizão pelo logaritmo do PIB e pelo índice de democracia acumulado avaliada segundo o critério ENTAR

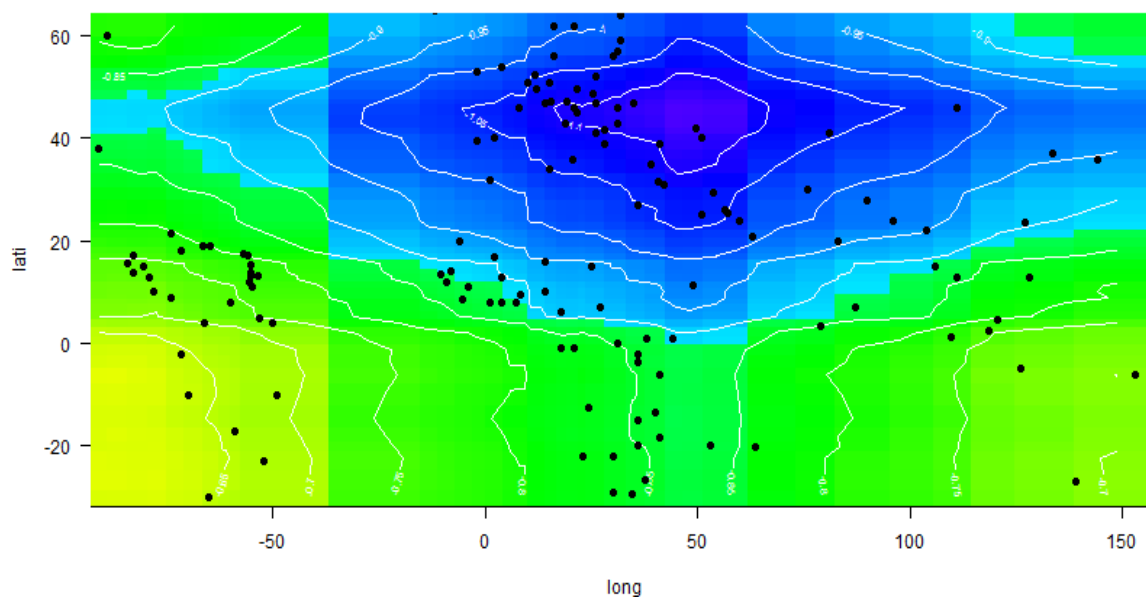


Figura 0.4 probabilidade estimada de entrar em coalizão pela posição geográfica avaliada segundo o critério ENTAR

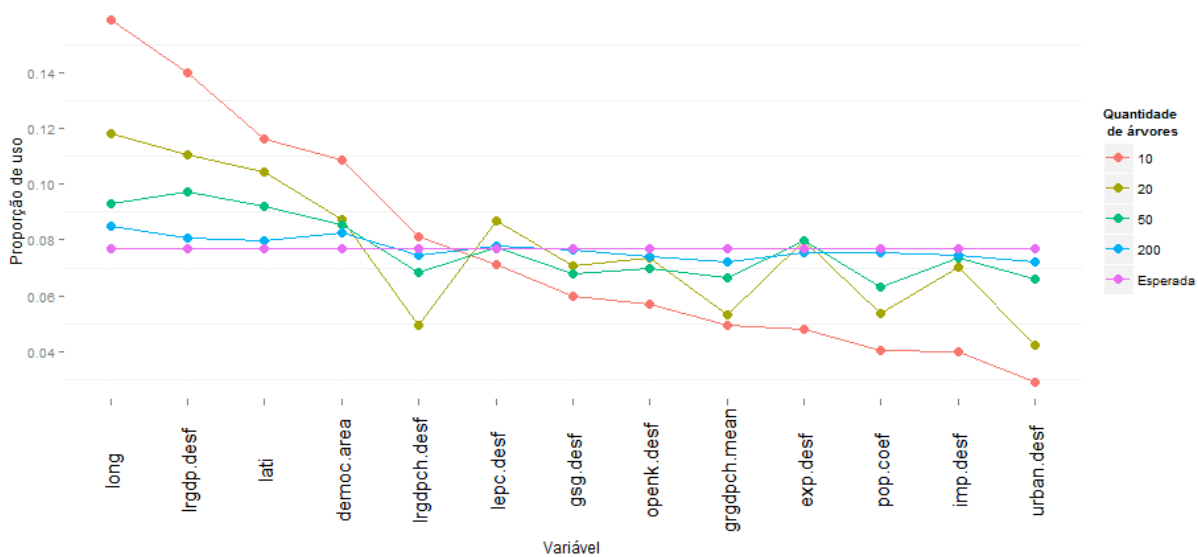


Figura 0.5 Gráfico da proporção de uso das variáveis nas árvores de decisão do modelo BART

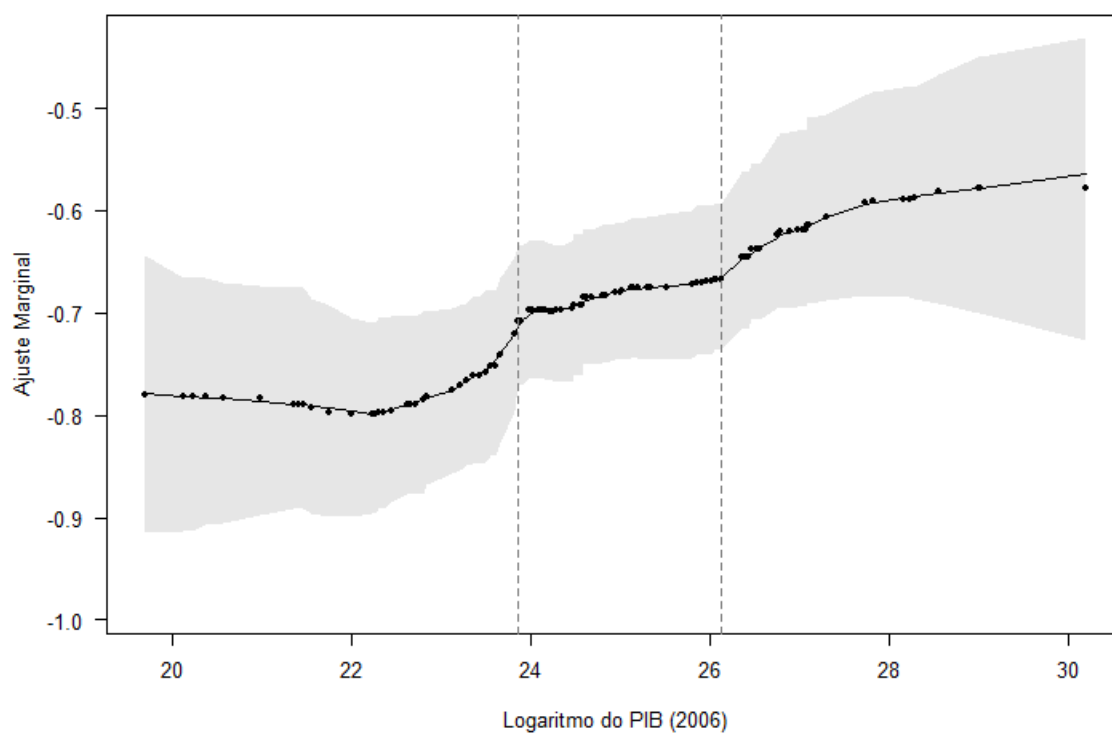


Figura 0.6 Efeito marginal do logaritmo do PIB no probito da probabilidade de entrar em coalizão para o primeiro critério (ENTAR)

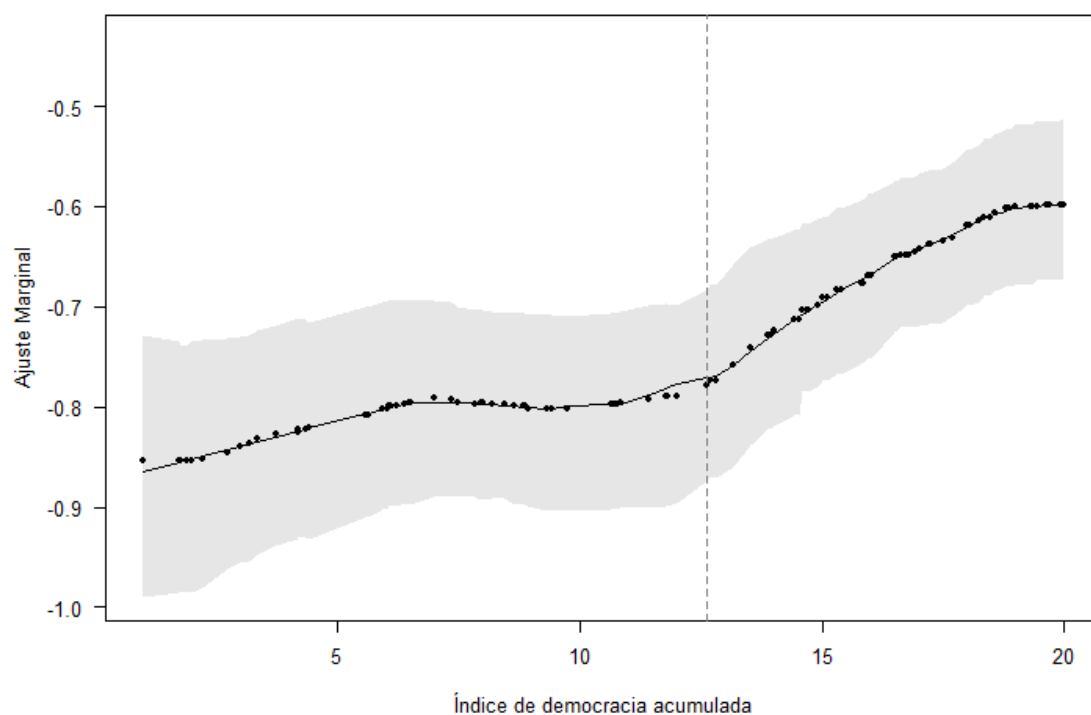


Figura 0.7 Efeito Marginal do índice de democracia acumulado no probito da probabilidade de entrar em coalizão (ENTAR)

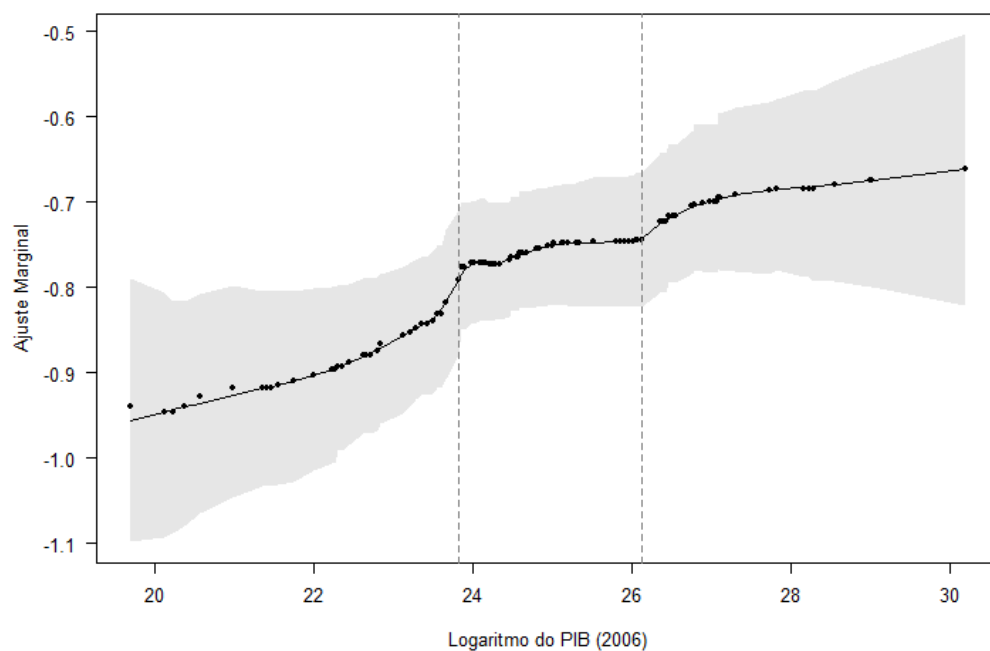


Figura 0.8 Efeito marginal do logaritmo do PIB no probito da probabilidade de entrar em coalizão para o segundo critério (ENSAR)

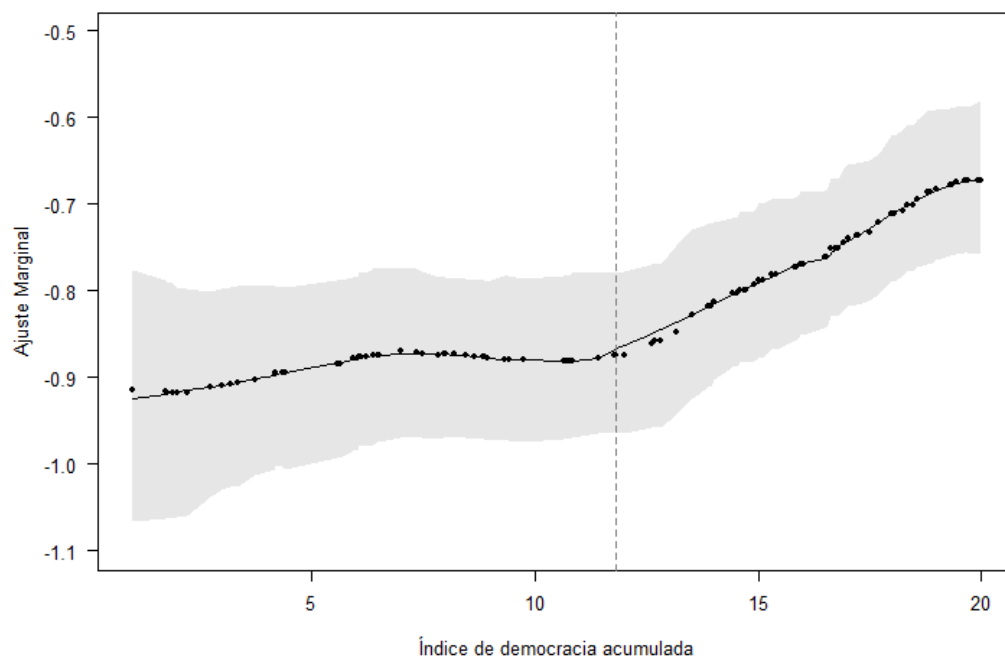


Figura 0.9 Efeito Marginal do índice de democracia acumulada no probito da probabilidade de entrar em coalizão (ENSAR)

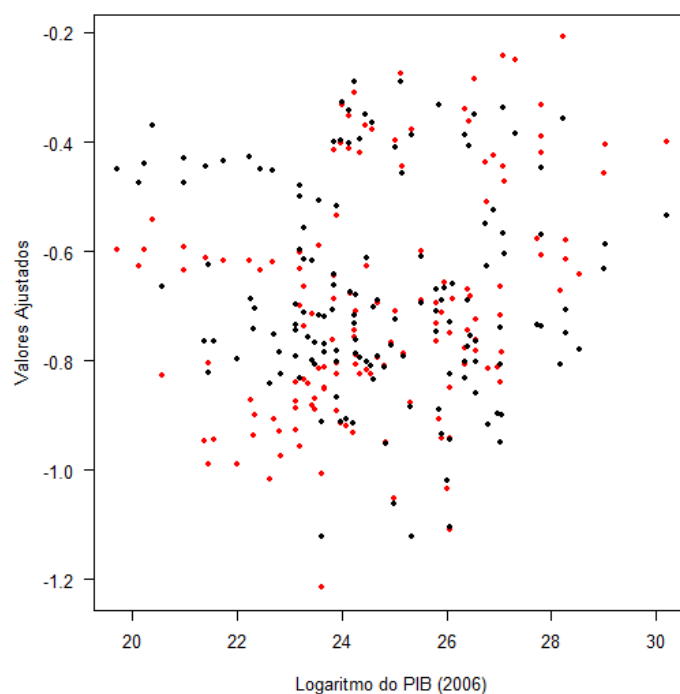


Figura 0.10 Valores ajustados com o real valor do PIB (em vermelho) e valores ajustados supondo PIB médio para todos os países (em preto) versus logaritmo do PIB de 2006 (ENTAR)

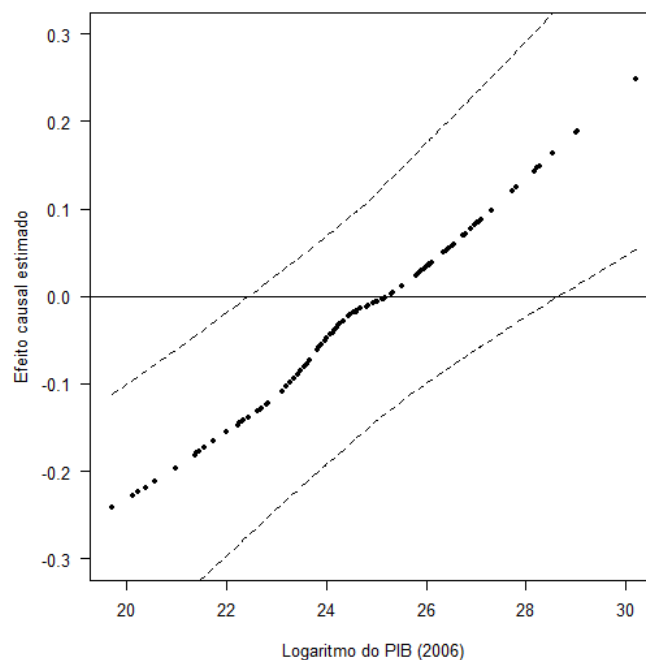


Figura 0.11 (Efeito causal) Diferença entre as probabilidades estimadas ao supor logaritmo do PIB médio aos países (ENTAR)

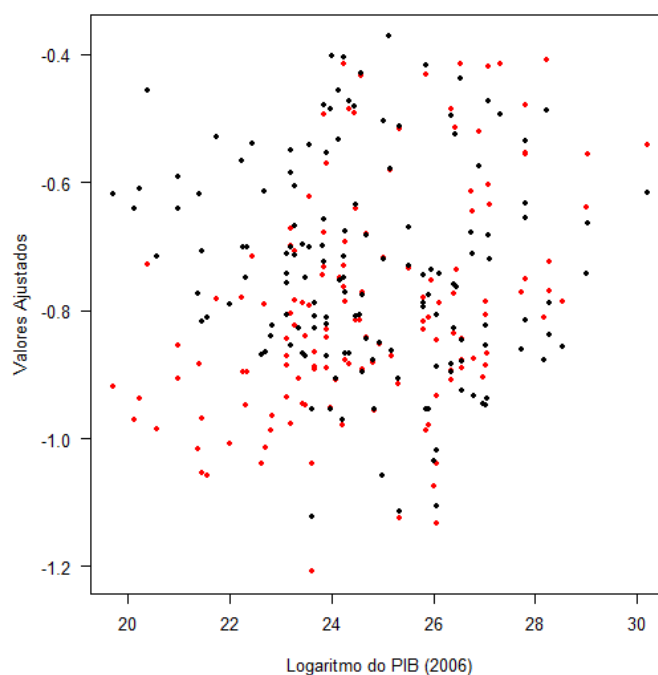


Figura 0.12 Valores ajustados com o real valor do logaritmo do PIB (em vermelho) e valores ajustados supondo PIB médio para todos os países (em preto) versus logaritmo do PIB de 2006 (ENSAR)

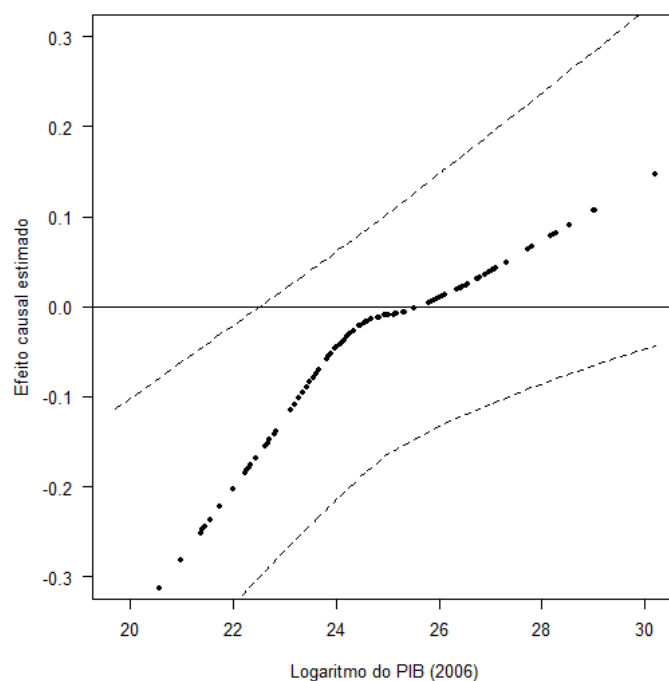


Figura 0.13 (Efeito causal) Diferença entre as probabilidades estimadas ao supor PIB médio aos países (ENSAR)

Tabela 0.1 Lista de países e suas respectivas probabilidades marginais estimadas de entrar em coalizão (para , ordenados pelo logaritmo do PIB total de 2006. O fator “grupo” agrupa países de mesma faixa delimitada na Figura C.2.6

país	lrgdp.def	grupo	mediana	IC 95% para a mediana	
				inf	sup
Dominica	19,70	1	0,176	0,137	0,218
St Kitts and Nevis	20,13	1	0,176	0,139	0,213
St Vincent and the Grenadines	20,25	1	0,176	0,141	0,212
Solomon Islands	20,38	1	0,175	0,142	0,212
Guinea-Bissau	20,57	1	0,175	0,142	0,210
Antigua and Barbuda	20,99	1	0,175	0,145	0,208
Grenada	20,99	1	0,175	0,145	0,208
Maldives	21,37	1	0,173	0,146	0,205
St Lucia	21,42	1	0,173	0,146	0,205
Djibouti	21,47	1	0,173	0,146	0,205
Guyana	21,47	1	0,173	0,146	0,205
Gambia	21,56	1	0,172	0,144	0,201
Belize	21,75	1	0,172	0,144	0,199
Central African Republic	22,00	1	0,171	0,143	0,199
Suriname	22,23	1	0,172	0,146	0,198

país	lrgdp.def	grupo	mediana	IC 95% para a mediana	
				inf	sup
Lesotho	22,27	1	0,172	0,146	0,198
Togo	22,31	1	0,172	0,146	0,197
Burundi	22,35	1	0,172	0,146	0,197
Fiji	22,45	1	0,173	0,148	0,198
Mauritania	22,63	1	0,175	0,150	0,199
Barbados	22,67	1	0,175	0,150	0,199
Mongolia	22,71	1	0,175	0,150	0,199
Malta	22,80	1	0,175	0,150	0,199
Swaziland	22,84	1	0,177	0,152	0,200
Benin	23,12	1	0,180	0,155	0,202
Gabon	23,12	1	0,180	0,155	0,202
Rwanda	23,12	1	0,180	0,155	0,202
Sierra Leone	23,12	1	0,180	0,155	0,202
Iceland	23,21	1	0,181	0,157	0,204
Nicaragua	23,21	1	0,181	0,157	0,204
Niger	23,21	1	0,181	0,157	0,204
Papua New Guinea	23,21	1	0,181	0,157	0,204
Congo	23,29	1	0,182	0,159	0,204
Haiti	23,29	1	0,182	0,159	0,204
Namibia	23,29	1	0,182	0,159	0,204
Macedonia	23,36	1	0,184	0,161	0,206
Mali	23,43	1	0,184	0,161	0,206
Moldova	23,43	1	0,184	0,161	0,206
Albania	23,50	1	0,185	0,163	0,208
Madagascar	23,50	1	0,185	0,163	0,208
Botswana	23,56	1	0,187	0,164	0,210
Malawi	23,56	1	0,187	0,164	0,210
Bahrain	23,61	1	0,187	0,164	0,210
Kyrgyzstan	23,61	1	0,187	0,164	0,210
Brunei	23,67	1	0,192	0,168	0,215
Cyprus	23,67	1	0,192	0,168	0,215
Burkina Faso	23,67	1	0,192	0,168	0,215
Zambia	23,81	1	0,199	0,173	0,222
Estonia	23,86	2	0,206	0,185	0,226
Jamaica	23,86	2	0,206	0,185	0,226
Zimbabwe	23,86	2	0,206	0,185	0,226
Chad	23,90	2	0,206	0,185	0,226
Congo, Democratic Republic	23,90	2	0,206	0,185	0,226
Mauritius	23,90	2	0,206	0,185	0,226
Senegal	23,90	2	0,206	0,185	0,226
Armenia	23,98	2	0,208	0,188	0,227
Honduras	23,98	2	0,208	0,188	0,227

país	lrgdp.def	grupo	mediana	IC 95% para a mediana	
				inf	sup
Panama	24,02	2	0,208	0,188	0,227
Jordan	24,09	2	0,209	0,188	0,228
Paraguay	24,12	2	0,209	0,188	0,228
Trinidad and Tobago	24,12	2	0,209	0,188	0,228
Latvia	24,16	2	0,209	0,188	0,228
Georgia	24,22	2	0,208	0,189	0,227
Bolivia	24,25	2	0,208	0,189	0,227
Guinea	24,25	2	0,208	0,189	0,227
Uganda	24,25	2	0,208	0,189	0,227
Ghana	24,28	2	0,208	0,189	0,227
Luxembourg	24,28	2	0,208	0,189	0,227
Tanzania	24,28	2	0,208	0,189	0,227
Cambodia	24,33	2	0,208	0,189	0,227
El Salvador	24,33	2	0,208	0,189	0,227
Uruguay	24,46	2	0,208	0,189	0,227
Cote d'Ivoire	24,48	2	0,210	0,190	0,228
Mozambique	24,48	2	0,210	0,190	0,228
Cameroon	24,55	2	0,210	0,190	0,228
Costa Rica	24,57	2	0,210	0,190	0,228
Lithuania	24,59	2	0,211	0,192	0,229
Slovenia	24,62	2	0,211	0,192	0,229
Angola	24,67	2	0,211	0,192	0,229
Nepal	24,67	2	0,211	0,192	0,229
Croatia	24,80	2	0,212	0,193	0,231
Qatar	24,83	2	0,212	0,193	0,231
Bulgaria	24,94	2	0,213	0,195	0,233
Oman	25,00	2	0,213	0,195	0,233
Guatemala	25,01	2	0,214	0,196	0,235
Kenya	25,01	2	0,214	0,196	0,235
Ecuador	25,13	2	0,214	0,196	0,236
Dominican Republic	25,14	2	0,214	0,196	0,236
Slovakia	25,19	2	0,214	0,196	0,236
Tunisia	25,31	2	0,215	0,196	0,236
Kuwait	25,33	2	0,216	0,196	0,237
New Zealand	25,33	2	0,216	0,196	0,237
Sri Lanka	25,51	2	0,216	0,198	0,237
Cuba	25,51	2	0,216	0,198	0,237
Liechtenstein	25,58	2	0,216	0,198	0,237
Finland	25,80	2	0,218	0,198	0,238
Ireland	25,80	2	0,218	0,198	0,238
Israel	25,80	2	0,218	0,198	0,238
Hungary	25,86	2	0,219	0,199	0,239

país	lrgdp.def	grupo	mediana	IC 95% para a mediana	
				inf	sup
Peru	25,86	2	0,219	0,199	0,239
Denmark	25,92	2	0,219	0,199	0,239
Morocco	25,92	2	0,219	0,199	0,239
Singapore	25,97	2	0,219	0,199	0,239
Romania	26,02	2	0,219	0,199	0,239
Czech Republic	26,07	2	0,220	0,200	0,241
Portugal	26,07	2	0,220	0,200	0,241
United Arab Emirates	26,07	2	0,220	0,200	0,241
Norway	26,12	2	0,220	0,200	0,241
Myanmar	26,13	2	0,220	0,200	0,241
Austria	26,36	3	0,228	0,205	0,254
Chile	26,36	3	0,228	0,205	0,254
Switzerland	26,36	3	0,228	0,205	0,254
Greece	26,39	3	0,228	0,205	0,254
Sweden	26,39	3	0,228	0,205	0,254
Venezuela	26,43	3	0,228	0,205	0,254
Nigeria	26,46	3	0,233	0,209	0,258
Colombia	26,52	3	0,233	0,209	0,258
Bangladesh	26,55	3	0,233	0,209	0,258
Belgium	26,55	3	0,233	0,209	0,258
Philippines	26,74	3	0,237	0,212	0,262
Malaysia	26,76	3	0,237	0,212	0,262
Egypt	26,79	3	0,238	0,213	0,264
South Africa	26,90	3	0,239	0,214	0,266
Poland	26,98	3	0,239	0,214	0,266
Netherlands	27,03	3	0,239	0,214	0,266
Saudi Arabia	27,03	3	0,239	0,214	0,266
Turkey	27,03	3	0,239	0,214	0,266
Pakistan (1972-)	27,05	3	0,239	0,214	0,266
Argentina	27,07	3	0,239	0,214	0,266
Taiwan	27,09	3	0,241	0,216	0,269
Thailand	27,10	3	0,241	0,216	0,269
Australia	27,30	3	0,242	0,217	0,271
Korea, South	27,73	3	0,246	0,218	0,277
Canada	27,81	3	0,247	0,218	0,278
Indonesia	27,81	3	0,247	0,218	0,278
Mexico	27,81	3	0,247	0,218	0,278
Spain	27,81	3	0,247	0,218	0,278
Italy	28,16	3	0,247	0,217	0,281
Brazil	28,22	3	0,247	0,217	0,281
France	28,27	3	0,248	0,216	0,282
United Kingdom	28,27	3	0,248	0,216	0,282

país	lrgdp.def	grupo	mediana	IC 95% para a mediana	
				inf	sup
Germany	28,55	3	0,248	0,216	0,284
Japan	28,99	3	0,250	0,214	0,289
India	29,02	3	0,250	0,214	0,289
United States	30,20	3	0,249	0,206	0,297

Tabela 0.2 Lista de países e suas respectivas probabilidades marginais estimadas de entrar em coalizão, ordenados pelo índice de democracia acumulado. O fator “grupo” agrupa países de mesma faixa delimitada na figura C.2.7

país	democ.area	grupo	mediana	IC 95% para a mediana	
				inf	sup
Swaziland	1	1	-0,9621	-1,12168	-0,83095
Oman	1,8	1	-0,9661	-1,11699	-0,83805
Bahrain	1,8	1	-0,9661	-1,11699	-0,84162
Congo	1,9	1	-0,9661	-1,11699	-0,84162
United Arab Emirates	2	1	-0,9661	-1,11699	-0,84162
Kuwait	2,2	1	-0,9661	-1,10768	-0,8452
Myanmar	2,8	1	-0,9621	-1,08482	-0,84162
Cuba	3	1	-0,9542	-1,08032	-0,83805
Morocco	3,2	1	-0,9502	-1,07584	-0,83805
Cote d'Ivoire	3,4	1	-0,9502	-1,07138	-0,83095
Mauritania	3,8	1	-0,9424	-1,06694	-0,83095
Egypt	4,2	1	-0,9385	-1,05812	-0,83095
Rwanda	4,2	1	-0,9385	-1,05812	-0,83095
Gabon	4,4	1	-0,9385	-1,05812	-0,83095
Cameroon	4,4	1	-0,9385	-1,04939	-0,83095
Burundi	5,6	1	-0,9269	-1,02365	-0,82389
Burkina Faso	5,6	1	-0,9269	-1,02365	-0,82389
Togo	5,9	1	-0,9192	-1,01943	-0,81687
Uganda	6	1	-0,9192	-1,01943	-0,81687
Zimbabwe	6,1	1	-0,9154	-1,01943	-0,81687
Congo	6,1	1	-0,9154	-1,01943	-0,81687
Chad	6,2	1	-0,9154	-1,01943	-0,81687
Tunisia	6,4	1	-0,9154	-1,01943	-0,81687
Sierra Leone	6,5	1	-0,9116	-1,01522	-0,81338
Tanzania	6,5	1	-0,9116	-1,01522	-0,81338
Nepal	7	1	-0,9116	-1,01103	-0,81687
Indonesia	7,4	1	-0,9116	-1,01103	-0,81338
Kenya	7,4	1	-0,9116	-1,01103	-0,81338
Angola	7,5	1	-0,9154	-1,01522	-0,81687
Nigeria	8	1	-0,9154	-1,01522	-0,82389
Jordan	8	1	-0,9154	-1,01522	-0,82389
Singapore	8	1	-0,9154	-1,01522	-0,82389
Kyrgyzstan	8,2	1	-0,9192	-1,02789	-0,83095
Qatar	8,3	1	-0,9192	-1,02789	-0,83095
Saudi Arabia	8,3	1	-0,9192	-1,02789	-0,83095
Haiti	8,5	1	-0,923	-1,03215	-0,8345
Malawi	8,6	1	-0,923	-1,03215	-0,83095
Guinea	8,8	1	-0,923	-1,02789	-0,82742

país	democ.area	grupo	mediana	IC 95% para a mediana	
				inf	sup
Djibouti	8,9	1	-0,923	-1,02789	-0,82742
Central African Republic	9	1	-0,9269	-1,02789	-0,8345
Ghana	9,3	1	-0,9269	-1,03215	-0,83095
Zambia	9,4	1	-0,9269	-1,03215	-0,83095
Niger	9,8	1	-0,9269	-1,03215	-0,83095
Pakistan (1972-)	10,7	1	-0,923	-1,03643	-0,82389
Gambia	10,7	1	-0,923	-1,03643	-0,82389
Guyana	10,9	1	-0,9269	-1,04073	-0,82389
Senegal	11,4	1	-0,923	-1,04073	-0,82389
Benin	11,8	1	-0,923	-1,03643	-0,82389
Bangladesh	11,8	1	-0,923	-1,03643	-0,82389
Cambodia	12	1	-0,923	-1,03643	-0,82038
Brunei	12,4	1	-0,9154	-1,03215	-0,8099
Maldives	12,4	1	-0,9154	-1,03215	-0,8099
Madagascar	12,7	2	-0,9002	-1,01522	-0,79606
Romania	12,8	2	-0,9002	-1,01522	-0,79606
Lesotho	13,2	2	-0,8816	-0,99446	-0,77897
Malaysia	13,5	2	-0,856	-0,97009	-0,7621
Mexico	13,9	2	-0,8452	-0,94238	-0,75542
Guinea-Bissau	14	2	-0,8452	-0,94238	-0,75542
Papua New Guinea	14	2	-0,8381	-0,93459	-0,74876
Poland	14,4	2	-0,8274	-0,92301	-0,73885
Nicaragua	14,5	2	-0,8274	-0,92301	-0,73885
Korea, South	14,6	2	-0,8169	-0,91156	-0,73556
Chile	14,7	2	-0,8169	-0,91156	-0,73556
Mozambique	14,7	2	-0,8169	-0,91156	-0,73556
Fiji	14,9	2	-0,8099	-0,90023	-0,729
Armenia	15	2	-0,803	-0,88529	-0,72574
Sri Lanka	15,1	2	-0,803	-0,88529	-0,72574
Peru	15,3	2	-0,7926	-0,87055	-0,71923
Hungary	15,4	2	-0,7926	-0,87055	-0,71923
Georgia	15,8	2	-0,7824	-0,86325	-0,70952
Philippines	15,8	2	-0,7824	-0,86325	-0,70952
Thailand	15,9	2	-0,779	-0,85239	-0,70952
Namibia	16	2	-0,779	-0,85239	-0,70952
Antigua and Barbuda	16,2	2	-0,7756	-0,85239	-0,70309
Barbados	16,2	2	-0,7756	-0,85239	-0,70309
Belize	16,2	2	-0,7756	-0,85239	-0,70309
Dominica	16,2	2	-0,7756	-0,85239	-0,70309
Grenada	16,2	2	-0,7756	-0,85239	-0,70309
St Kitts and Nevis	16,2	2	-0,7756	-0,85239	-0,70309
St Lucia	16,2	2	-0,7756	-0,85239	-0,70309

país	democ.area	grupo	mediana	IC 95% para a mediana	
				inf	sup
St Vincent and the Grenadines	16,2	2	-0,7756	-0,85239	-0,70309
Suriname	16,2	2	-0,7756	-0,85239	-0,70309
Guatemala	16,5	2	-0,7621	-0,84162	-0,69668
Mali	16,6	2	-0,7588	-0,8345	-0,68713
Honduras	16,6	2	-0,7588	-0,8345	-0,68713
Brazil	16,7	2	-0,7588	-0,8345	-0,68713
Dominican Republic	16,8	2	-0,7588	-0,8345	-0,68713
South Africa	16,9	2	-0,7554	-0,8345	-0,68713
Albania	17	2	-0,7554	-0,83095	-0,68713
Ecuador	17	2	-0,7554	-0,83095	-0,68713
El Salvador	17	2	-0,7554	-0,83095	-0,68713
Paraguay	17,2	2	-0,7521	-0,82742	-0,68713
Argentina	17,3	2	-0,7521	-0,82742	-0,68713
Turkey	17,3	2	-0,7521	-0,82742	-0,68713
Venezuela	17,3	2	-0,7521	-0,82742	-0,68713
Botswana	17,5	2	-0,7488	-0,82389	-0,67764
Colombia	17,7	2	-0,7454	-0,81687	-0,67449
Solomon Islands	18	2	-0,7388	-0,81338	-0,66821
Latvia	18	2	-0,7388	-0,81338	-0,66821
Moldova	18	2	-0,7388	-0,81338	-0,66821
Uruguay	18,1	2	-0,7388	-0,81338	-0,66821
Croatia	18,3	2	-0,7356	-0,7995	-0,66508
Slovakia	18,3	2	-0,729	-0,79606	-0,65884
India	18,5	2	-0,729	-0,79606	-0,65884
Bulgaria	18,6	2	-0,7257	-0,79606	-0,65573
Iceland	18,7	2	-0,7257	-0,79606	-0,65573
Liechtenstein	18,7	2	-0,7257	-0,79606	-0,65573
Luxembourg	18,7	2	-0,7257	-0,79606	-0,65573
Malta	18,7	2	-0,7257	-0,79606	-0,65573
Bolivia	18,8	2	-0,7257	-0,79262	-0,65262
France	18,9	2	-0,7257	-0,79262	-0,65262
Estonia	18,9	2	-0,7257	-0,79262	-0,65262
Panama	19	2	-0,7225	-0,79262	-0,65262
Macedonia	19	2	-0,7225	-0,79262	-0,65262
Israel	19,3	2	-0,7225	-0,79262	-0,64952
Trinidad and Tobago	19,3	2	-0,7225	-0,79262	-0,64952
Jamaica	19,4	2	-0,7192	-0,79262	-0,64643
Taiwan	19,6	2	-0,7192	-0,79262	-0,64643
Greece	19,7	2	-0,7192	-0,79262	-0,64643
Czech Republic	19,9	2	-0,7225	-0,79262	-0,64643
Australia	20	2	-0,7225	-0,79262	-0,64643
Austria	20	2	-0,7225	-0,79262	-0,64643

país	democ.area	grupo	mediana	IC 95% para a mediana	
				inf	sup
Belgium	20	2	-0,7225	-0,79262	-0,64643
Canada	20	2	-0,7225	-0,79262	-0,64643
Costa Rica	20	2	-0,7225	-0,79262	-0,64643
Cyprus	20	2	-0,7225	-0,79262	-0,64643
Denmark	20	2	-0,7225	-0,79262	-0,64643
Finland	20	2	-0,7225	-0,79262	-0,64643
Germany	20	2	-0,7225	-0,79262	-0,64643
Ireland	20	2	-0,7225	-0,79262	-0,64643
Italy	20	2	-0,7225	-0,79262	-0,64643
Japan	20	2	-0,7225	-0,79262	-0,64643
Lithuania	20	2	-0,7225	-0,79262	-0,64643
Mauritius	20	2	-0,7225	-0,79262	-0,64643
Mongolia	20	2	-0,7225	-0,79262	-0,64643
Netherlands	20	2	-0,7225	-0,79262	-0,64643
New Zealand	20	2	-0,7225	-0,79262	-0,64643
Norway	20	2	-0,7225	-0,79262	-0,64643
Portugal	20	2	-0,7225	-0,79262	-0,64643
Slovenia	20	2	-0,7225	-0,79262	-0,64643
Spain	20	2	-0,7225	-0,79262	-0,64643
Sweden	20	2	-0,7225	-0,79262	-0,64643
Switzerland	20	2	-0,7225	-0,79262	-0,64643
United Kingdom	20	2	-0,7225	-0,79262	-0,64643
United States	20	2	-0,7225	-0,79262	-0,64643

APÊNDICE D

Uma breve descrição do modelo BART

D Uma breve descrição do modelo BART

D.1 O modelo BART

O modelo BART (sigla inglesa para árvores de regressão bayesianas aditivas), publicado em 2010 no *The Annals of Applied Statistics*, foi proposto pelos pesquisadores Hugh A. Chipman, Edward I. George e Robert E. McCulloch, com o objetivo de introduzir uma nova metodologia para relacionar variáveis. A principal análise deste relatório utilizou as técnicas descritas neste artigo e por isso sugere-se fortemente sua leitura para o completo entendimento do mecanismo do modelo e seus recursos.

O modelo consiste no ajuste de **árvores de decisão binárias**, que levam as variáveis preditoras (no caso logaritmo do PIB, escore de democracia acumulado, etc.) em estimativas para a variável resposta (proporção de coalizões, soma ponderada, etc.). A principal novidade do BART em comparação com outros já introduzidos que utilizam árvores de decisão é a utilização de uma quantidade elevada de árvores que compõem as estimativas (na ordem de centenas), ao passo que nos modelos usuais, apenas algumas (ou mesmo uma) árvore é ajustada. A seleção de variáveis de decisão é feita aleatoriamente e, com uma grande quantidade de árvores, este modelo combina efeitos aditivos e efeitos de interação das características explicativas sem necessariamente especificá-las.

O BART é uma técnica inteiramente especificada pelo enfoque bayesiano e por isso envolve a especificação de uma distribuição a *priori*. O artigo discute o uso de uma *priori* reguladora padrão que funciona bem em grande parte das aplicações práticas e não influencia nos resultados. Na análise apresentada neste relatório esta *priori* foi deixada inalterada por não haver motivo para modificá-la. Em suma, a priori supõe que as árvores são geradas independentemente em todos os seus aspectos: quantidade de nós, profundidade, variáveis de decisão, valor de corte das variáveis de decisão.

Para ilustrar a estrutura de uma árvore, suponha que $g(x, T, M)$ seja uma função que atribua um valor μ para um certo $x = (x_1, \dots, x_n)$, onde T denota toda a estrutura da árvore e $M = (\mu_1, \dots, \mu_b)$ é o vetor de todos os terminais de T .

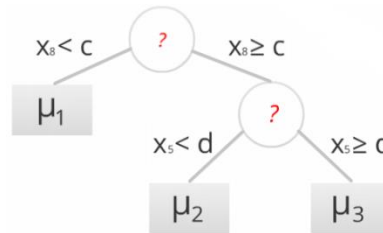


Figura D.1.1: Exemplo de uma árvore T com $M = (\mu_1, \mu_2, \mu_3)$ com profundidade igual a 2. Caso tenhamos um $x_{obs} = (\dots, x_5 = d, \dots, x_8 = c)$, $g(x_{obs}, T, M) = \mu_3$.

A soma de m árvores constitui a estrutura geral do modelo e cada nó é uma pequena contribuição à estimativa de $f(x) = E(y|x)$.

D.2 Seleção de modelos

Para a ilustração deste tópico, ver Figura C.2.5.

A seleção de modelos para o BART consiste numa técnica empírica, baseada na construção de um gráfico a partir de várias simulações do modelo. Neste gráfico, é observada a proporção de vezes que cada variável participa das árvores de decisão, isto é, a razão de quantas decisões baseia-se na variável pelo número total de decisões. Para cada simulação, o número m de árvores utilizadas é modificado.

Estudos realizados com o modelo BART (McCulloch, Chipman e George, 2010) revelam que, quanto maior o valor de m , mais as proporções tendem a se tornarem iguais. Para valores menores de m , no entanto, as variáveis que mais contribuem para a explicação da resposta participam com maior frequência. Como regra de corte para a seleção de variáveis no modelo, utilizaram-se as variáveis que possuíram proporção de uso acima da proporção esperada (caso todas as variáveis fossem igualmente importantes). Na Figura C.2.5 observa-se as 4 variáveis acima da linha de corte (em lilás).