

Apêndice

Este é o apêndice do artigo: “América Latina: a quantidade de acordos internacionais influencia o recebimento de investimento estrangeiro direto? Comparando análises nos dados”.

Apresenta-se a seguir a (1) análise exploratória dos dados, (2) as técnicas de tratamento utilizadas, (3) a análise dos pressupostos (4) o ajuste dos modelos, (5) modelos estimados e (6) referências

1) Análise exploratória dos dados

A análise exploratória dos dados procura oferecer um panorama dos dados e da relação entre as variáveis, revelando suas tendências e padrões, além de possíveis anomalias. Realiza a estatística descritiva dos dados.

O artigo apresenta como variável dependente a quantidade de investimento estrangeiro direto recebido como porcentagem do Produto Interno Bruto pelos países da América Latina entre 1970 e 2009, dados fornecidos pela UNCTAD e selecionados dessa forma por tornar possível a comparação ao longo do tempo entre os países, dispensando a deflação (BÜTHER e MILNER, 2008, p. 748), ou seja, a necessidade de um índice que isole o crescimento real do produto do artificial causado pelo aumento de preços na economia (IPEA, 2007).

A justificativa para a escolha dos dados da UNCTAD é que estes seriam menos afetados por omissões intencionais em seus relatos quando se trata de países em desenvolvimento (BÜTHER e MILNER, 2008, p. 748). A região América Latina refere-se a 20 países (MAINWARING e PÉREZ-LIÑÁN, 2009, p.532): Argentina, Bolívia, Brasil, Uruguai, Venezuela, Paraguai, Costa Rica, El Salvador, Guatemala, Honduras, México, Nicarágua, Panamá, Haiti, República Dominicana e Cuba.

O resumo da quantidade de IED como porcentagem recebida na América Latina varia de – 10,6749% do PIB de um país até 16,3549% do PIB de um país e tem como valor médio dos 20 países 1,8754% do PIB, ou seja, a média percentual de investimento estrangeiro direto na América Latina no período de 1970 a 2009 chega a quase 2% do PIB dos países. Ainda, o desvio-padrão da

variável é de aproximadamente 2,45 e o coeficiente de variação (d.p./média) aproxima-se 1,30 ou 130%, demonstrando uma grande variância entre os dados que pode ser observada pelo histograma. A maioria dos valores de IED concentram-se, no entanto, entre 0 e 4% do PIB dos países.

As variáveis independentes quantitativas utilizadas são: Número de PTAs em vigor, Número de BITs que um país é signatário, Índice de constrição política em um país - POLCON 2010, Instabilidade Política, Tamanho do Mercado, Desenvolvimento Econômico e Crescimento do PIB per capita. Ademais, usa-se a variável independente categórica Participação no GATT ou na OMC, após sua criação.

Número de PTAs em vigor

A quantidade de PTAs em vigor representa o número de acordos preferencias de comércio em vigor no intervalo de 1970 a 2009 para os países da América Latina. É importante ressaltar que o número total de tratados em vigor na data da publicação do banco de dados (2013) era de 2431 PTAs em vigor, comparado com 2305 PTAs do artigo de 2008. PTAs assinados somam 2659, sendo 228 assinados e não em vigor.

Um das hipóteses é de que quanto maior o número de PTAs assinados pelos países, maior será o recebimento de IED na região. O valor mínimo é 0, ou seja, nenhum acordo assinado e o valor máximo é o de 21, i.e., 21 acordos assinados e em vigor pelo Chile. A média de acordos assinados encontra-se em torno de 3 e o desvio-padrão em torno de 3 também. Ao plotar a variável no gráfico, percebe-se que a maioria dos países assinam até 3 acordos preferenciais por ano. Visualiza-se que a relação da quantidade de PTAs em vigor com o recebimento de IED em proporção ao PIB é positiva e concentrada em torno de 0 a 5 acordos.

Número de BITs

O número de tratados de investimentos bilaterais dos quais um país é signatário, fornecido pela UNCTAD, contém disposição específicas sobre o tratamento de investidores estrangeiros e sobre a definição de investimento (BÜTHE e MILNER, 2008, p.749), o que os diferencia em restritos ou não. Assim,

há uma diferença de 119 tratados quando são diferenciados em cumulativos e cumulativos restritos.

O número mínimo de acordo é 0 e o número máximo é o de 61 tratados, assinados por Cuba. A média de tratados fica perto de 9 e o desvio-padrão, um pouco acima de 13, sinalizando uma grande variância nos dados, com um coeficiente de variação em 1,56.

A maior frequência dos tratados ocorre até 5, demonstrando que, assim como PTAs, o número de BITs assinados por ano em cada país é baixo, o que é natural em virtude de sua complexidade. A relação da variável dependente com BITs é positiva e concentrada no início do gráfico, indicando uma grande quantidade de pouco BITs firmados por ano e o impacto desses BITs em IED encontra-se em uma faixa percentual de 0 a 5% do PIB dos países.

Índice de constrição política - POLCON

A outra variável é o índice “Political Constraint Index” (POLCON), produzido por Witold J. Henisz para medir constrições políticas em um país, identificando estruturas políticas secundárias e capacidade de apoio a compromissos políticos críveis (BÜTHE e MILNER, 2008, p.749). Considera-se a medida ponderada no número de players com poder de veto no sistema político nacional. O índice foi publicado em 2000 e este artigo usa o POLCON 2010, que possui dados até 2007.

O índice vai de 0 a 1, estabelecendo constrições políticas domésticas em uma escala em que 0 representa um alto nível de risco e instabilidade política e 1, baixo nível de risco político. Neste *dataset*, Cuba tem um POLCON de 0, enquanto o Brasil, no início dos anos 90, atinge o valor máximo para os países da América Latina dentro do intervalo temporal. Pelo índice, percebe-se que a média dos países para o índice fica em torno de 0.3, demonstrando um alto nível de constrições políticas domésticas na região. A relação com a variável dependente é levemente positiva, com a maior parte dos dados concentrados logo acima da média do índice na faixa de intervalo de recebimento de IED de até 5% do PIB.

Instabilidade Política

A próxima variável é instabilidade política, medida composta dos dados de Arthur Banks (1999) para eventos políticos que indicam violência política e instabilidade. A base apresenta-se como dados transnacionais de séries temporais para mais de 200 países abrangendo variáveis econômicas, políticas, populacionais e demográficas, com ênfase para os dados de regimes e sistemas políticos (BÜTHE e MILNER, 2008, p.749).

O valor máximo é do índice para os países é 33 e o mínimo é 0, com a média em torno de 3,5, ou seja, a maior parte dos países possuem baixa instabilidade política. O valor máximo foi para a Argentina coincidindo com a morte do então presidente Perón. O desvio-padrão chega a 4,66 e o coeficiente de variação em 1,37, significando grande variação nos dados. Ao plotar-se a frequência da variável, vê-se que ela está toda concentrada à esquerda do gráfico, representando a baixa instabilidade política dos países da região. Sua relação com a variável dependente é levemente negativa, ou seja, quando maior a instabilidade política, menor será a quantidade de IED recebida.

Segue-se a seguir as três variáveis padrão de modelos econômicos de controle: tamanho do mercado, representado pela log da população do país; desenvolvimento econômico, representado pelo log do PIB – valor total de bens e serviços finais produzidos na economia, normalmente em um período de um ano (KRUGMAN e WELLS, 2007, p.488) - per capita em dólares do ano 2000; e crescimento do PIB, representando pela a variação percentual no PIB real do país em relação ao ano anterior (BÜTHE e MILNER, 2013, p.102). A base de dados original foi atualizada com dados de 2010 fornecidos pelo Banco Mundial.

Tamanho do mercado

O tamanho do mercado nos países é relativamente próximo. O valor mínimo do log da população é 14,23, enquanto o valor máximo é 19,08. A média fica em torno de 16 e o desvio-padrão em torno de 1, o que é reafirmado pelo coeficiente de variação, também abaixo de 1. Ao plotar-se no histograma, apenas reafirma-se a distribuição com a maior concentração populacional em torno da média. A relação entre a variável dependente e independente apresenta-se como quase inexistente, levemente negativa. A faixa de IED no PIB permanece concentrada entre 0 e 5% com a média populacional entre 15 e 17.

Desenvolvimento Econômico

Desenvolvimento econômico também é próximo entre os países, apresentando-se 5,93 como valor mínimo do log do PIB per capita e 9,2 como valor máximo. A média fica em torno de 7,6 e tanto o desvio-padrão como coeficiente são menores que 1, percebendo-se uma variação pequena os dados. Ao plotar-se no histograma, a concentração em torno da média é visível e, novamente, a relação entre a variável dependente e independente apresenta-se como quase inexistente, levemente positiva, sugerindo que quanto maior o desenvolvimento econômico, maior seria o recebimento de IED. Os dados em sua maioria concentram-se na faixa de desenvolvimento de 7 a 8 log do PIB per capita.

Crescimento do PIB

O crescimento do PIB, auferido em variação percentual apresenta grande variação entre os países indo de -26,479 (valor mínimo), ou seja, um crescimento negativo, demonstrando que o país produziu menos quando comparado ao ano anterior. O PIB real serve para determinar o crescimento do produto agregado tomando por base algum ano-base em contraste ao PIB nominal que é calculado a preços correntes, “a medida do produto agregado médio por pessoa” (KRUGMAN e WELLS, 2007, p.497).

O valor máximo é de 18,28, ficando a média em 3,44. O desvio-padrão é de 4,5 e o coeficiente de variação de 1,31, ou seja, há grande variância no crescimento do PIB entre os países da região. Ao plotarmos no histograma, percebe-se que a maioria das variações são entre 0 e 10%. Além disso, haveria uma relação positiva entre recebimento de IED e crescimento do PIB, sugerindo que quanto maior o crescimento do PIB, maior o recebimento de IED. No gráfico, percebe-se uma concentração massiva dos dados entre 0 e 10% para a variação percentual no PIB e de 0 para cima em relação ao recebimento de IED.

Participação no GATT/OMC

Por fim, tem-se a variável independente categórica GATT/OMC, computando a participação dos países no GATT/OMC como 1 e 0 como não. A ideia é que a participação nesses organismos multilaterais promova maior

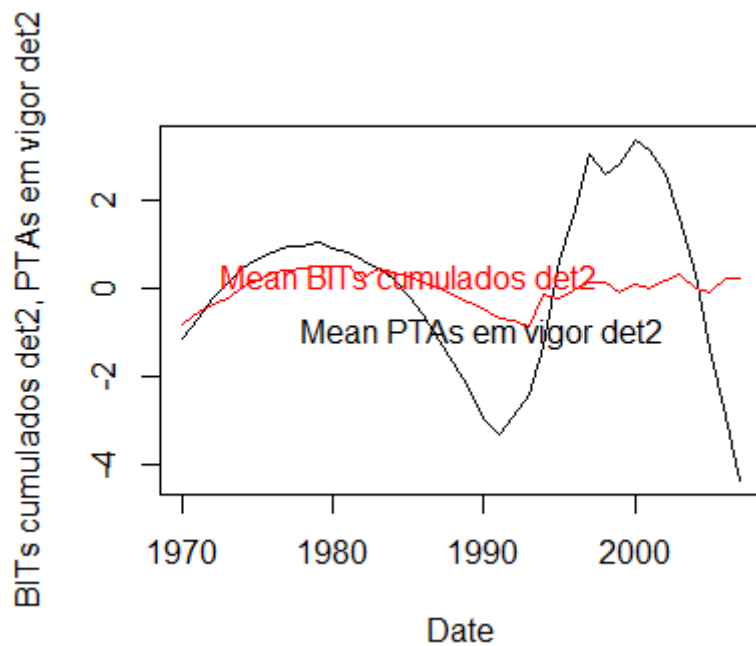
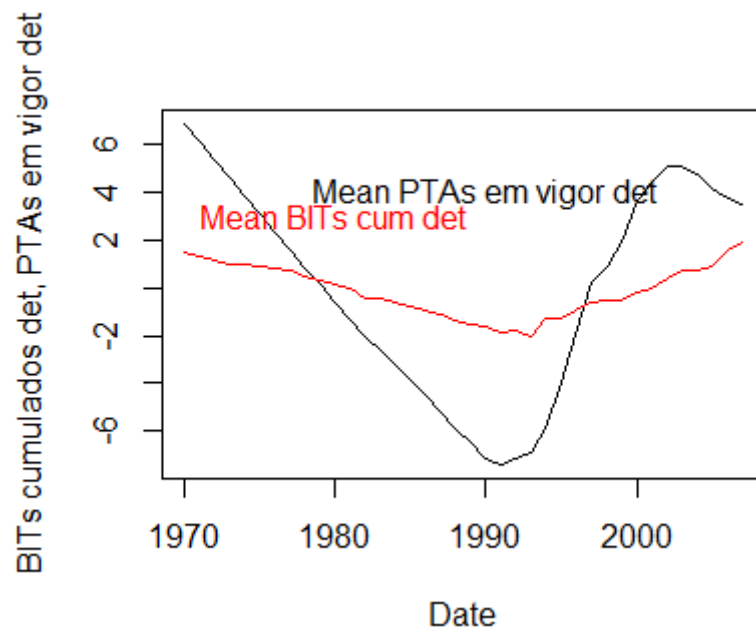
entrada de IED nos países da região. Ao todo são 535 aferições, sendo 535 o número de países participantes por ano e 250, o número de países não participantes. Pela análise do boxplot, constata-se uma maior variação no número de países participantes, mas a mediana em ambos é quase a mesma, não indicando visualmente grande influência de IED na distribuição de participação em GATT/OMC.

2) Técnicas de tratamento

Os dados utilizados neste artigo são séries temporais para cada membro em um conjunto de dados, neste caso, países. Os dados foram coletados para as mesmas variáveis ao longo do tempo e para um mesmo grupo de países, países da América Latina. Trata-se, então, de dados em painel. (WOOLDRIDGE, 2002, p. 6-10).

Séries temporais são séries com observações de variáveis ao longo do tempo. Como eventos passados frequentemente influenciam eventos futuros, não é possível considerar as observações colhidas como independentes ao longo do tempo. Quando se estabelece no tempo e apresenta uma tendência de crescimento, esta tendência pode levar a inferências que não correspondam a realidade, induzindo a correlações entre tendências positivas, por exemplo, que estão ligadas a fatores não observados (WOOLDRIDGE, 2002, p. 331).

Como neste artigo, isto é exatamente o que se observa, ou seja, o recebimento de IED e o número de tratados internacionais crescem, as técnicas para tratamento de dados utilizadas foram um *detrend* linear e um *detrend* quadrático. A seguir um exemplo de *detrend* linear seguido de quadrático:



A análise comparada de dados na adoção de uma estratégia de *detrend* linear por parte de Büthe e Milner (2008) foi criticada por King e Roberts (2014) sob a alegação de que os autores realizaram o *detrend* linear em tendências quadráticas e aplicaram a mesma tendência para todos os países, contrária a diversidade de países em desenvolvimento de que trata seu artigo, o que geraria um novo padrão espúrio na série-temporal (KING e Roberts, 2014, p. 174).

A solução de King e Roberts (2014) é o *detrend* quadrático para cada país, o que geraria resultados nas variáveis mais próximos de um modelo estacionário aproximando com erros padrões robustos e clássicos no modelo. O resultado fundamental é que após a mudança na técnica, as variáveis de interesse deixaram de ser significativamente correlacionadas com o recebimento de FDI, i.e., o maior número de PTAs, por exemplo, não influenciaria a quantidade de recebimento de IEDs de maneira significativa (KING e ROBERTS, 2014, p. 174).

Este artigo, então, gera 3 modelos e compara-os sem *detrend*, com *detrend* linear e *detrend* quadrático para perceber o efeito das técnicas do modelo em um contexto regional em que o arcabouço teórico é próximo. Testa-se se mantendo as técnicas usadas pelos autores inicialmente para um *subset* do banco de dados, os resultados mantem-se ou devido às proximidades de contexto, os dados contextualmente mais próximos possibilitarão uma correlação significativa entre as variáveis independente da técnica.

Realizou-se o *detrend* linear pelo pacote *pracma* do R que especificamente “calcula o ajuste dos mínimos quadrados de uma linha reta (ou linha composta para tendências lineares por partes) nos dados e subtrai a função resultante dos dados” (R DOCUMENTATION, 2019, tradução minha). A diferença principal entre o *detrend* quadrático e o linear é que para o primeiro assume-se um comportamento exponencial dos dados e adiciona-se o tempo ao quadrado.

Sobre as variáveis, é importante ressaltar as modificações do banco do artigo original, levantado pelos autores e usado por Gary King para a contestação do modelo, e do banco utilizado neste artigo. Os dados utilizados neste artigo vêm de um artigo de Büthe e Milner em que o modelo 4 é a base para o artigo e a partir dele, os autores atualizam as variáveis em tempo e interpretação, o que também foi realizado aqui.

As principais alterações são na precisão da escolha da variável dependente, na atualização do índice do POLCON e na atualização da medição de desenvolvimento econômico, representado pelo log do PIB per capita em dólares. Sobre a variável dependente, o artigo usa PTAs em vigor, assinados e

ratificados, e não apenas PTAs assinados, sob a justificativa de que apenas os acordos efetivamente em vigor teriam um impacto no recebimento de IED (BÜTHE e MILNER, 2014, p. 103). Em relação ao POLCON, atualiza-se seus dados para 2010 ao invés de 2002, utilizado no primeiro modelo; e em relação a desenvolvimento econômico, usa-se o log do PIB de 2000 no lugar do de 1995.

As variáveis também estão defasadas sob a justificativa de qualquer efeito das variáveis independentes na variável dependente pode levar algum tempo para ser percebida e, conseqüentemente, mensurada. A opção do modelo é, então, defasar as variáveis independentes em 1 ano (BÜTHE e MILNER, 2008, p.748). Assume-se também que o valor de IED é determinado por seu valor passado.

Outro conceito importante para o modelo é o de processo estacionário para séries temporais. Este ocorre quando as distribuições das probabilidades mantêm-se estáveis ao longo do tempo, indicando que uma sequência temporal para determinado valor é igualmente distribuída e a natureza das correlações entre os termos permanece a mesmo em todos os períodos de tempo (WOOLDRIDGE, 2002, p. 348).

Um modelo que produza resíduos próximo a um processo estacionário, aproxima erros padrão robustos – usados para corrigir o modelo de estimação - dos clássicos, indicando maior precisão para o modelo. Distância entre ambos, podem indicar falhas e má especificação na adoção do modelo estatístico (KING et ROBERTS, 2014, p.159).

Assim, os erros padrão robustos servem a possibilitar uma adequação ao pressupostos de um modelo linear, por exemplo, que garantam, ao máximo, consistência e ausência de vieses, ainda que a produção dos dados não se encaixe, a priori, no modelo (KING et ROBERTS, 2014, p.161).

Isto exposto, segue-se analisando os pressupostos dos 3 modelos.

3) Análise dos pressupostos

A análise de regressão múltipla serve à análise da relação de uma variável dependente com várias variáveis independentes, estimando-se o grau

de associação que pode haver entre elas (HAIR et al., 2014). A notação do modelo em sua forma mais simples apresenta a seguinte fórmula:

$$Y = \alpha + \beta_1 X_1 + \epsilon$$

Y é a variável dependente e o que queremos explicar no modelo, X1 acima, representa a variável explicativa do modelo que em sua fórmula multivariada seria composta de mais de um termo, e analisa-se a influência de cada variável no modelo. Alfa, o intercepto, representa o valor de Y quando X1 é zero, β representa o aumento marginal de X1 em Y e ϵ representa a diferença entre os valores observados e os preditos, os erros ou resíduos do modelo.

Estes têm pressupostos que devem ser analisados nos dados para que se possa adotar a regressão linear multivariada, pressupondo-se que quanto menor o resíduo, mais próximo está o modelo da realidade. Elenca-se a seguir os que serão analisados neste apêndice, ressaltando-se que não há uniformidade sobre quais ou quantos pressupostos são necessários entre os autores para a validação de um modelo (FILHO et al., 2011, p.51):

- (1) a expectativa da média do termo de erro é igual a zero;
- (2) homocedasticidade, ou seja, a variância do termo de erro é constante para os diferentes valores da variável independente;
- (3) as variáveis independentes não apresentam alta correlação, o chamado pressuposto da não multicolinearidade;
- (4) assume-se que o termo de erro tem uma distribuição normal;
- (5) há uma adequada proporção entre o número de casos e o número de parâmetros estimados;
- (6) nenhuma variável teoricamente relevante para explicar Y foi deixada de fora do modelo e nenhuma variável irrelevante para explicar Y foi incluída no modelo;
- (7) as variáveis foram medidas adequadamente, ou seja, assume-se que não há erro sistemático de mensuração; e
- (8) a relação entre a variável dependente e as variáveis independentes deve ser linear.

Explica-se a função teórica de cada pressuposto no modelo seguida de sua aplicação nos modelos do artigo. O modelo adotado é:

$$IED_{it} = \alpha + \gamma_1 (\text{Tamanho do Mercado})_{i(t-1)} + \gamma_2 (\text{Des. Econ.})_{i(t-1)} + \gamma_3 (\text{Cresc. do PIB})_{i(t-1)} + (\text{PTAs em vigor})_{i(t-1)} + (\text{GATT/WTO})_{i(t-1)} + (\text{BITs cumul.})_{i(t-1)} + (\text{POLCON})_{i(t-1)} + (\text{Instab. Pol.})_{i(t-1)} + \delta_i + \varepsilon_{it},$$

em que δ_i indica efeitos fixos do país implementados por meio de um conjunto de $n - 1$ dummies dos países.

O primeiro pressuposto estabelece que o valor médio do erro deve ser zero, indicando que valores de erro não incluídos no modelo não afetaram seu resultado, porque os valores anulam-se ao serem equidistantes. Nos três modelos este pressuposto passa, com a média dos resíduos muito próxima de 0.

O segundo pressuposto, homocedasticidade, pressupõe que os resíduos variam uniformemente no modelo, caso não o façam, interferem no intervalo de compatibilidade e testes de significância do modelo. Uma das causas de heterocedasticidade é erro de mensuração nas variáveis independentes (TABACHICK E FIDELL, 2007, *in* FILHO et al., 2011).

A homocedasticidade é garantida pela aplicação da função de Arellano (1987) na geração dos modelos 2 e 3. Como após a aplicação da função não é possível testar o comportamento dos resíduos, avalia-se os pressupostos dos modelos 2 e 3 antes da aplicação da função e verifica-se que antes dela não há homocedasticidade nos dados.

Após a aplicação, assume-se que a função é robusta para heterocedasticidade e autocorrelação (o valor de uma variável auferido em determinado momento não afeta o valor de uma medição posterior). Antes da aplicação da função, porém, os testes realizados têm um p-valor menor que um nível de 0,01, indicando que a variância dos resíduos não se dá de forma uniforme, o que é confirmado pela análise gráfica.

O terceiro pressuposto refere-se à multicolinearidade. “Multicolinearidade refere-se à correlação excessiva entre as variáveis preditoras” (GARSON, 2011 *in* FILHO et al. 2011). Essa correlação excessiva gera informação redundante sobre a resposta do modelo. A técnica de análise utilizada para medição de

multicolinearidade é *Variance Inflation Factor* (VIF), indicando que quanto maior, pior, ou seja, maior a correlação das variáveis sob condições ideais.

O valor limite do VIF para os modelos varia entre os autores. Adotando-se 4 (HAIR et. al, 2014), analisa-se os três modelos, percebendo-se que só o 3 apresentaria ausência de multicolinearidade com todos os VIFs abaixo de 2. Para o primeiro modelo, há alta multicolinearidade entre Tamanho do Mercado e Desenvolvimento Econômico. Para o segundo modelo, apenas para Desenvolvimento Econômico. Para o terceiro, não há.

O quarto pressuposto é a distribuição normal dos resíduos. Nos três modelos percebe-se que esse pressuposto é atendido. Observa-se tanto pela análise gráfica, quanto pelo teste de normalidade Shapiro que os modelos apresentam distribuição normal dos resíduos. O teste Shapiro nos três modelos, deu resultados para o p-valor abaixo de 0,001, indicando um forte indício na normalidade da distribuição dos resíduos. Esta é uma premissa do teorema de Gauss-Markov para garantir que os estimadores sejam não-viesados e eficientes (FILHO et al, 2011, p. 59).

O quinto pressuposto requer uma proporção mínima entre o número de casos e os estimadores. Utilizando a fórmula de Tabachnick e Fidell (2007 *in* FILHO, 2011) em que $N \geq 50 + 80X$ (sendo X o número de variáveis independentes incluídas na análise), calcula-se que para o modelo o número mínimo de observações seria 690. A análise considera 785 observações.

O sexto pressuposto requer uma especificação adequada do modelo com a justificativa teórica para inclusão das variáveis relevantes ao modelo e exclusão das irrelevantes, para que não se aumente o erro nas estimativas.

O sétimo pressuposto assume ausência de erro de medição. Como os dados utilizados são secundários de fontes com respaldos de credibilidade, tais quais UNCTAD, Banco Mundial, além de índice usualmente adotados na literatura, assume-se que não houve erro na medição das variáveis.

Por fim, o oitavo pressuposto estabelece que haja a linearidade dos parâmetros, i.e., que haja uma relação linear entre as variáveis do modelo. Sobre a associação das variáveis dependentes quantitativas com a variável

independente, utilizando-se a análise gráfica e o teste de correlação de Pearson com a indicação do p-valor, tem-se que para o:

1 - Número de PTAs em vigor - há uma relação linear com os dados, que se concentram mais à esquerda do gráfico, diminuindo o intervalo de compatibilidade da amostra em relação ao lado do esquerdo. O valor p do teste é menor que o nível de significância de 0,05, indicando que há uma correlação entre as variáveis, porém a relação linear é positiva e moderada, sendo 0,47.

2 - Número de BITs - há uma relação linear com os dados, que se concentram mais à esquerda do gráfico, diminuindo o intervalo de compatibilidade da amostra em relação ao lado do esquerdo. O valor p do teste é menor que o nível de significância de 0,05, indicando que há uma correlação entre as variáveis, porém, a relação linear é positiva e moderada, sendo 0,33.

3 - Índice de constrição política - POLCON - há uma relação linear com os dados, que se concentram mais no meio do gráfico, diminuindo o intervalo de compatibilidade da amostra em relação ao meio. O valor p do teste é menor que o nível de significância de 0,05, indicando que há uma correlação entre as variáveis, porém a relação linear é positiva e fraca, sendo 0,24.

4 - Instabilidade Política - há uma relação linear com os dados, que se concentram mais à esquerda do gráfico, diminuindo o intervalo de compatibilidade da amostra em relação ao lado do esquerdo. O valor p do teste é menor que o nível de significância de 0,05, indicando que há uma correlação entre as variáveis, porém a relação linear é negativa e fraca, sendo - 0,12.

5 - Tamanho do mercado - a relação linear entre as variáveis é muito próxima de 0, próxima à inexistência de relação. O valor p do teste é menor que o nível de significância de 0,05, indicando que há uma correlação mínima entre as variáveis, porém a relação linear é negativa e muito próxima de 0, sendo - 0,083.

6 - Desenvolvimento Econômico - a relação linear entre as variáveis é muito próxima de 0, próxima à inexistência de relação. O valor p do teste é menor que o nível de significância de 0,05, indicando que há uma correlação

mínima entre as variáveis, porém a relação linear é positiva e muito próxima de 0, sendo 0,092.

7 - Crescimento do PIB - há uma relação linear com os dados, que se concentram mais no meio do gráfico, diminuindo o intervalo de compatibilidade da amostra em relação ao meio. O valor p do teste é menor que o nível de significância de 0,05, indicando que há uma correlação entre as variáveis, porém a relação linear é positiva e fraca, sendo 0,19.

Sobre os outliers, o modelo 1 apresenta os pontos no início e no final da curva fora do intervalo de compatibilidade, indicando a necessidade de uma regressão mais robusta para corrigir possíveis efeitos de outliers (FOX, 2008 *in* FILHO, 2011). O modelo 2, com o *detrend* linear nas variáveis, consegue ajustar-se melhor a curva do intervalo de compatibilidade, com poucos valores fora dele em relação ao modelo 1. O modelo 3, com o *detrend* quadrático nas variáveis, também adequa-se a reta em sua maior parte, sendo muito próxima sua distribuição ao modelo 3.

4) Ajuste dos modelos

Avaliando a capacidade explicativa do modelo 1, tem-se que o *R-squared* informa o valor percentual para o conteúdo explicativo do modelo, de modo que quanto mais alto, melhor o modelo se ajusta aos dados. A capacidade explicativa do modelo é de 34,6%, i.e., explica-se 34,6% da variação percentual do PIB em recebimento de IED.

O erro quadrático médio (RMSE) fornece a medida da precisão média do modelo com a métrica da variável dependente (KELLSTEDT e WHITTEN, 2009, p. 201). Apresenta-se na mesma unidade que a variável dependente, i.e., para o modelo 1, em média, a variação dos valores preditos no modelo é de 1,98 percentual em relação a porcentagem do PIB de recebimento de IED nos países em desenvolvimento da América Latina.

Para o modelo 2, observa-se uma diminuição da capacidade explicativa do modelo, nele de 19,86%, ou seja, explica-se agora, após o *detrend linear* quase 20% da variação percentual do PIB em recebimento de IED. O RMSE também diminui, em média, a variação dos valores preditos no modelo 2

é de 1,66 percentual em relação a porcentagem do PIB de recebimento de IED nos países em desenvolvimento da América Latina.

Para o modelo 3, observa-se outra diminuição da capacidade explicativa do modelo, aqui de 6,98%, ou seja, explica-se agora, após o *quadrático* apenas 7% da variação percentual do PIB em recebimento de IED. Ocorre, porém, um aumento da qualidade da precisão média do modelo para a variação da variável dependente, passando a 1,71 percentual em relação a porcentagem do PIB de recebimento de IED nos países em desenvolvimento da América Latina, conseguindo explicar uma maior variação no recebimento de IED em relação ao 1.

Interessante notar como o tratamento das variáveis afeta a capacidade explicativa do modelo, sugerindo que a tendência nos dados aumenta, e muito, o poder explicativo de um modelo pela influência do tempo em potencializar as relações entre as variáveis.

Em relação ao intervalo de compatibilidade dos modelos, demonstrando uma probabilidade de 95% de que os valores populacionais estejam contidos na amostra, muitos cruzam o valor 0, indicando que o valor p para o teste de hipótese de que os parâmetros serão iguais a 0, será maior que 0,05. As variáveis que mantêm efeitos próximos nos três modelos são: PTAs em vigor, GATT/OMC. BITs no terceiro modelo cruzam o 0, o que não acontece nos dois primeiros.

O comportamento dos parâmetros no modelo 2 e 3 é próximo, coincidindo para a maioria dos parâmetros e denotando maior precisão. O comportamento dos parâmetros no modelo 1, apresenta faixas de intervalo maiores, mas que também englobam a dos outros modelos, indicando que não estão em conflito, ainda que para alguns haja significância estatística e para outros não (NATURE, 2019).

5) Modelos estimados:

Modelo 1:

```
Call:
lm(formula = fdi_inflow_unctad_gdp ~ lag_pta_force + lag_gattwto +
```

```

lag_bits_cum1_restricted + lag_polconiii_2010 + lag_polinstability
+ lag_ln_pop + lag_ln_gdp_pc_00d + lag_gdp_gr + factor(country),
data = fdi)

```

Residuals:

	Min	1Q	Median	3Q	Max
	-14.4076	-0.7973	-0.1847	0.5928	12.7539

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.19905	11.40555	-0.280	0.779196	
lag_pta_force	0.19657	0.06272	3.134	0.001802	**
lag_gattwto	0.42357	0.29195	1.451	0.147303	
lag_bits_cum1_restricted	0.04530	0.01247	3.632	0.000303	***
lag_polconiii_2010	0.23047	0.48457	0.476	0.634506	
lag_polinstability	-0.02036	0.01640	-1.241	0.215024	
lag_ln_pop	0.33486	0.63459	0.528	0.597902	
lag_ln_gdp_pc_00d	-0.33923	0.51563	-0.658	0.510840	
lag_gdp_gr	0.07489	0.01796	4.169	3.47e-05	***
factor(country)2	2.00118	1.46427	1.367	0.172195	
factor(country)3	-0.30270	1.17015	-0.259	0.795961	
factor(country)4	1.29320	0.89018	1.453	0.146775	
factor(country)5	0.04184	0.87915	0.048	0.962060	
factor(country)6	0.67395	1.18618	0.568	0.570114	
factor(country)7	0.35336	1.54057	0.229	0.818654	
factor(country)8	0.19406	0.84050	0.231	0.817479	
factor(country)9	0.86603	1.55413	0.557	0.577551	
factor(country)10	0.08278	0.66497	0.124	0.900968	
factor(country)11	1.92583	1.61521	1.192	0.233570	
factor(country)12	0.96730	1.35256	0.715	0.474764	
factor(country)13	0.70167	1.21141	0.579	0.562641	
factor(country)14	1.06780	1.57389	0.678	0.497727	
factor(country)15	0.25921	0.92029	0.282	0.778294	
factor(country)16	1.04203	1.75473	0.594	0.552822	
factor(country)17	4.14504	1.72166	2.408	0.016334	*
factor(country)18	-0.70279	1.76808	-0.397	0.691139	
factor(country)19	1.32009	1.21924	1.083	0.279330	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.023 on 655 degrees of freedom
(103 observations deleted due to missingness)

Multiple R-squared: 0.346, Adjusted R-squared: 0.32
F-statistic: 13.33 on 26 and 655 DF, p-value: < 2.2e-16

Modelo 2:

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.450324	0.247650	1.8184	0.0696704	.
lag_pta_force_det	0.163945	0.115116	1.4242	0.1550934	
lag_gattwto_det	0.994567	0.456922	2.1767	0.0300259	*
lag_bits_cum1_restricted_det	0.067051	0.024562	2.7299	0.0065843	*
lag_polconiii_2010_det	0.823389	0.966460	0.8520	0.3946885	
lag_polinstability_det	-0.042139	0.019238	-2.1905	0.0290051	*
lag_ln_pop_det	-0.297094	12.467785	-0.0238	0.9809996	
lag_ln_gdp_pc_00d_det	-0.651725	1.149122	-0.5672	0.5708952	
lag_gdp_gr_det	0.057957	0.018577	3.1199	0.0019260	*
factor(country)2	0.297167	0.062300	4.7699	2.496e-06	*
factor(country)3	-0.457498	0.082163	-5.5682	4.438e-08	*
factor(country)4	0.050243	0.163649	0.3070	0.7589715	


```

factor(country)5          -0.606043    0.098800   -6.1340  1.882e-09 *
**
factor(country)6          -0.596545    0.188756   -3.1604  0.0016821 *
*
factor(country)7           0.197158    0.145435    1.3556  0.1758942
factor(country)8          -0.197233    0.122418   -1.6111  0.1078502
factor(country)9          -0.906327    0.112083   -8.0862  5.778e-15 *
**
factor(country)10         -0.021891    0.128181   -0.1708  0.8644751
factor(country)11         -0.830202    0.152714   -5.4363  8.944e-08 *
**
factor(country)12         -0.596262    0.128033   -4.6571  4.231e-06 *
**
factor(country)13         -0.194000    0.079306   -2.4462  0.0148188 *
factor(country)14         -0.800020    0.118193   -6.7688  4.073e-11 *
**
factor(country)15         -0.379589    0.193240   -1.9643  0.0501065 .
factor(country)16         -0.640420    0.172980   -3.7023  0.0002403 *
**
factor(country)17         -2.396916    0.109052  -21.9795 < 2.2e-16 *
**
factor(country)18          3.434957    6.921825    0.4963  0.6199607
factor(country)19         -0.678059    0.070225   -9.6556 < 2.2e-16 *
**
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Modelo 3:

t test of coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.23104520 0.15411785  1.4991 0.134326
pta_det2       0.29368017 0.09313216  3.1534 0.001689 **
gattwto_det2   -0.02158015 0.50063503  -0.0431 0.965631
bits_det2      0.01927869 0.02303007  0.8371 0.402842
pol_det2       0.18769249 0.67763846  0.2770 0.781884
lag_polinstability -0.02735114 0.01824448  -1.4991 0.134326
pop_det2       0.84902951 1.87878641  0.4519 0.651491
gdppc_det2     0.01229045 0.67963908  0.0181 0.985578
gdprgr_det2    0.06851225 0.01702592  4.0240 6.403e-05 ***
factor(country)2 -0.08763649 0.05845760  -1.4991 0.134326
factor(country)3 -0.11804178 0.07873934  -1.4991 0.134326
factor(country)4 -0.10076738 0.06721651  -1.4991 0.134326
factor(country)5 -0.04174648 0.02784684  -1.4991 0.134326
factor(country)6 -0.14529566 0.09691894  -1.4991 0.134326
factor(country)7 -0.19505685 0.13011196  -1.4991 0.134326
factor(country)8 -0.10242089 0.06831948  -1.4991 0.134326
factor(country)9 -0.17994174 0.12002948  -1.4991 0.134326
factor(country)10 -0.15618943 0.10418559  -1.4991 0.134326
factor(country)11 -0.21736962 0.14499561  -1.4991 0.134326
factor(country)12 -0.11732202 0.07825922  -1.4991 0.134326
factor(country)13 -0.07285209 0.04859572  -1.4991 0.134326
factor(country)14 -0.17490337 0.11666866  -1.4991 0.134326
factor(country)15 -0.18887885 0.12599095  -1.4991 0.134326
factor(country)16 -0.15978827 0.10658618  -1.4991 0.134326
factor(country)17 -0.17431690 0.11627745  -1.4991 0.134326
factor(country)18  0.00052783 0.00035209   1.4991 0.134326
factor(country)19 -0.16482664 0.10994701  -1.4991 0.134326
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

```

6) Referências

ARELLANO, M. **Computing Robust Standard Errors for Within-groups Estimators**. Oxford Bulletin of Economics and Statistics 49(4): 1987, p. 431–34. Disponível em: https://www.cemfi.es/~arellano/OBES_1987.pdf. Acesso em 02.09.2019.

BANKS, ARTHUR S. **Cross-National Time-Series Data Archive**. 1999. Binghamton, NY: Banner Software. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/K2SSTK>. Acesso em 02.09.2019.

BÜTHE, TIM E MILNER, HELEN V. **The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI through International Trade Agreements?** American Journal of Political Science, Vol. 52, No. 4, October 2008, p. 741–762. Disponível em: https://people.duke.edu/~buthe/downloads/ButheMilner_AJPS_Oct2008.pdf. Acesso em 02.09.2019.

_____. **Foreign Direct Investment and Institutional Diversity in Trade Agreements: Credibility, Commitment, and Economic Flows in the Developing World, 1971–2007**. World Politics, 66, 2014, pp 88-122. doi:10.1017/S0043887113000336. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1796506. Acesso em 02.09.2019.

FILHO, D. F., NUNES, F., ROCHA, E. C. DA, SANTOS, M. L., BATISTA, M., & JÚNIOR, J. A. S. **O Que Fazer e o Que Não Fazer Com a Regressão: pressupostos e aplicações do modelo linear de Mínimos Quadrados Ordinários (MQO)**. 2011. Revista Política Hoje - ISSN: 0104-7094, 20(1). Disponível em: <https://periodicos.ufpe.br/revistas/politica hoje/article/view/3808>. Acesso em 02.09.2019.

HAIR JR., JOSEPH; ANDERSON, RALPH; TATHAM, RONALD; BLACK, WILLIAN. **Multivariate data analysis**. 17ª Edição. Pearson Education Limited.

2014. Disponível em: https://is.muni.cz/el/1423/podzim2017/PSY028/um/_Hair_-_Multivariate_data_analysis_7th_revised.pdf. Acesso em 30.08.2019.

IPEA (Instituto de Pesquisa Econômica Aplicada) por SOUZA, JORGE LUIZ DE. O que é? Deflator implícito. Desafios do Desenvolvimento. Ano 4. Edição 37. 2007. Disponível em: http://desafios.ipea.gov.br/index.php?option=com_content&view=article&id=2140:catid=28&Itemid=23 . Acesso em 02.09.2019.

KELLSTEDT, PAUL M. e WHITTEN, GUY D. **Fundamentos da Pesquisa em Ciência Política**. (2ª ed.) Blucher, 2009.

KING, GARY E ROBERTS, MARGARET E. **How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It**. Political Analysis 23, 2015, p.159–179. doi:10.1093/pan/mpu015. Disponível em: https://gking.harvard.edu/files/gking/files/robust_0.pdf. Acesso em 20.08.2019.

KRUGMAN, PAUL e WELLS, ROBERTS. **Introdução à Economia**. Rio de Janeiro, Elsevier: 2007.

NATURE. AMRHEIN, VALENTIN; GREENLAND, SANDER; MCSHANE, BLAKE. Scientists rise up against statistical significance. 20 March 2019. Disponível em: <https://www.nature.com/articles/d41586-019-00857-9>. Acesso em 02.09.2019.

MAINWARING, SCOTT E PÉREZ-LIÑÁN, ANIBAL. **¿Por qué son importantes las regiones? Especificidades regionales y difusión de la democracia a escala regional**. Revista SAAP (ISSN 1666-7883) Vol. 3, Nº 3, agosto 2009, p. 523-566. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=5238129>. Acesso em 22.08.2019.

R DOCUMENTATION. **Detrend**. De pracma v1.9.9 por HwB. Disponível em: <https://www.rdocumentation.org/packages/pracma/versions/1.9.9/topics/detrend> Acesso em 18.08.2019.

WOOLDRIDGE, J. **Introductory Econometrics: A Modern Approach** (2^o ed).
South-Western College Pub. 2002.