

Capítulo 7

Séries temporais e análise comparada em Ciência Política e Relações Internacionais

Natália Poiatti (IRI-USP) e Pedro Feliú (IRI-USP)

Introdução

Muitos dos fenômenos estudados em ciência política (CP) e relações internacionais (RI) variam de maneira significativa ao longo do tempo. Muitas vezes, estamos interessados em compreender mudanças no objeto de pesquisa no decorrer do tempo (Pevehouse e Brozek, 2008), testar os efeitos de uma variação temporal em outra ou utilizar dados passados para prever eventos futuros (Brandt, Freeman e Schrodt, 2011). A análise de séries temporais (*time series*) permite testar hipóteses calcadas na influência do tempo no comportamento das observações, organizando os dados sequencialmente ao longo de uma série temporal (Barberia, 2019). Um atributo singular deste tipo de estruturação dos dados é a modelagem da dependência das observações vizinhas, auxiliando a teorização de eventos dinâmicos (Ehlers, 2005).

Séries temporais são muito utilizadas em distintas agendas de pesquisa nas áreas de CP e RI, constituindo-se em uma das ferramentas metodológicas primordiais na formação das pesquisadoras e dos pesquisadores desses campos. Nas relações internacionais, por exemplo, muitos estudos quantitativos analisaram os países em séries temporais por díade (Maoz e Russett, 1993), na qual pares de nações são anualmente observados por largos períodos de tempo, possibilitando testar os efeitos de regime político, interdependência econômica e nível de poder militar relativo na propensão de conflito entre os países (Russett e Oneal, 2001). De maneira geral, a utilização de diferentes variedades de séries temporais nos estudos comparados em CP e RI tem contribuído ao desenvolvimento empírico e teórico, configurando-se como uma técnica muito discutida (Beck e Katz, 1995) e aplicada (Whitten e Williams, 2011) nessas áreas.

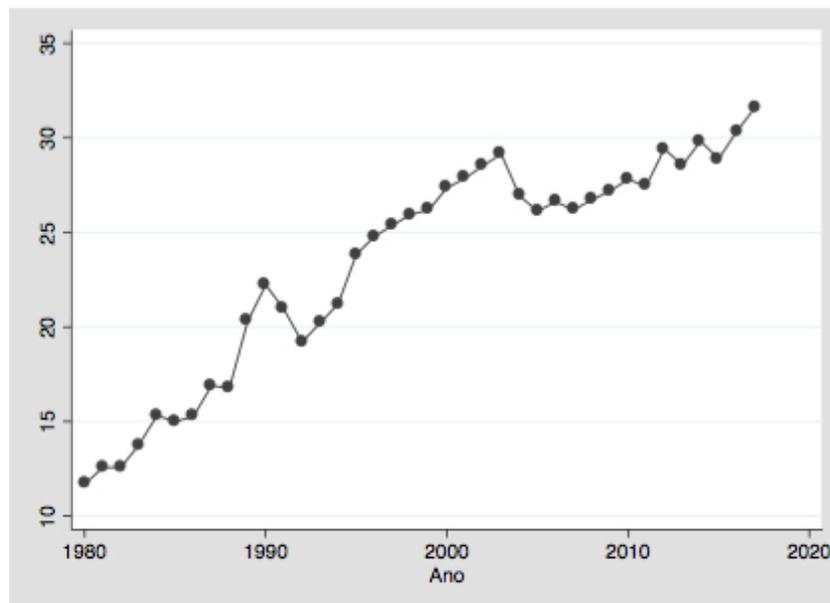
O objetivo deste capítulo é apresentar procedimentos metodológicos de séries de tempo capazes de mensurar as relações entre duas ou mais variáveis identificadas por modelos

teóricos da ciência política e das relações internacionais. Esses procedimentos metodológicos permitem mensurar as relações contemporâneas e passadas entre as variáveis propostas pelo modelo teórico escolhido e testar os impactos de alterações em uma das variáveis nas demais. Por esse motivo, permitem avaliar, por exemplo, os impactos de políticas públicas em indicadores sociais. Uma característica muito importante das séries temporais é a viabilização de uma análise comparada tendo como objeto de pesquisa apenas um país, subdividido em várias unidades no tempo. Uma vantagem dos estudos comparados de um único país, quando confrontados aos estudos de muitos países, é o menor nível de abstração requerido, permitindo uma análise mais intensiva sobre o objeto de pesquisa e a inclusão de elementos específicos do país (Landman, 2008). Um exemplo da abordagem comparada de um único país é o trabalho de Lipset e Marks (2000) sobre a ausência contra intuitiva de um partido socialista nos Estados Unidos. Os autores utilizam séries temporais de imigração e proporção da força de trabalho sindicalizada no país, por exemplo, para compreender o excepcionalismo norte-americano quanto à ausência de um partido socialista relevante.

Para ilustrar os principais passos na estimação de modelo com séries temporais vamos utilizar dados de homicídios por 100 mil habitantes do Brasil entre 1980 e 2017. A ideia central é averiguar a associação entre a série temporal de violência no Brasil e séries temporais socioeconômicas, especificamente a desigualdade de renda. Quatro motivos centrais nos motivaram a escolher a relação temporal entre homicídios e dados socioeconômicos como ilustração da análise de séries temporais. O primeiro diz respeito à relevância social do tema no Brasil. O país apresenta a décima terceira maior taxa de homicídio no mundo (Igarapé, 2018) e em 2017, pela primeira vez, o principal problema do país apontado pela opinião pública foi a violência (Latinobarômetro, 2018). Segundo, o tema perpassa distintas áreas do conhecimento nas ciências sociais, com destaque para sociologia, economia, direito, filosofia, psicologia e ciência política, constituindo-se em um programa de pesquisa interdisciplinar. Terceiro, há um crescente interesse na área de ciência política do Brasil em compreender a eficácia das políticas públicas estaduais de combate ao crime violento (Ferreira, 2011; Portella e Nascimento, 2014; Zaverucha e Nóbrega Júnior, 2015; Freire, 2018). Por fim, o estudo proposto ilustra um modelo comparado para apenas um país, cujos dados em séries temporais costumam ser modelados estatisticamente e podem ser aproveitados para incluir uma variedade de variáveis explicativas (Landman, 2008: 92).

Denominaremos uma série de tempo como uma sequência temporal $\{y_t\}$ que pode tomar valores no conjunto dos números inteiros ou no conjunto dos números reais. Apesar de muitas variáveis socioeconômicas assumirem valores em um intervalo contínuo, observamos os seus valores somente na frequência de divulgação dos dados pelo governo ou pelas instituições responsáveis por suas divulgações. O Instituto de Pesquisa Econômica Aplicada (IPEA) divulga o Atlas da Violência na frequência mensal ou anual (Cerqueira, 2019). A taxa de homicídios totais (por 100 mil habitantes) está disponível na frequência anual para o território nacional, de 1980 a 2017. Dessa forma, a série temporal ‘homicídios’ $\{y_t\}$ é equivalente às realizações $y_{1980}, y_{1981}, y_{1982}, \dots, y_{2017}$. O gráfico 1.1 abaixo mostra a oscilação anual ao longo de uma tendência de crescimento de longo prazo dos homicídios no Brasil:

Gráfico 1.1: Crescimento persistente de homicídios



Fonte: Elaborado pelos autores com dados de IPEA (2019)

A presença de tendência de longo prazo implica não convergência da série temporal para um valor finito, ou seja, um crescimento ou decréscimo infinito quando o horizonte temporal aumenta indefinidamente, o que seria equivalente a um comportamento explosivo da série. Nesse caso, o modelo clássico de regressão pode gerar resultados enganosos. Suponha, por exemplo, que tenhamos duas variáveis quaisquer, y_{1t} e y_{2t} , com tendência crescente ao longo do tempo, como o consumo de pães e o consumo de jornais. Possivelmente, se fizermos uma regressão simples pelo

método dos mínimos quadrados ordinários de consumo de jornais como uma função do consumo de pães, encontraremos que o aumento do consumos de pães tem um impacto positivo e estatisticamente significativo em consumo de jornais. Isso é o que chamamos de regressão espúria, uma regressão com parâmetros e R-quadrado estatisticamente significantes, mas cujos resultados não encontram respaldo em modelo teórico algum. Apesar de não existir uma teoria predizendo uma relação entre consumo de pães e jornais, as variáveis tendem a crescer juntamente com uma terceira variável, o crescimento econômico.

Muitos modelos econométricos, como os univariados auto-regressivos de médias móveis (*autoregressive moving average - ARMA*), os multivariados auto-regressivos de defasagens distribuídas (*autoregressive distributed lag - ADL*) e os modelos de vetor auto-regressivos (*vector autoregressions - VAR*), foram desenvolvidos para o entendimento das oscilações de curto prazo das séries temporais ao longo de suas tendências de longo prazo. Nesse caso, primeiramente testa-se pela presença de tendência em cada uma das variáveis apontadas pelo modelo teórico. Se houver evidência de tendência e não houver uma combinação das variáveis do modelo que seja estacionária, retira-se as tendências de cada variável, e estima-se o modelo com os componentes estacionários de cada variável. Em contrapartida, se as variáveis apresentarem um comportamento explosivo quando analisadas individualmente, mas houver uma relação convergente entre elas, estima-se um modelo de vetor de cointegração (VEC). Nesse capítulo de introdução a séries temporais, abordaremos apenas os modelos ARMA e ADL.

Portanto, faz-se necessário entender se as séries temporais apresentam tendências de longo prazo antes de se estudar e escolher os modelos de séries temporais multivariados apropriados para cada caso. Na seção 1.2, apresentaremos a definição de processos covariância-estacionários, os tipos de tendência exibidas pelas séries temporais, bem como os testes possíveis para a detecção dessas tendências. Na seção 1.3, descreveremos os testes de raiz unitária. Na seção 1.4, apresentaremos os modelos univariados ARMA, bem como os métodos para a estimação dos parâmetros do modelo. Na seção 1.5, exibiremos os modelos multivariados ADL, e finalizaremos o capítulo na seção 1.6, com a ilustração do procedimento para a estimação de um modelo ADL para as relações entre a desigualdade socioeconômica e a taxa de homicídios no Brasil.

1.2 Tendências presentes em séries temporais

1.2.1 Processos covariância-estacionários

Embora o desenho de pesquisa experimental esteja presente na ciência política e nas relações internacionais (McDermott, 2002), é muito comum a disponibilidade apenas de uma única série temporal de dados históricos ano a ano para cada variável, não sendo possível realizar qualquer tipo de desenho experimental. Por exemplo, se desejamos estudar as causas dos homicídios no Brasil, obtemos uma única série histórica de homicídios em território nacional ano a ano. A partir dessa série, desejamos obter conclusões sobre o padrão de homicídios. Intuitivamente, seria como se estivéssemos coletando inúmeras amostras, uma para cada período temporal. Já que temos uma única série temporal disponível, seria ideal que o comportamento de estatísticas amostrais de relevância para o estudo convergisse para o valor verdadeiro das estatísticas populacionais, dado o número de períodos amostrais disponíveis. Para que pudéssemos inferir aspectos sobre o comportamento populacional, gostaríamos que o comportamento dessas amostras tivesse um mesmo padrão ao longo do tempo, como ocorre com os processos chamados covariância-estacionários.

Um processo estocástico é dito fracamente estacionário ou covariância-estacionário se a sua média, variância e auto-covariância são finitas e independentes do instante 't' em que são mensuradas. Denotando-se a média, a variância e a covariância por $E()$, $Var()$ e $Cov()$, respectivamente, uma série temporal é covariância-estacionária se, e somente se, as três condições a seguir são satisfeitas:

1. $E(y_t) = \mu$ para qualquer instante 't'.
2. $Var(y_t) = \sigma^2$ para qualquer instante 't'.
3. $E(y_t y_{t-j}) = \gamma_j$ para qualquer instante 't' e 'j', sendo 'j' um número qualquer do conjunto dos números naturais.

A primeira e a segunda condições dizem que a média e a variância do processo são constantes e finitas, não dependendo do momento em que são analisadas. A terceira condição diz que a auto-covariância ou covariância entre um valor e o seu valor em um momento anterior depende apenas da distância temporal 'j' entre elas, e não do momento 't' de análise. Por exemplo, se a série de homicídios no Brasil for covariância-estacionária, então podemos inferir a média de homicídios no Brasil a partir do valor médio amostral, calculado da seguinte forma:

$$\bar{y} = \frac{\sum_{t=0}^n y_t}{n}$$

em que 'n' é o número de períodos disponíveis.

Cabe-se destacar um caso especial de processo covariância-estacionário, chamado de processo ruído branco (*white-noise*). Um processo *white-noise* é aquele que possui média equivalente a zero, variância constante e ausência de autocorrelação entre um momento 't' e um momento anterior qualquer. Dessa forma, uma série temporal é *white-noise* se, e somente se, as três condições a seguir são satisfeitas:

1. $E(e_t) = 0$ para qualquer instante 't'.
2. $Var(e_t) = \sigma^2$ para qualquer instante 't'.
3. $E(e_t e_{t-j}) = 0$ para qualquer instante 't' e 'j', sendo 'j' um número qualquer do conjunto dos números naturais.

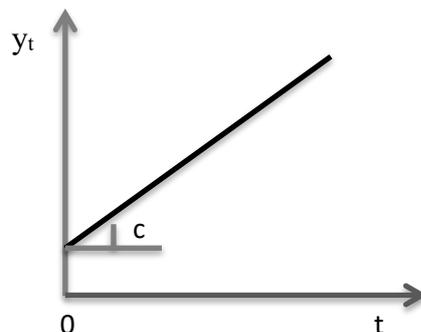
Veremos que, de acordo com os critérios de Box e Jenkins (1976), uma série temporal está adequadamente modelada somente se o resíduo do modelo se comportar como um processo *white-noise*.

Algumas séries de tempo apresentam tendências determinísticas e/ou estocásticas, as quais as tornam não-estacionárias, sendo o tópico apresentado nas subseções 1.2.2 e 1.2.3.

1.2.2 Tendências determinísticas

Suponha que uma série temporal $\{y_t\}$ sempre aumente no valor de uma constante 'c', de um período para o outro, podendo ser representada através de uma linha reta, com inclinação 'c' em relação ao eixo temporal 't':

Gráfico 1.2: Tendência determinística



Fonte: Elaborado pelos autores

Suponha que no primeiro momento quando $t=0$, o valor da série seja y_0 . No momento $t=1$, seu valor y_1 será dado por y_0+c , e assim sucessivamente:

$$\begin{aligned}y_0 &= y_0 \\y_1 &= y_0 + c \\y_2 &= y_0 + 2c\end{aligned}$$

Dessa forma, pode-se representar essa série da seguinte forma:

$$y_t = y_0 + c(t)$$

Além de uma linha reta, uma tendência determinística também pode ter o formato de uma parábola ou de uma curva qualquer, cujos valores podem ser descritos precisamente ou aproximados por um polinômio através, por exemplo, de uma expansão em série de Taylor da função determinística desconhecida:

$$y_t = y_0 + c_1(t) + \dots + c_n(t^n)$$

A série $\{y_t\}$ apresentada consiste, exclusivamente, em uma tendência determinística. A maioria das séries em ciência política e relações internacionais não é exclusivamente determinística, visto que não podem ser previstas com exatidão. Uma série temporal é geralmente equivalente a realizações de uma variável aleatória. Uma variável aleatória ou estocástica é aquela que assume valores aleatórios ou não determinísticos em períodos futuros. Dessa forma, não existe um valor único 'b' tal que a probabilidade da série tomar esse valor no próximo período é exatamente 1 ou 100%. Nesse caso, a variável seria determinística, pois saberíamos ao certo o seu valor no próximo período. Diferentemente, uma variável aleatória é aquela para a qual existe pelo menos um valor 'b' tal que a probabilidade da série assumir esse valor é maior do que zero e menor do que 1. Ou seja, existe pelo menos um valor tal que não sabemos se a variável assumirá ou não esse valor.

Na série temporal de homicídios apresentada, os valores y_{1980} , y_{1981} , y_{1982} y_{2017} são tidos, um a um, como realizações de uma variável aleatória. Isso porque sabemos os seus valores *ex-post* ou depois de terem ocorrido, mas não sabíamos ao certo esses valores *ex-ante* ou antes de terem ocorrido. Dessa forma, uma série temporal, quando apresenta uma tendência determinística, geralmente apresenta conjuntamente um componente aleatório e covariância-estacionário u_t , podendo ser escrita como:

$$y_t = y_0 + c_1(t) + \dots + c_n(t^n) + u_t$$

Muitos modelos econométricos de séries de tempo necessitam de uma série convergente para que possam ser estimados. Nesses casos, anteriormente à estimação do modelo, deve-se testar a presença conjunta de tendência determinística e estocástica, como apresentado na subseção 1.2.3, sobre testes de tendência estocástica ou de raiz unitária. Se a evidência empírica não rejeitar a hipótese de tendência estocástica, deve-se eliminar a tendência estocástica, trabalhando-se com a primeira diferença da série ($y_t - y_{t-1}$), em vez de y_t no procedimento descrito nesta subseção. Se a evidência empírica rejeitar a hipótese de tendência estocástica mas apontar para a presença de tendência determinística, pode-se retirar essa tendência da série, adotando-se o seguinte procedimento:

1. Estima-se os parâmetros c_1, c_2, \dots, c_n do polinômio, considerando-se o grau máximo 'n' plausível. Na maioria dos casos, o polinômio tem grau máximo 3. Para essa estimação, faz-se a regressão da série y_t contra a série de valores temporais $t, t-1, t-2, t-3, \dots$, no nível, ao quadrado, ao cubo e até o expoente máximo do polinômio considerado.
2. Aplica-se testes estatísticos t padrões para determinar o grau do polinômio, com hipótese nula de que o valor do parâmetro é equivalente a zero. Primeiramente, aplica-se um teste t no último parâmetro estimado c_n . Se esse teste implicar na rejeição da hipótese nula, há evidências de que o parâmetro é estatisticamente significativo e identificamos que a tendência determinística tem grau 'n'. Caso contrário, se a hipótese nula não for rejeitada, aplica-se um teste t no parâmetro anterior c_{n-1} . Se houver evidência de que o parâmetro é estatisticamente significativo, então a tendência determinística parece ter grau 'n-1'. Caso contrário, aplica-se um teste t no parâmetro c_{n-2} . Finaliza-se esse processo, quando é testada a significância de todos os parâmetros do polinômio. Concluimos que a tendência determinística parece ser dada por um polinômio de grau equivalente ao maior expoente da série temporal associado a um parâmetro estatisticamente significativo.

Por exemplo, se iniciamos com um polinômio de grau 2, testamos a significância de c_2 . Caso seja estatisticamente significativo, concluimos que a tendência determinística parece ser dada por um polinômio de grau 2. Caso contrário, testa-se a significância de c_1 . Caso seja estatisticamente significativo,

concluimos que a tendência determinística parece ser dada por um polinômio de grau 1, caso contrário, não há evidência de que a série temporal apresenta tendência determinística.

3. O componente aleatório da série é equivalente ao valor da série subtraído do valor de sua tendência determinística. A estimativa do componente aleatório da série é dada pela diferença entre o valor da série temporal y_t e a estimativa da tendência determinística:

$$\hat{u}_t = y_t - (\hat{y}_0 + \hat{c}_1(t) + \dots + \hat{c}_n(t^n))$$

4. Muitos modelos econométricos são utilizados para modelar essa estimativa do componente aleatório como os modelos univariados auto-regressivos de médias móveis (ARMA) e os modelos multivariados auto-regressivos de defasagens distribuídas (ADLs).

Por exemplo, se testarmos pela existência de tendência determinística para a série do índice de desigualdade brasileiro GINI, de 1980 a 2018, obtemos que apenas os termos t e t^2 são importantes para explicar a tendência determinística:

$$GINI_t = c0 + 38.28(t) - 0.01(t^2) + u_t$$

(3.81) (-3.83)

Com as estatísticas t entre parênteses, indicamos a significância dos parâmetros ao nível de 1%. Porém, esses resultados são enganosos, visto que não podemos rejeitar a hipótese nula de presença de tendência estocástica em $GINI_t$ quando realizamos um teste conjunto para presença de tendência determinística e estocástica. Como veremos na próxima subseção, nesse caso, devemos tomar a primeira diferença da série ($GINI_t - GINI_{t-1}$) para torná-la estacionária. Podemos testar também se a série possui uma tendência determinística, além da tendência estocástica, estimando-se a regressão da sua primeira diferença contra os termos de tendência. Os resultados são:

$$GINI_t - GINI_{t-1} = c0 - 7.47(t) + 0.002(t^2) + u_t$$

(-0.07) (0.07) ,

e obtemos que os parâmetros são todos estatisticamente insignificantes.

As séries com tendência determinística são também chamadas de *trend stationary* (tendência estacionária), já que se tornam estacionárias ao se eliminar a sua tendência determinística, subtraindo-a da série original. Além de tendências determinísticas, as

séries temporais podem apresentar tendências estocásticas, das quais trataremos agora.

1.2.3 Tendências estocásticas

Suponha que o valor de uma série temporal $\{y_t\}$ seja sempre equivalente ao seu valor no período anterior $\{y_{t-1}\}$ adicionado a um termo aleatório *white-noise* e_t :

$$y_t = y_{t-1} + e_t$$

Esse processo é uma tendência estocástica, sendo denominado **passeio aleatório** ou *random walk*. Muitas séries temporais apresentam esse comportamento. Por exemplo, de acordo com a teoria da aleatoriedade dos retornos das ações popularizada por Cootner (1964) e Fama (1965), os preços das ações p_t seguiriam uma *random walk*:

$$p_t = p_{t-1} + e_t,$$

em que e_t é um processo *white-noise*.

Portanto, as variações dos preços das ações $(p_t - p_{t-1})$ seriam aleatórias ou imprevisíveis e dadas pelo componente *white-noise* e_t :

$$p_t - p_{t-1} = e_t$$

Se essa hipótese não fosse refutada empiricamente, seria impossível obter retornos positivos apostando em ações. Entretanto, diversos autores mostraram evidências contrárias a essa hipótese. Lo e Mackinlay (2002) refutam a hipótese de *random walk* através de testes baseados na volatilidade dos retornos das ações. Jegadeesh e Titman (1993) apresentam evidência de que a compra de ações que deram retornos relativamente altos nos últimos meses e a venda de ações que tiveram retornos ruins podem ser rentáveis.

Para entender o comportamento não convergente de uma *random walk*, construiremos uma série y_t , tal que no primeiro momento $t=0$ o valor da série seja y_0 , no momento $t=1$, seu valor y_1 seja dado por y_0+e_1 , e assim sucessivamente:

$$\begin{aligned} y_0 &= y_0 \\ y_1 &= y_0 + e_1 \\ y_2 &= y_1 + e_2 \end{aligned}$$

⋮

$$y_t = y_{t-1} + e_t$$

Podemos reescrever y_2 , substituindo-se pelo valor de y_1 , como:

$$y_2 = y_0 + e_1 + e_2$$

Da mesma forma,

$$\begin{aligned} y_3 &= y_2 + e_3 \\ y_3 &= y_0 + e_1 + e_2 + e_3 \end{aligned}$$

Dessa forma, pode-se representar essa série da seguinte forma:

$$y_t = y_0 + \sum_{n=1}^t (e_n)$$

O que caracteriza um processo *random walk* é a persistência dos choques ao longo do tempo. Observa-se que o efeito de um choque aleatório qualquer e_n não se dissipa com o passar do tempo, pois tem um impacto permanente no valor da série y_t , a partir do momento 'n' em que ocorre. Por isso, esse processo é caracterizado como uma tendência estocástica.

Se tomarmos a expectativa não-condicional da série y_t obtemos:

$$\begin{aligned} E(y_t) &= E(y_0) + E \sum_{n=1}^t (e_n) \\ E(y_t) &= y_0 + \sum_{n=1}^t E(e_n) \\ &= y_0 \end{aligned}$$

O valor esperado de y_0 é exatamente y_0 , visto ser uma constante qualquer. O valor esperado da soma dos termos aleatórios é equivalente à soma dos seus valores esperados, propriedade característica do operador esperança. Além disso, o valor esperado de cada um dos termos aleatórios é zero. Dessa forma, o valor esperado da série é equivalente ao valor da condição inicial y_0 para qualquer horizonte temporal t . Porém, a variância da série depende do horizonte temporal, sendo maior quanto maior o valor de t . Dado o pressuposto de processo e_t *white-noise*, que implica em ausência de correlação entre os choques, tem-se que:

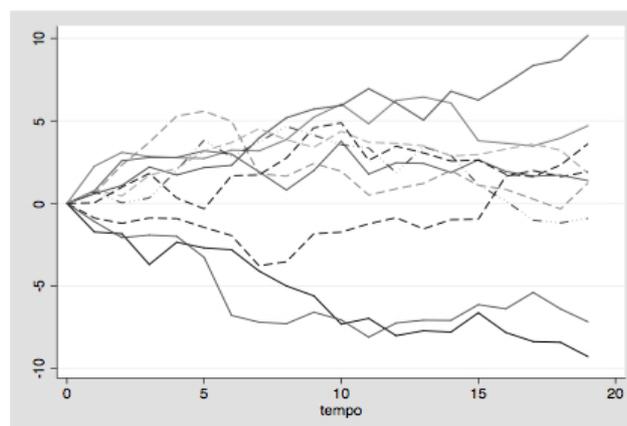
$$\begin{aligned} \text{Var}(y_t) &= \text{Var}(y_0) + \text{Var} \sum_{n=1}^t (e_n) \\ \text{Var}(y_t) &= \sum_{n=1}^t \text{Var}(e_n) \\ &= \sigma^2 + \sigma^2 + \dots + \sigma^2 = t\sigma^2 \end{aligned}$$

Dessa forma, vemos que a variância de um processo *random walk* tende ao infinito quando o tempo tende ao infinito. Isso leva a série a apresentar desvios persistentes em relação a qualquer valor finito ou ausência de convergência à sua média não-condicional nos horizontes temporais observados.

Vimos que um processo estocástico é dito fracamente estacionário ou covariância-estacionário se a sua média, variância e auto-covariância são finitas e independentes do instante 't' em que são mensuradas. Dessa forma, tem-se que *random walk* não é um processo estacionário.

O gráfico 1.3 mostra a simulação de 10 séries *random walk* com parâmetro y_0 equivalente a 0 e $e_t \sim N(0, 1)$. As séries foram simuladas no programa STATA. Cada uma das 10 linhas corresponde a um processo *random walk* diferente, sujeito a choques aleatórios distintos e independentes $e_t \sim N(0, 1)$. Vejam que as séries com linhas sólidas foram sujeitas aleatoriamente a choques de maior magnitude e não apresentam tendência de convergência a qualquer valor finito no período analisado. Já quanto às séries em tracejado, apesar de aparentarem um comportamento mais convergente ao valor da média não-condicional, isso ocorre por sorte de não terem sido afetadas por choques elevados no curto período analisado. Deve-se, portanto, ter cautela e não se tirar conclusões sobre a convergência de uma série através de uma análise visual de seu comportamento em um curto período de tempo.

Gráfico 1.3: Processos *random walk*



Fonte: Elaborado pelos autores.

Para entender a diferença entre o processo gerador de uma série convergente e uma *random walk*, considere uma série temporal qualquer $\{y_t\}$ e componente aleatório *white-noise* e_t , que seja função apenas de seu valor no período anterior y_{t-1} :

$$y_t = \rho y_{t-1} + e_t,$$

em que $|\rho| < 1$.

No primeiro momento $t=0$, o valor da série é dado por y_0 . No momento $t=1$, seu valor y_1 será dado por $\rho y_0 + e_1$, e assim sucessivamente:

$$\begin{aligned} y_0 &= y_0 \\ y_1 &= \rho y_0 + e_1 \\ y_2 &= \rho y_1 + e_2 \\ &\vdots \end{aligned}$$

$$y_t = \rho y_{t-1} + e_t$$

Podemos reescrever y_2 , substituindo-se pelo valor de y_1 , como:

$$y_2 = \rho^2 y_0 + \rho e_1 + e_2$$

Da mesma forma,

$$\begin{aligned} y_3 &= \rho y_2 + e_3 \\ y_3 &= \rho^3 y_0 + \rho^2 e_1 + \rho e_2 + e_3 \end{aligned}$$

Portanto, pode-se representar essa série da seguinte forma:

$$y_t = \rho^t y_0 + \sum_{n=0}^{t-1} \rho^n (e_{t-n})$$

A convergência de uma série autoregressiva de ordem um ou que seja função apenas de seu valor no período anterior y_{t-1} depende do valor do parâmetro ρ . Tem-se três casos possíveis:

1. Se $|\rho| < 1$, temos que o termo ρ^t será menor quanto maior for o valor de 't'. Assim, o impacto da condição inicial y_0 no valor da série y_t vai se reduzindo com o passar do tempo e tende a zero quando o tempo tende ao infinito. Da mesma forma, o impacto de

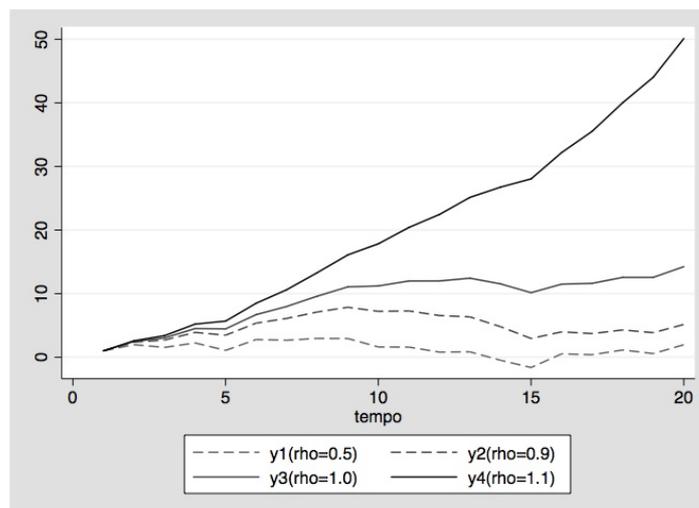
choques passados na série tendem a se dissipar com o passar do tempo. Esse seria o caso de processos covariância-estacionários.

2. Se $|\rho| = 1$, tem-se uma *random walk* em que o impacto da condição inicial em y_t não se reduz ao longo do tempo. Da mesma forma, os choques estocásticos passados têm um efeito permanente na série, o que leva a não convergência da série.

3. Se $|\rho| > 1$, tem-se uma série explosiva em que o termo ρ^t será maior quanto maior for o valor de 't'. Nesse caso, o impacto da condição inicial y_0 em y_t aumenta ao longo do tempo. Da mesma forma, os choques estocásticos passados têm um efeito crescente na série, o que leva a um comportamento não-convergente e explosivo da série.

O Gráfico 1.4 apresenta a simulação em que são plotadas as séries $y_t = \rho y_{t-1} + e_t$, para quatro valores positivos de ρ : 0,5; 0,9; 1; 1,1. A condição inicial y_0 foi tomada como sendo equivalente a 0 e foram gerados 20 choques aleatórios $e_t \sim N(0, 1)$ para a construção da série $y_t = \rho y_{t-1} + e_t$; $t=1,2,\dots,20$. Os choques aleatórios são os mesmos para as quatro séries. Vê-se que as séries y_1 e y_2 , em tracejado, com ρ equivalente a 0,5 e 0,9, respectivamente, possuem comportamento convergente ao redor de zero, com taxa ou rapidez de convergência maior para a série y_1 , com menor valor de ρ . As séries y_3 e y_4 , com ρ equivalente a 1 e 1,1, não apresentam comportamento convergente. A série y_3 é uma *random walk*, apresentando choques persistentes, com impacto que não se reduz ao longo do tempo, e a série y_4 , de caráter explosivo, apresenta choques com impacto crescente ao decorrer do tempo.

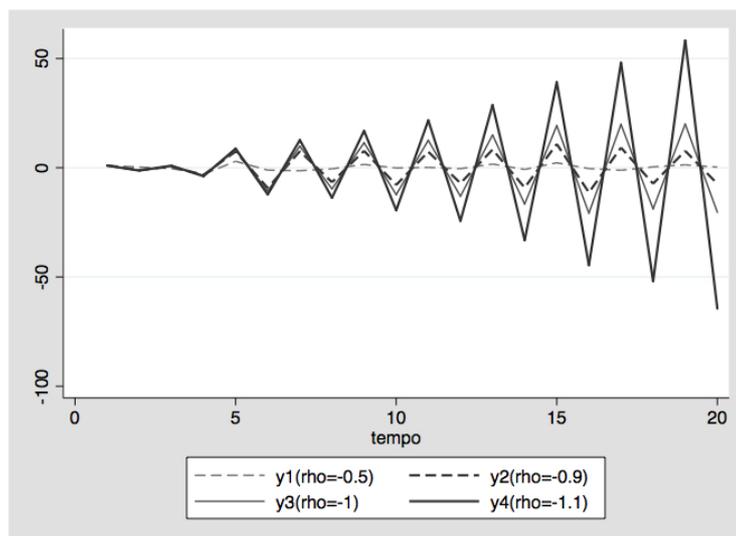
Gráfico 1.4: Convergência de séries temporais com $\rho > 0$



Fonte: Elaborado pelos autores.

No gráfico 1.5 realizamos o mesmo exercício apresentado no gráfico 1.4, porém com valores negativos de ρ ($-0,5; -0,9; -1; -1,1$), o que implica em caráter oscilatório da série. A condição inicial y_0 é igual a 0 e foram gerados 20 choques aleatórios $e_t \sim N(0,1)$ para a construção da série $y_t = \rho y_{t-1} + e_t$ para $t=1,2,\dots,20$. Os choques aleatórios são os mesmos para as quatro séries apresentadas. Nesse caso as séries em tracejado, com $|\rho| < 1$, apresentam comportamento convergente. Já as séries com ρ equivalente a -1 e $-1,1$, apresentam comportamento divergente. Pode-se notar que essas séries apresentam variância que cresce ao longo do tempo, com picos de magnitude cada vez mais elevada, diferentemente das séries convergentes, com variância limitada ao longo do tempo.

Gráfico 1.5: Convergência de séries temporais com $\rho < 0$



Fonte: Elaborado pelos autores.

Os procedimentos para eliminação da tendência e estimação de um modelo para o componente convergente da série dependem do tipo de tendência que a série possui. No caso de uma tendência determinística, vimos que devemos estimar os parâmetros da tendência e subtrair a tendência da série. Já no caso de uma tendência estocástica dada por um *random walk*, tem-se que a primeira diferença da série é equivalente ao processo *white-noise* e_t e, portanto, covariância-estacionária:

$$y_t - y_{t-1} = e_t$$

Dessa forma, a média, a variância e a autocovariância da primeira diferença da série

respeitam a definição de processo covariância-estacionário, sendo dados por:

$$E(y_t - y_{t-1}) = E(e_t) = 0$$

$$Var(y_t - y_{t-1}) = Var(e_t) = \sigma^2$$

$$E(e_t, e_{t=j}) = 0$$

para qualquer instante t e número inteiro j .

Elimina-se, pois, a tendência de uma *random walk*, tomando-se a sua primeira diferença. Por isso, a série é denominada *difference stationary* (diferença-estacionária). Em contraposição ao processo de eliminação de tendência de uma série *trend stationary* (tendência estacionária), que se torna estacionária se subtrairmos a sua tendência determinística.

1.3 Teste de raiz unitária

Um teste de raiz unitária em $\{y_t\}$ seria equivalente ao teste da hipótese nula de que $\rho = 1$ em:

$$y_t = \rho y_{t-1} + X_t + e_t,$$

em que X_t são todos os demais fatores que explicam variações de y_t .

Porém, sob a hipótese nula, o modelo é *random walk*, não sendo convergente. Nesse caso, o método dos mínimos quadrados ordinários levaria a uma estimação enviesada de ρ . Dessa forma, não podemos testar a hipótese nula através de um teste t padrão.

Dickey e Fuller (1979) desenvolveram testes de raiz unitária que contornam o problema de não-convergência da série sob a hipótese nula, sobre os quais discorreremos agora. Primeiramente, considere um modelo geral para $\{y_t\}$, que permite à série apresentar tendência determinística 't' e estocástica, além do termo de *drift* c_0 :

$$y_t = c_0 + c_1(t) + \rho y_{t-1} + e_t$$

Subtraindo-se y_{t-1} de ambos os lados, podemos reescrever da seguinte forma:

$$y_t - y_{t-1} = c_0 + c_1(t) + (\rho - 1)y_{t-1} + e_t$$

$$\text{ou } \Delta y_t = c_0 + c_1(t) + \alpha y_{t-1} + e_t$$

Como vimos, um teste de raiz unitária equivale ao teste com hipótese nula $\rho = 1$, o qual seria equivalente a um teste com hipótese nula $\alpha = 0$, visto que $\alpha = \rho - 1$. Essa seria a hipótese nula do teste de Dickey e Fuller (1979). Nota-se que, nesse caso, a variável dependente do modelo Δy_t é covariância-estacionária sob a hipótese nula do teste, e pode-se realizar a estimação do modelo por mínimos quadrados ordinários. Após a estimação, obtém-se o valor da estatística t do teste, dada pela razão entre α e o desvio-padrão de α . O valor da estatística t do teste deve ser comparado aos valores críticos do teste, encontrados por meio de estudos de Monte Carlo.

O teste de raiz unitária é um teste unicaudal, visto que a hipótese alternativa de estacionaridade implica em $\rho < 1$ ou $\alpha < 0$. Dessa forma, somente valores da estatística t menores do que um valor crítico negativo levam à rejeição da hipótese nula de *random walk* e provêm evidência a favor de um processo convergente.

Os estudos de Monte Carlo mostram que os valores críticos dependem do número de observações utilizadas e da presença ou não do termo *drift* e da tendência '*t*'. O valor crítico seria menor, ou maior em magnitude (dado ser negativo), na presença do termo *drift* c_0 e ainda menor na presença do termo de tendência '*t*'. Dessa forma, a inclusão desnecessária desses termos reduz o poder estatístico do teste, de rejeição da hipótese nula se realmente for falsa. Ao mesmo tempo, a não inclusão em um modelo que os possui pode nos levar a estimativas inconsistentes de todos os parâmetros do modelo e a testes de conclusões equivocadas. Veremos que o procedimento padrão para os testes de raiz unitária envolve o teste da hipótese conjunta de presença dos termos de *drift* e de tendência.

Os parâmetros estimados desse modelo podem ser inconsistentes e não adequados para o teste de raiz unitária se outras defasagens da variável dependente, além daquela do período anterior y_{t-1} , também impactarem as variações de y_t . De fato, muitas séries de tempo dependem não somente da defasagem do período anterior, como também de defasagens adicionais, além do termo de tendência '*t*' e de outros fatores não controlados e_t :

$$y_t = f(t, y_{t-1}, y_{t-2}, \dots, y_{t-n}, e_t),$$

em que f seria uma função qualquer.

Esse modelo geralmente pode ser aproximado por uma função linear, podendo assumir a seguinte representação:

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_n y_{t-n} + \lambda t + e_t,$$

em que $b_0, b_1, b_2, \dots, b_n, \lambda$ seriam parâmetros a serem estimados.

Dickey e Fuller (1981) desenvolveram *augmented Dickey-Fuller tests*, os chamados testes de raiz unitária expandidos. Eles mostram que esse modelo pode ser escrito com as séries de tempo na primeira diferença e estacionárias sob a hipótese nula de presença de raiz unitária, com exceção de y_{t-1} , cujo parâmetro α será testado:

$$\Delta y_t = c_0 + c_1(t) + \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_n \Delta y_{t-n} + e_t$$

Novamente, após o controle de todas as demais defasagens da série, o teste de raiz unitária equivale ao teste de $\alpha=0$. Os valores críticos para o teste seriam os mesmos dos testes não aumentados.

O procedimento para testes de raiz unitária de Dickey e Fuller (1981) inicia-se com a escolha do número de defasagens e posterior teste de presença conjunta de tendência determinística e estocástica, finalizando-se com o teste de presença apenas de tendência estocástica. Considere o modelo:

$$\Delta y_t = c_0 + c_1(t) + \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_n \Delta y_{t-n} + e_t$$

1. Primeiramente, estima-se o modelo por mínimos quadrados ordinários, e escolhe-se o número de defasagens n através de testes t ou F convencionais, com hipótese nula de que o valor do parâmetro é nulo e valores críticos apropriados das tabelas da distribuição *t-student*, Normal ou F . Começa-se testando a defasagem mais longa plausível 'n'. Se o valor da estatística t , em módulo, for menor do que o valor crítico, em módulo, retira-se a defasagem do modelo e testa-se a próxima defasagem. Se for maior, mantém-se a defasagem, e testa-se a próxima defasagem. O procedimento continua até que o valor do parâmetro β_1 tenha sido testado.
2. Testa-se a hipótese nula $H_0: c_0 = c_1 = \alpha = 0$, comparando-se com os valores críticos apropriados de Dickey e Fuller (1979). Para uma amostra de tamanho 100, tem-se 4.88 para nível de significância equivalente a 5% e 6.50 para nível de significância equivalente a 1%. Se a hipótese nula for rejeitada, é evidência de que pelo menos um desses parâmetros é diferente de zero. Porém, ainda não se pode rejeitar a hipótese de raiz unitária, e segue-se ao passo 3. Se a hipótese nula não for rejeitada, não se pode concluir que há evidência a favor de raiz

unitária, visto que o termo de tendência reduz o poder do teste quando presente desnecessariamente na regressão. Retira-se o termo de tendência do modelo, e segue-se ao passo 5.

3. Testa-se a hipótese nula $H_0: c_1 = \alpha = 0$, e compara-se com os valores críticos apropriados de Dickey e Fuller (1979). Para uma amostra de tamanho 100, tem-se 6.49 para nível de significância equivalente a 5% e 8.73 para nível de significância equivalente a 1%. Se a hipótese nula for rejeitada, é evidência a favor de tendência determinística e contrária à presença de raiz unitária. Porém, ainda não se pode rejeitar a hipótese de raiz unitária, e segue-se ao passo 4. Se a hipótese nula não for rejeitada, retira-se o termo de tendência, e segue-se ao passo 5.
4. Testa-se a hipótese nula $H_0: \alpha = 0$, e compara-se com os valores críticos apropriados de Dickey e Fuller (1979). Para uma amostra de tamanho 100, tem-se -3.45 para nível de significância equivalente a 5% e -4.04 para nível de significância equivalente a 1%. Se a hipótese nula for rejeitada, é evidência contrária à presença de raiz unitária. Se a hipótese nula não for rejeitada, é evidência a favor da presença de raiz unitária, podendo-se retirar a tendência estocástica da série tomando-se a sua primeira diferença. Conclui-se o teste.
5. Nos passos 2 e 3, vimos que se não houver evidência a favor da tendência 't', deve-se retirar esse termo, e estima-se o seguinte modelo:

$$\Delta y_t = c_0 + \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_n \Delta y_{t-n} + e_t$$

Testa-se a hipótese nula $H_0: c_0 = \alpha = 0$, comparando-se com os valores críticos apropriados de Dickey e Fuller (1979). Para uma amostra de tamanho 100, tem-se 4.71 para nível de significância equivalente a 5% e 6.70 para nível de significância equivalente a 1%. Se a hipótese nula for rejeitada, é evidência de que pelo menos um desses parâmetros é diferente de zero. Porém, ainda não se pode rejeitar a hipótese de raiz unitária e segue-se ao passo 6. Se a hipótese nula não for rejeitada, não se pode concluir que há evidência a favor de raiz unitária, visto que o termo *drift* reduz o poder do teste quando presente desnecessariamente na regressão. Retira-se o termo *drift* do modelo, e segue-se para o passo 7.

6. Testa-se a hipótese nula $H_0: \alpha = 0$, e compara-se com os valores críticos apropriados de Dickey e Fuller (1979). Para uma amostra de tamanho 100, tem-se -2.89 para nível de significância equivalente a 5% e -3.51 para nível de

significância equivalente a 1%. Se a hipótese nula for rejeitada, é evidência contrária à presença de raiz unitária. Se a hipótese nula não for rejeitada, é evidência a favor da presença de raiz unitária, podendo-se retirar a tendência estocástica da série tomando-se a sua primeira diferença. Conclui-se o teste.

7. No passo 5, vimos que se não houver evidência a favor da presença de tendência ‘t’ e *drift*, deve-se retirar esses termos e estimar-se o seguinte modelo:

$$\Delta y_t = \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_n \Delta y_{t-n} + e_t$$

Testa-se a hipótese nula $H_0: \alpha = 0$, comparando os valores críticos apropriados de Dickey e Fuller (1979). Para uma amostra de tamanho 100, tem-se -1.95 para nível de significância equivalente a 5% e -2.60 para nível de significância equivalente a 1%. Se a hipótese nula for rejeitada, é evidência contrária à presença de raiz unitária. Se a hipótese nula não for rejeitada, é evidência a favor da presença de raiz unitária, podendo-se retirar a tendência estocástica da série tomando a sua primeira diferença. Conclui-se o teste.

Por exemplo, se testarmos a hipótese nula de existência de tendência estocástica para a série de taxas de homicídios totais brasileiros, de 1980 a 2017:

$$\Delta hom_t = c_0 + c_1(t) + \alpha hom_{t-1} + \beta_1 \Delta hom_{t-1} + \dots + \beta_8 \Delta hom_{t-8} + e_t$$

Primeiramente, obtemos que nenhuma defasagem ($\Delta y_{t-1}, \dots, \Delta y_{t-8}$) deve ser incluída na regressão, de acordo com testes t em cada um dos coeficientes testados.

Em segundo lugar, rejeita-se: $H_0: c_0 = c_1 = \alpha = 0$, porém não se pode concluir que não há raiz unitária. Segue-se ao passo 3, e testa-se $H_0: c_1 = \alpha = 0$. Nesse caso, não se rejeita a hipótese nula, o que seria evidência a favor de raiz unitária e contraria a presença de tendência determinística. Seguimos ao passo 5 e estimamos:

$$\Delta hom_t = 1.67 - 0.05 hom_{t-1} + e_t$$

(2.14) (-1.49)

Testa-se a hipótese nula $H_0: c_0 = \alpha = 0$, comparando-se com os valores críticos apropriados de Dickey e Fuller (1979). Nesse caso, rejeitamos a hipótese nula sendo evidência de que pelo menos um desses parâmetros é diferente de zero. Porém, ainda não se pode rejeitar a hipótese de raiz unitária, e seguimos para o passo 6. Testamos a hipótese nula $H_0: \alpha = 0$. O valor da estatística t reportado, relacionado ao parâmetro α , é de -1.49, não sendo inferior ao valor crítico apontado por Dickey e Fuller de -2.96,

dependente do tamanho amostral e do *drift* no modelo. Não rejeitamos a hipótese nula, sendo evidência a favor da presença de raiz unitária. Retiramos a tendência estocástica da série tomando-se a sua primeira diferença e concluímos o teste. Para o índice GINI, o mesmo procedimento também leva a não rejeição da hipótese nula de raiz unitária.

1.4 Modelos autoregressivos de médias móveis (*Autoregressive Moving Average (ARMA)*)

Muitos processos temporais univariados $\{y_t\}$ podem ser representados através de uma função linear de seus valores passados $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ e de choques aleatórios exógenos correntes e passados $(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q})$, o que denominamos modelos autoregressivos de médias móveis ou ARMA. Um processo ARMA(p,q) pode ser representado da seguinte forma:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t + \delta_1 \varepsilon_{t-1} + \delta_2 \varepsilon_{t-2} + \dots + \delta_q \varepsilon_{t-q}$$

Nesse modelo, os termos autoregressivos são as defasagens da variável dependente $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$. Esses termos permitem que uma variável y_t dependa de seus valores nos períodos anteriores, os quais são influenciados pelos choques passados $(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$. Os termos de médias móveis são os choques correntes e passados $(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q})$. Quanto menores, em valor absoluto, forem os valores dos parâmetros β e δ , mais temporários são os efeitos dos choques passados nos valores presentes da variável y_t e maior a rapidez de reversão do processo à sua média de longo prazo.

De acordo com as técnicas de Box e Jenkins (1976), os processos ARMA(p,q) são caracterizados e podem ser identificados através de um processo de três etapas:

1. Realização de testes de não-estacionariedade em y_t : se houver evidência de tendência estocástica, trabalhar com a primeira-diferença da série; se houver evidência de tendência determinística, trabalhar com o resíduo da regressão contra o termo de tendência;

2. Visualização dos gráficos das funções de autocorrelação e autocorrelação parcial a fim de considerar os modelos mais plausíveis;
3. Estimação dos modelos mais plausíveis e escolha por meio dos critérios AIC e SBC, que serão explicados a continuação;
4. Exame de diagnóstico dos resíduos. Um modelo adequadamente identificado implica em resíduo seguindo um processo *white-noise*.

Para que possamos estimar os parâmetros de um modelo ARMA(p,q) de forma consistente, devemos ter uma série convergente. Dessa forma, iniciamos a análise testando pela presença de tendências estocástica e/ou determinística e eliminando-as, caso haja evidência de existência. Já estudamos os testes de estacionariedade e partiremos para a explicação da segunda etapa de Box e Jenkins (1976).

Na segunda etapa, devemos plotar as funções de autocorrelação e autocorrelação parcial. A função de autocorrelação é formada pela autocorrelação da série entre o momento corrente e cada uma de suas defasagens anteriores. Considerando-se um processo covariância-estacionário, vimos que a variância e autocovariância são dados, respectivamente, por:

- $Var(y_t) = \sigma^2$ para qualquer instante 't'.
- $E(y_t y_{t-j}) = \gamma_j$ para qualquer instante 't' e 'j', sendo 'j' um número qualquer do conjunto dos números naturais.

Nesse caso, γ_j representa a covariância ou autocovariância entre a série y_t e a sua defasagem 'j', y_{t-j} , para qualquer instante 't' de avaliação.

A autocorrelação de ordem 'j' é definida como a razão da covariância γ_j pela variância σ^2 :

$$\rho_j = \frac{\gamma_j}{\sigma^2}$$

Por sua vez, a **função de autocorrelação** (*Autocorrelation Function (ACF)*) é definida como o valor da autocorrelação como função da defasagem 'j'. A autocovariância entre uma série qualquer y_t e ela própria é sempre equivalente à sua variância. Dessa forma, a autocorrelação de ordem 0 geralmente não é apresentada, pois é equivalente a 1 para qualquer série temporal.

A autocorrelação entre a série y_t e a sua defasagem y_{t-j} depende não somente da correlação direta entre as duas séries, mas da correlação entre y_t e $y_{t-1}, \dots, y_{t-j+1}$, caso $y_{t-1}, \dots, y_{t-j+1}$ estejam correlacionados com y_{t-j} . Por exemplo, em um processo ARMA(p,0) a autocorrelação entre y_t e $y_{t-(p+1)}$ não é nula, mesmo que a defasagem ‘p+1’ não entre diretamente na regressão. Isso ocorre porque, por exemplo, y_t e y_{t-p} são diretamente correlacionados e, ao mesmo tempo, a defasagem y_{t-p} é diretamente correlacionada com as suas ‘p’ defasagens, incluindo a defasagem ‘p+1’.

Surge então a definição de autocorrelação parcial para captar a correlação direta entre série y_t e a sua defasagem y_{t-j} . A **função de autocorrelação parcial** (*Partial Autocorrelation Function (PACF)*) é definida como o valor da autocorrelação parcial como função da defasagem ‘j’. A autocorrelação parcial entre y_t e a sua defasagem y_{t-j} é estimada pela regressão de y_t subtraído de sua média \bar{y} contra todas as suas defasagens, até a defasagem ‘j’. Dessa forma, a autocorrelação entre y_t e a sua defasagem y_{t-1} é dada pelo valor estimado de ρ_{11} na regressão de $y_t - \bar{y}$ contra $y_{t-1} - \bar{y}$:

$$y_t - \bar{y} = \rho_{11}(y_{t-1} - \bar{y}) + u_{1t}$$

A autocorrelação parcial entre y_t e a sua defasagem y_{t-2} é dada pelo valor estimado de ρ_{22} na regressão de $y_t - \bar{y}$ contra $y_{t-1} - \bar{y}, y_{t-2} - \bar{y}$:

$$y_t - \bar{y} = \rho_{21}(y_{t-1} - \bar{y}) + \rho_{22}(y_{t-2} - \bar{y}) + u_{2t}$$

E a autocorrelação parcial entre y_t e a sua defasagem y_{t-j} é dada pelo valor estimado de ρ_{jj} na regressão de $y_t - \bar{y}$ contra $y_{t-1} - \bar{y}, y_{t-2} - \bar{y}, \dots, y_{t-j} - \bar{y}$:

$$y_t - \bar{y} = \rho_{j1}(y_{t-1} - \bar{y}) + \rho_{j2}(y_{t-2} - \bar{y}) + \dots + \rho_{jj}(y_{t-j} - \bar{y}) + u_{jt}$$

A **função de autocorrelação parcial** (*Partial Autocorrelation Function (PACF)*) é definida como o valor da autocorrelação parcial como função da defasagem ‘j’. Os gráficos da função de autocorrelação e da função de autocorrelação parcial já estão programados em softwares econométricos como Eviews, Stata e R.

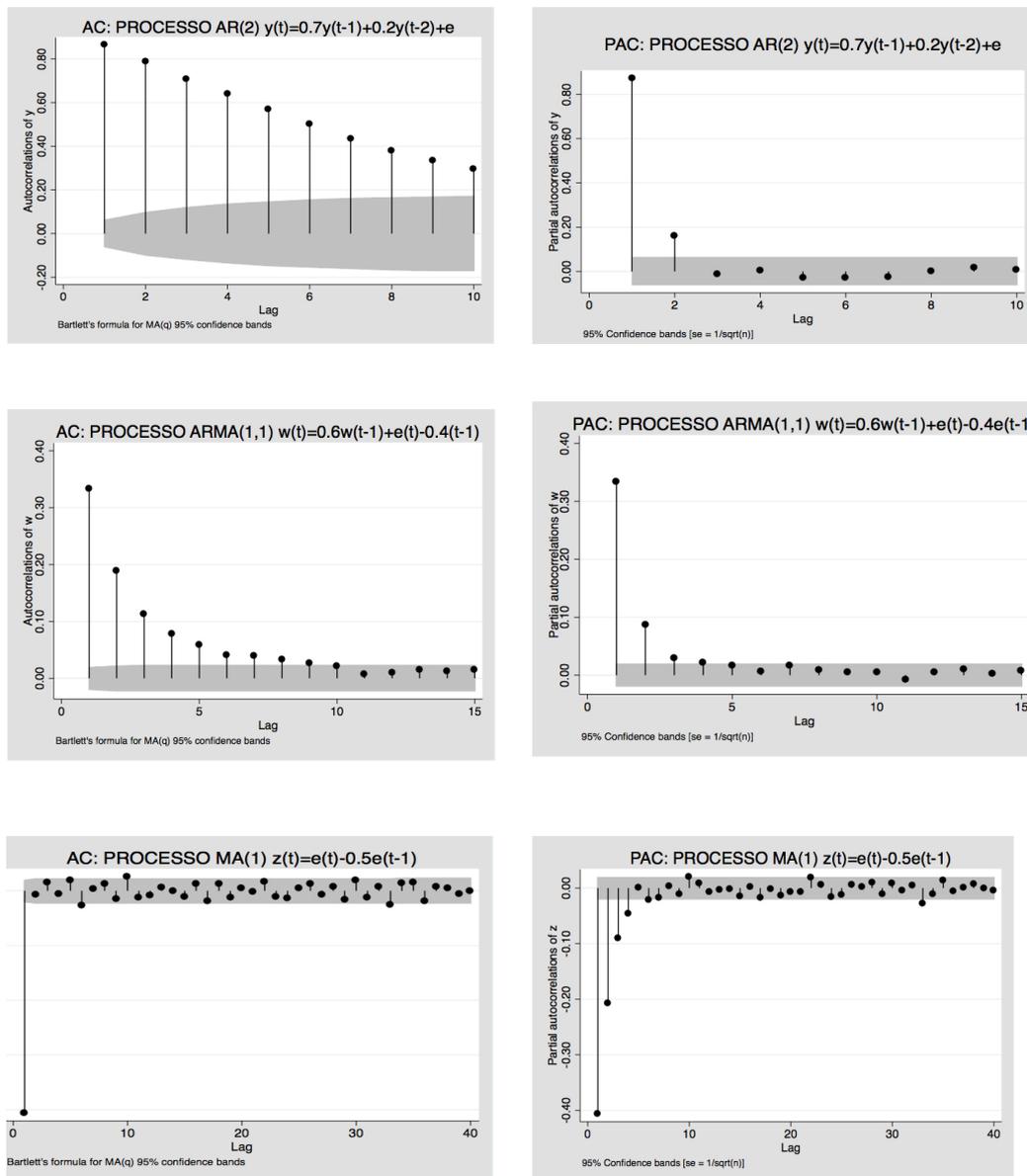
As funções de autocorrelação e de autocorrelação parcial têm características específicas que podem auxiliar na identificação do melhor modelo univariado ARMA(p,q) para uma série de tempo. Pode-se demonstrar (Enders, 2008) que um modelo autoregressivo puramente, caracterizado por q=0, apresenta uma função de autocorrelação que decresce

suavemente, em valor absoluto, com o aumento da defasagem j e, ao mesmo tempo, apresenta uma função de autocorrelação parcial com valores estatisticamente significantes somente nas ordens p do processo. O gráfico 1.6, gerado no Stata, mostra as funções de autocorrelação e autocorrelação parcial, respectivamente, para um processo simulado ARMA(2,0) ou AR(2), gerados pela seguinte equação:

$$y_t = 0.7y_{t-1} + 0.2y_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim N(0,1)$$

Gráfico 1.6: Funções AC e PAC para processos AR(p), ARMA(p,q) e MA(q)



Fonte: Elaborado pelos autores.

Nota-se que, se não soubéssemos que a série de tempo segue um modelo AR(2), poderíamos inferir esse processo gerador através do comportamento do gráfico PACF, com valores estatisticamente diferentes de zero apenas para as defasagens 1 e 2, e do gráfico ACF, decrescente ao longo do tempo.

Em contraste, um modelo puramente de médias móveis é caracterizado por gráfico PACF gradualmente decrescente e gráfico ACF com picos estatisticamente significantes nas ordens q do processo. O gráfico 1.6 mostra as funções de autocorrelação e autocorrelação parcial, respectivamente, para um processo ARMA(0,1) ou MA(1), gerado pela seguinte equação:

$$z_t = \varepsilon_t - 0.5\varepsilon_{t-1}$$
$$\varepsilon_t \sim N(0,1)$$

Da mesma forma, podemos inferir o processo MA(1) por meio da observação do gráfico ACF, com o primeiro pico estatisticamente significante, somente, e do gráfico PACF, de declínio gradual.

As funções de autocorrelação para um processo autoregressivo e de média-móvel conjunto ARMA(p,q) apresentam picos do gráfico ACF significantes nas defasagens ' q ' do processo e declinantes a partir da defasagem $q+1$ e picos de PACF significantes nas defasagens ' p ' do processo e declinantes a partir da defasagem $p+1$. O gráfico 1.6 mostra as funções ACF e PACF para o processo ARMA(1,1):

$$w_t = 0.6w_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$$
$$\varepsilon_t \sim N(0,1)$$

Nota-se que os picos de maior magnitude são os primeiros picos de ACF e PACF. Posteriormente, os picos começam a declinar. Porém, seria difícil distinguir visualmente entre processos ARMA(1,1) e ARMA(2,1). Por isso, seguimos para a próxima etapa.

A terceira etapa de Box e Jenkins (1976) consiste na escolha do modelo que melhor se adequa aos dados, estatisticamente. Existem alguns critérios de seleção de modelos que prezam pela redução da soma do quadrado dos resíduos (SQR), uma medida da parte

inexplicada do modelo. Porém, a adição de uma variável explicativa qualquer ao modelo sempre leva a uma redução de SQR ao custo de redução dos graus de liberdade para a estimação de todos os parâmetros do modelo devido à necessidade de estimação de um parâmetro adicional. Por isso, esses critérios penalizam pela perda de graus de liberdade devido à inclusão de uma variável explicativa adicional no modelo.

O *Akaike Information Criterion* (AIC) e o *Schwartz Bayesian Criterion* (SBC) são dados por:

$$AIC = N(\ln(SQR)) + 2k$$
$$SBC = N(\ln(SQR)) + k(\ln(N)),$$

em que N corresponde ao número total de observações disponíveis para a estimação e k corresponde ao número total de parâmetros estimados na regressão, ou seja (p+q+1) em um ARMA(p,q) com constante. Quanto menores forem os valores de AIC e SBC, melhor é a qualidade da estimação.

Suponham que não soubéssemos o processo gerador da série w(t) acima, dado por um ARMA (1,1). Visualmente poderíamos suspeitar de um ARMA (2,1), pelo tamanho do segundo pico da PACF. Podemos comparar os valores dos critérios AIC e SBC gerados por uma regressão ARMA(1,1) e por uma regressão ARMA(2,1) no Stata:

$$AIC(ARMA(1,1)) = 28640 < AIC(ARMA(2,1)) = 28642$$
$$SBC(ARMA(1,1)) = 28669 < SBC(ARMA(2,1)) = 28678$$

De fato, temos valores inferiores de AIC e SBC para o processo ARMA(1,1). Além disso, o coeficiente β_2 relacionado ao termo AR(2) não é estatisticamente significativo. Essas evidências nos levariam à escolha correta do modelo ARMA(1,1).

A quarta etapa de Box e Jenkins (1976) consiste na análise dos resíduos, os quais devem se comportar como um processo aleatório denominado *white-noise* em um modelo adequadamente especificado. Se os resíduos formarem um processo *white-noise*, não devem apresentar evidência de autocorrelação. Dessa forma, as funções de autocorrelação e autocorrelação parcial não devem possuir picos de magnitudes elevadas e estatisticamente significantes. Se houver evidência de alguns picos estatisticamente significantes, deve-se considerar um modelo alternativo univariado ou multivariado. Por exemplo, é possível incluir variáveis adicionais apontadas na literatura sobre o objeto de pesquisa, em modelos autoregressivos de defasagens distribuídas, apresentados na seção 1.5.

1.5 Modelos autoregressivos de defasagens distribuídas (*Autoregressive Distributed Lags (ADL)*)

Um modelo em que a variável dependente é função de suas próprias defasagens (autoregressivo) e de valores correntes e passados de um processo exógeno qualquer x_t (defasagens distribuídas) é chamado de modelo autoregressivo de defasagens distribuídas (*Autoregressive Distributed Lags (ADL)*). Podemos escrevê-lo da seguinte forma:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_m x_{t-m} + \varepsilon_t,$$

em que ε_t é um processo *white noise*.

Suponham que exista um modelo predizendo que alterações na variável x_t podem ter um impacto corrente e/ou de longo prazo na variável y_t . Além disso, a teoria disponível não prediz que a variável x_t seja, por sua vez, impactada por y_t ou correlacionada com o termo de erro ε_t . Dessa forma, podemos considerar x_t como um processo exógeno a y_t . Nessas circunstâncias, podemos estimar um modelo ADL para entender os impactos de x_t em y_t no curto e no longo prazos.

Em contraposição, se a teoria indicasse que x_t poderia também ser impactada por y_t , poderíamos, por exemplo, encontrar instrumentos válidos para x_t , a fim de estimar o seu impacto em y_t , em um modelo de variáveis instrumentais. Outra possibilidade seria estimar os impactos simultâneos de uma variável na outra em um modelo de vetor autoregressivo (VAR) ou em um modelo de vetor de correção de erros (VEC). Modelos de variáveis instrumentais, de vetor autoregressivo ou de vetor de correção de erros não serão abordados nesse capítulo, mas podem ser estudados em Enders (2008).

Para que os parâmetros $\{\beta_0, \beta_1, \dots, \beta_p, a_0, a_1, a_2, \dots, a_m\}$ sejam consistentemente estimados pelo método dos mínimos quadrados ordinários, é necessário que o processo x_t seja exógeno ou que $\{x_t, x_{t-1}, \dots, x_{t-m}\}$ não sejam correlacionados com o ruído do modelo ε_t .

Nesse modelo, uma variação marginal em x_t tem um impacto presente em y_t , equivalente a a_0 , também chamado de multiplicador de impacto (*impact multiplier*),

mas também possui um impacto na variável dependente nos próximos períodos $\{y_{t+1}, y_{t+2}, \dots\}$, que podemos medir se reescrevermos o processo no próximo período:

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \dots + \beta_p y_{t-p+1} + a_0 x_{t+1} + a_1 x_t + a_2 x_{t-1} + \dots + a_m x_{t-m+1} + \varepsilon_t$$

Substituindo-se por y_t , obtemos:

$$y_{t+1} = \beta_0 + \beta_1 (\beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_m x_{t-m} + \varepsilon_t) + \beta_2 y_{t-1} + \dots + \beta_p y_{t-p+1} + a_0 x_{t+1} + a_1 x_t + a_2 x_{t-1} + \dots + a_m x_{t-m+1} + \varepsilon_t$$

Dessa forma, uma variação marginal em x_t tem um impacto de $\beta_1 a_0 + a_1$ em y_{t+1} . Se continuarmos essa análise, reescrevendo o processo daqui a dois períodos y_{t+2} , e assim sucessivamente, concluiremos que o impacto marginal de x_t na variável dependente é de longo prazo, perdurando ao longo dos próximos períodos.

Se a teoria prediz que o processo x_t é exógeno, podemos encontrar o melhor modelo ADL para explicar as relações entre y_t e x_t . Em outras palavras, queremos encontrar a ordem autoregressiva ‘p’ e a ordem das defasagens distribuídas ‘m’ que melhor explicam as relações entre as duas variáveis.

De forma similar ao processo de Box e Jenkins (1976), seria recomendável a adoção do seguinte procedimento:

1. Realização de testes de não-estacionariedade em cada série, y_t e x_t , separadamente.

Se houver evidência de tendência estocástica, trabalhar com a primeira-diferença da série; se houver evidência de tendência determinística, trabalhar com o resíduo da regressão contra o termo de tendência.

2. Visualização da função de autocorrelação cruzada.

Há dois caminhos possíveis. Examinam-se os gráficos ACF e PACF de x_t . Se x_t parecer seguir um processo *white-noise*, adota-se o passo ‘a’. Caso contrário, adota-se o passo ‘b’.

- a. Se x_t for um processo *white-noise*, examina-se as correlações entre y_t e x_{t-k} para $k = 0, 1, 2, \dots$, as quais são denominadas correlações cruzadas (*cross-correlations*). Os pacotes estatísticos R, Stata e Eviews já possuem rotinas prontas para o cálculo

da função de correlação cruzada, com os valores das correlações no eixo y e os valores das defasagens no eixo x.

Em um modelo ADL, a função de correlação cruzada apresenta picos de magnitude elevada que se iniciam na primeira defasagem estatisticamente significativa k de x , x_{t-k} , importante para explicar y_t , e decaem posteriormente. Entretanto, não se pode concluir que devemos incluir no modelo todas as defasagens equivalentes a todos os picos $k, k+1, k+2$ de magnitude elevada, visto que esses picos também são influenciados pela ordem autoregressiva da variável dependente. Nesse passo, pode-se considerar todas as defasagens de magnitude relativamente elevada como candidatas a variáveis explicativas.

b. Se x_t não for um processo *white-noise*, não podemos aplicar o passo ‘a’. Por exemplo, se x_t for um processo AR(s), as defasagens ($x_{t-1}, x_{t-2}, \dots, x_{t-s}$) podem parecer bastante correlacionadas com y_t mesmo que apenas os valores correntes de x_t sejam importantes para explicar y_t . Nesse caso, pode-se extrair os choques exógenos de x_t , estimando-se, primeiramente, o melhor processo autoregressivo univariado AR(s) para explicar as suas variações:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-s} + u_t$$

Vimos que podemos determinar a ordem ‘s’ através das funções de autocorrelação e autocorrelação parcial. Podemos reescrever esse modelo através dos operadores defasagem ‘ L^1, L^2, \dots, L^i ’, que toma a primeira, segunda, ..., i-ésima diferença da série:

$$x_t = \beta_0 + \beta_1 L x_t + \beta_2 L^2 x_t + \dots + \beta_s L^s x_{t-s} + u_t$$

ou, equivalentemente,

$$(1 - \beta_0 - \beta_1 L - \beta_2 L^2 - \dots - \beta_s L^s) x_t = u_t,$$

em que u_t é um processo *white-noise* de choques exógenos de x_t , se o processo AR(p) for adequadamente especificado. O operador $\gamma^s = (1 - \beta_0 - \beta_1 L - \beta_2 L^2 - \dots - \beta_s L^s)$ transforma a sequência x_t em *white-noise*, visto que, se aplicado, resulta em u_t .

Considerando-se novamente o modelo ADL entre y_t e x_t :

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_m x_{t-m} + \varepsilon_t,$$

em que ε_t é um processo *white-noise*.

Equivalentemente, podemos aplicar o operador γ^s nos dois lados da equação a fim de escrever o processo y modificado, como uma função dos choques exógenos de x_t , correntes e defasados, $u_t, u_{t-1}, \dots, u_{t-m}$:

$$\gamma^s y_t = \beta_0 + \beta_1 \gamma^s y_{t-1} + \beta_2 \gamma^s y_{t-2} + \dots + \beta_p \gamma^s y_{t-p} + a_0 \gamma^s x_t + a_1 \gamma^s x_{t-1} + a_2 \gamma^s x_{t-2} + \dots + a_m \gamma^s x_{t-m} + \varepsilon_t$$

ou, equivalentemente,

$$\gamma^s y_t = \beta_0 + \beta_1 \gamma^s y_{t-1} + \beta_2 \gamma^s y_{t-2} + \dots + \beta_p \gamma^s y_{t-p} + a_0 u_t + a_1 u_{t-1} + a_2 u_{t-2} + \dots + a_m u_{t-m} + \varepsilon_t$$

Vemos que o processo modificado $\gamma^s y_t$ é uma função de choques *white-noise* de x_t . Ressalta-se que os parâmetros (a_0, a_1, \dots, a_m) são os mesmos que governam a relação entre y_t e x_t . Já que $(u_t, u_{t-1}, \dots, u_{t-m})$ são processos *white-noise*, podemos adotar a metodologia explicada no passo '2.a.', examinando-se as correlações cruzadas entre $\gamma^s y_t$ e $\gamma^s x_{t-k}$ para $k = 0, 1, 2, \dots$, a fim de compreender as defasagens de x_t que explicam y_t .

3. Uma vez determinadas as defasagens de x_t importantes para explicar y_t , resta entender as defasagens de y_t que entram como variáveis explicativas no modelo. Para isso, estima-se os parâmetros da regressão:

$$y_t = a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_m x_{t-m} + \omega_t,$$

em que ω_t é o termo de erro.

Examinam-se os gráficos ACF e PACF do processo estimado de ω_t , a fim de se entender o melhor modelo AR(p) (ou ARMA(p,q)) para explicar y_t .

4. Estima-se o modelo completo, incluindo-se as defasagens de x_t que explicam y_t , determinadas no passo 2, e as defasagens de y_t , determinadas no passo 3:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_m x_{t-m} + \varepsilon_t$$

Devido às possíveis correlações entre as defasagens de y_t e $(x_t, x_{t-1}, \dots, x_{t-m})$, os valores estimados de (a_0, a_1, \dots, a_m) podem ser diferentes daqueles obtidos

no passo 3. Se o modelo for corretamente especificado, o processo ε_t deve se comportar como *white-noise*.

1.6 A relação entre homicídios e desigualdade de renda: um modelo ADL

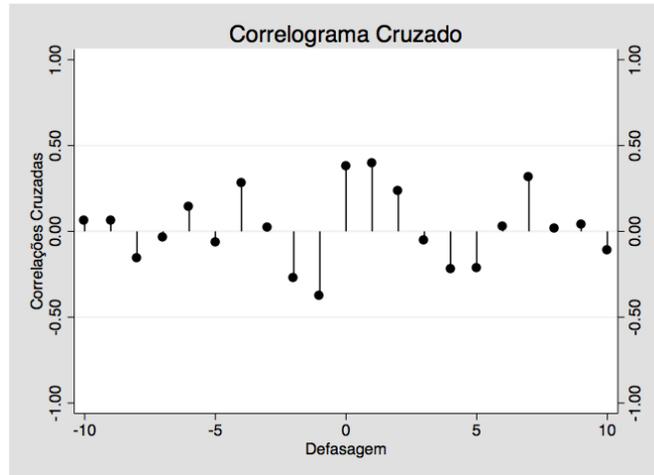
A literatura acadêmica sobre as relações entre violência e desigualdade socioeconômica é vasta. Fajnzylber, Lederman e Loayza (2002) apresentam evidência empírica do impacto positivo de um aumento da desigualdade socioeconômica em crimes violentos, tanto dentro de países quanto em uma análise entre diferentes países. Os resultados se mantêm mesmo quando se controla por outros fatores importantes para explicar crimes e pela possível endogeneidade da variável desigualdade. Wilkinson (2004) aponta que o fator determinante da violência mais bem estabelecido na literatura é o nível de desigualdade. O estudo mostra que uma possível explicação é o menor nível de qualidade das relações sociais em ambientes mais hierárquicos e desiguais.

Neste último tópico do capítulo, estudaremos as relações entre a taxa anual de homicídios no Brasil (por 100 mil habitantes) de 1980 a 2017, publicadas pelo IPEA no Atlas da Violência (Cerqueira et al, 2019), e o Índice de GINI no mesmo período, publicado pelo Banco Mundial.

Primeiramente, já vimos que não podemos rejeitar a hipótese nula de raiz unitária para as séries de homicídios e desigualdade de renda no Brasil. Dessa forma, aplicaremos o modelo ADL para a primeira diferença de homicídios e de GINI, Δhom_t e $\Delta GINI_t$.

Em segundo lugar, examinaremos as funções ACF e PACF de $\Delta GINI_t$. Como nenhuma autocorrelação ou autocorrelação parcial se apresenta estatisticamente significativa, não podemos rejeitar a hipótese nula de que a série se comporta como um processo *white-noise*. Nesse caso, utilizamos a etapa 2.a. em que examinaremos as correlações cruzadas (*cross-correlations*) entre Δhom_t e $\Delta GINI_{t-k}$ para $k = 0, 1, 2, \dots$. Observa-se o seguinte padrão:

Gráfico 1.7. Correlações Cruzadas



Fonte: Elaborado pelos autores.

No gráfico 1.7, estão plotados os valores das correlações entre Δhom_t e $\Delta GINI_{t-k}$ no eixo y e as defasagens $k = 0,1,2, \dots$ no eixo x. Nota-se que as correlações de magnitudes maiores ocorrem para $k=0,1$ e 2 . Por isso, considera-se um modelo possível com o valor corrente da desigualdade $\Delta GINI_t$, além de duas defasagens: $\Delta GINI_{t-1}, \Delta GINI_{t-2}$.

Passamos para a etapa 3, em que determinamos as defasagens de Δhom_t ou o melhor modelo ARMA(p,q) para explicar Δhom_t , além dos valores correntes e defasados de $\Delta GINI_t$. Para isso, estimamos a seguinte regressão:

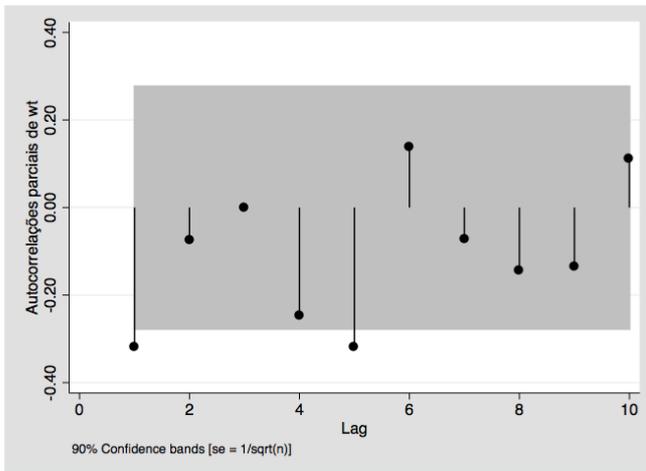
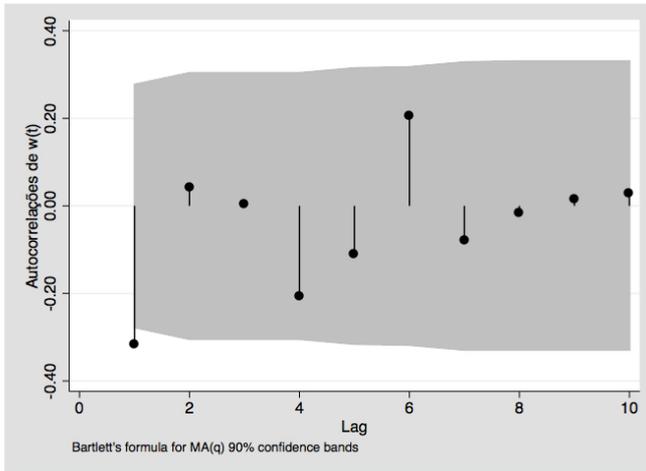
$$\Delta hom_t = 0.67 + 0.31\Delta GINI_t + 0.29\Delta GINI_{t-1} + 0.24\Delta GINI_{t-2} + \hat{\omega}_t \quad (3.57) \quad (3.36) \quad (2.70) \quad ,$$

em que $\hat{\omega}_t$ é o termo de erro estimado.

Assim, como apontado pelo correlograma estimado na etapa 2, $\Delta GINI_t, \Delta GINI_{t-1}$ e $\Delta GINI_{t-2}$ são estatisticamente significantes. Podemos incluir também defasagens adicionais, mas os parâmetros não são estatisticamente significantes.

Examinando-se as funções ACF e PACF de $\hat{\omega}_t$, observamos que nenhuma autocorrelação ou autocorrelação parcial se apresenta estatisticamente significativa ao nível de significância de 5%. Entretanto, dado o pequeno tamanho amostral, de 38 anos, estimamos as funções ACF e PACF, respectivamente, ao nível de significância de 10%.

Gráfico 1.8 Funções ACF e PACF ao nível de significância de 10%.



Fonte: Elaborado pelos autores.

A função ACF indica que deveríamos utilizar um modelo MA(1), ao passo que a PACF indica que deveríamos utilizar um modelo AR com defasagens 1, 4 e 5. Entretanto, como as defasagens 1, 4 e 5 se apresentam estatisticamente insignificantes, mesmo ao nível de 10%, estimamos o modelo apenas com $\Delta GINI_t, \Delta GINI_{t-1}, \Delta GINI_{t-2}$ e MA(1):

$$\Delta hom_t = 0.69 + 0.288\Delta GINI_t + 0.289\Delta GINI_{t-1} + 0.277\Delta GINI_{t-2} + \varepsilon_t - 0.399\varepsilon_{t-1},$$

em que todos os parâmetros estimados são significantes ao nível de 5% e os resíduos ε_t se comportam como um processo *white-noise*.

Finalizamos a nossa análise com a avaliação do impacto de um aumento da desigualdade, medida pelo índice GINI, no número de homicídios por 100.000 habitantes. O leitor deve se atentar que devido ao tamanho amostral, de apenas 38 anos, os resultados devem ser interpretados com cautela. Recomenda-se a utilização de uma amostra de tamanho superior a 100 para que os valores das estatísticas possam

convergir para os verdadeiros parâmetros populacionais em um modelo ARMA(p,q) ou ADL.

De acordo com o modelo estimado, o aumento de uma unidade na variação anual do índice GINI leva a um aumento corrente da variação de homicídios de 0.288. Esse aumento seria considerável, representando um aumento de 53.5% em relação ao valor médio de Δhom_t no banco de dados, equivalente à 0.538. Entretanto, o impacto perdura pelos próximos períodos, levando a um aumento da variação de homicídios de 0.289 no próximo período e a um aumento de 0.277 em dois anos. O impacto total seria de 0.854 (0.288+0.289+0.277) em Δhom_t em 2 anos, sendo nulo posteriormente. Entretanto, tendo em vista que a série original de homicídios hom_t possui raiz unitária, o aumento do número de homicídios devido a choques positivos em $\Delta GINI_t$ perduraria indefinidamente.

Conclusões

A análise de séries temporais é uma ferramenta extremamente útil para distintos estudos nas áreas de relações internacionais e ciência política, auxiliando no teste de hipóteses e na construção de teorias nesses campos. O desenho de pesquisa comparada de um único país abordado neste capítulo, baseado em séries temporais, permite a comparação da incidência de fenômenos políticos e sociais em diferentes momentos históricos de um mesmo país, como eleições, protestos, ciclos de negociações políticas, criminalidade, entre outras flutuações. Assim, as séries temporais fornecem dados para aprofundar o conhecimento da relação de variáveis relevantes para o contexto mais específico de um país, beneficiando-se de um nível de abstração menor das variáveis instrumentalizadas na análise desse país.

A utilização de dados históricos demanda uma série de cuidados metodológicos para que se estabeleça uma relação estatística adequada entre as variáveis. O presente capítulo buscou apresentar de maneira sistemática e sucinta os principais passos para a utilização estatística de séries temporais, focando na estimação de modelos univariados ARMA e multivariados ADL. Para ilustrar os passos fundamentais de estimação, utilizamos séries temporais de homicídios e índice de GINI no Brasil.

Dessa forma, este capítulo oferece um guia sucinto e ilustrado para que pesquisadoras e pesquisadores possam trabalhar com séries temporais e construir modelos estatísticos para testar hipóteses teóricas da ciência política e das relações internacionais.

Referências bibliográficas

- Barberia, Lorena. *Desenho de pesquisa em política comparada*. Brasília: ENAP, 2019.
- Beck, Nathaniel e Katz, Jonathan N. “What to do (and not to do) with time-series cross-section data”. *American political science review*, vol.89, n.3, p.634-647, 1995.
- Box, George E. e Jenkins, Gwilym M. *Time series analysis: Forecasting and control*. San Francisco: Calif: Holden-Day, 1976.
- Brandt, Patrick T., Freeman, John R. e Schrodt, Philip A. “Real time, time series forecasting of inter- and intra-state political conflict”. *Conflict Management and Peace Science*, vol.28, n.1, p.40–63, 2011.
- Cerqueira, Daniel et al. *Atlas da violência 2019*. Brasília: IPEA, 2019. Disponível em <https://www.ipea.gov.br/atlasviolencia/>. Acesso em 11 jun. 2021.
- Dickey, David e Fuller, Wayne. “Distribution of the estimators for autoregressive time series with a unit root”. *Journal of the American statistical association*, vol. 74 (366a), p. 427-431, 1979.
- Dickey, David e Fuller, Wayne. “Likelihood ratio statistics for autoregressive time series with a unit root”. *Econometrica: journal of the Econometric Society* vol. 49 (4), p. 1057-1072, 1981.
- Ehlers, Ricardo S. *Análise de Séries Temporais*. Curitiba: Departamento de Estatística, UFPR, 2005. Disponível em: <http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf> Acesso em 11 jun. 2021.
- Enders, Walter. *Applied econometric time series*. Columbia: Wiley, 2008.
- Fajnzylber, Pablo; Lederman, Daniel e Loayza, Norman. “Inequality and violent crime”. *The Journal of Law and Economics*, vol.45, n.1, p.1-39, 2002.
- Ferreira, Sérgio. “Segurança pública nas grandes cidades”. In Schwartzman, Simon e Bacha, Edmar (Orgs.). *Brasil: A nova agenda social*. Rio de Janeiro: LTC, 2011, p.287-318.
- Freire, Danilo. “Evaluating the effect of homicide prevention strategies in Sao Paulo, Brazil: A synthetic control approach”. *Latin American Research Review*, vol.53, n.2, 2018.
- Igarapé, Instituto. “Observatório de Homicídios”. 2018. Disponível em: <https://igarape.org.br/apps/observatorio-de-homicidios/>. Acesso em 24/10/2020.
- Jegadeesh, Narasimhan e Titman, Sheridan. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency". *The Journal of Finance*, vol.48, n.1, p.65–91, 1993.

- Landman, Todd. *Issues and methods in comparative politics: an introduction*. New York: Routledge. 3ed, 2008.
- Latinobarômetro. “Informe Latinobarômetro 2018”. Buenos Aires: Corporación Latinobarômetro, 2018.
- Lipset, Seymour Martin e Marks, Gary. *It Didn't Happen Here: Why Socialism Failed in the United States*. New York: W. W. Norton, 2000.
- Lo, Andrew W. e MacKinlay, A. C. *A Non-Random Walk Down Wall Street*. Princeton: Princeton University Press, 2002.
- Maoz, Zeev e Russett, Bruce. “Normative and structural causes of democratic peace, 1946-1986”. *American Political Science Review*, vol. 87 (3), p. 624-638, 1993.
- McDermott, Rose. "Experimental methods in political science." *Annual Review of Political Science*, n.51, p.31-61, 2002.
- Pevehouse, Jon e Brozek, Jason. “Times Series Analysis”. In Box-Steffensmeier, Janet M., Brady, Henry E., e Collier, David. (Eds.) *The Oxford Handbook of Political Methodology*, vol. 10, p. 456-474, 2008.
- Portella, Ana Portella e do Nascimento, Marília G. “Impactos de Gênero na Redução da Mortalidade Violenta: Reflexões sobre o Pacto pela Vida em Pernambuco”. *Revista Brasileira de Segurança Pública*, vol.8, n.1, p.48-68, 2014.
- Russett, Bruce e Oneal, John. *Triangulating peace. Democracy, Interdependence, and International Organizations*. New York : Norton, 2001.
- Zaverucha, Jorge e Nóbrega Júnior, José Maria Pereira da. “O Pacto pela Vida, os tomadores de decisão e a redução da violência homicida em Pernambuco”. *Dilemas-Revista de Estudos de Conflito e Controle Social*, vol.8, n.2, p.235-252, 2015.
- Whitten, Guy D. e Williams, Laron K. “Buttery guns and welfare hawks: The politics of defense spending in advanced industrial democracies”. *American Journal of Political Science*, vol.55, n.1, p.117-134, 2011.
- Wilkinson, Richard. “Why is violence more common where inequality is greater?” *Annals of the New York Academy of Sciences*, vol.1036, n.1, p.1-12, 2004.