



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia



IBAPE BAHIA
Instituto Brasileiro de
Avaliações e Perícias de
Engenharia da Bahia

Teste de Durbin Watson: Aplicação para Variável data do Evento

Maria Teresa Marques de Cerqueira Sobral; Larissa Mendes Rocha



TESTE DE DURBIN WATSON
APLICAÇÃO PARA VARIÁVEL DATA DO EVENTO

RESUMO

Um dos Métodos Avaliatórios mais respeitáveis dentro da Engenharia de Avaliações é a Inferência Estatística, que analisa através de modelos de regressão, com uma amostra representativa do mercado imobiliário local, os efeitos e influências das características intrínsecas de cada imóvel na determinação de seu valor de mercado, atendendo todo os parâmetros estatísticos e rigor Normativo. Para esta análise estatística, são realizados diversos testes que permitem balizar os resultados apresentados. Dentre os testes previstos em Norma, estão os testes de verificação de autocorrelação do modelo de regressão utilizado. O principal objetivo do presente trabalho é estudar o Teste de Durbin Watson e se o mesmo pode ser utilizado na análise de variáveis não temporais, a exemplo da variável “Data do evento” nos macromodelos, bem como diferenciar a Autocorrelação Espacial da Autocorrelação temporal e sugerir testes mais modernos que possam verificar a existência de Autocorrelação dos modelos de regressão avaliatórios.

***Avaliação de imóveis; Teste de Durbin Watson; Autocorrelação;
Autocorrelação Espacial; Variável “data do evento”***

1 INTRODUÇÃO

As Avaliações de Imóveis desempenham um papel de suma importância dentro da sociedade, vez que parametriza distintas negociações nos mais diversos setores da economia. Atualmente, existem diferentes metodologias utilizadas que determinam o valor de um bem de forma criteriosa e que atendem os preceitos Normativos, no entanto, a escolha da metodologia mais adequada para determinação do valor de Mercado de um bem depende das condições mercadológicas com que o Engenheiro avaliador se defronta e da quantidade de observações disponíveis no mercado. Verifica-se, contudo, que a grande maioria das Avaliações imobiliárias utilizam o Método Comparativo Direto de dados de Mercado - MCDDM, com tratamento de Dados através da inferência estatística. A aplicabilidade da inferência estatística em Engenharia de Avaliações é explicar o comportamento do mercado imobiliário em estudo, com base em uma amostra coletada neste mercado, à partir da premissa que conhecendo uma parte representativa do mercado pode-se concluir sobre o seu comportamento geral, com uma probabilidade de erro.

Dentro da Engenharia de Avaliações é muito usual o acúmulo de dados ao longo do tempo, como forma de aumentar a amostra do mercado imobiliário e por esta razão, se torna imprescindível o uso da variável "data do evento", o que pode trazer problemas para modelagem. Um dos Testes mais utilizados para verificação desta variável nos modelos avaliatórios é o Teste de Durbin-Watson. O objetivo deste trabalho é estudar este teste, a sua aplicabilidade, a influência sobre o resultado quando da resposta positiva e quais as soluções que podem ser utilizadas.

Para elucidar as questões até aqui apresentadas, se faz necessária também uma abordagem acerca dos modelos de regressão linear, estimados através de Mínimos quadrados Ordinários - MQO e suas propriedades, com destaque para o estudo da Autocorrelação, se fazendo mister diferenciar a autocorrelação temporal da autocorrelação espacial e qual delas é a existente nos modelos de regressão utilizados em avaliações de imóveis.

O presente estudo trará reflexões acerca do descarte de um modelo inferencial que atenda a todos os demais parâmetros estatísticos em função de uma resposta à auto-regressão positiva através do teste de Durbin-Watson, tal qual apresentar alternativas que minimizem os efeitos da presença da auto-correlação dentro de um modelo econométrico.

Portanto, este trabalho tem como base a pesquisa bibliográfica, além de trazer alguns exemplos práticos visando alcançar os objetivos propostos.

2 MODELOS DE REGRESSÃO COM O USO DA VARIÁVEL DATA DO EVENTO

É muito usual dentro da Engenharia de Avaliações os macromodelos que utilizam de diversas variáveis explicativas coletadas em períodos distintos para explicarem o valor de um imóvel. Cada tipologia de imóvel possui características próprias que influenciam no valor de mercado daquele bem e que é inferenciado através de um modelo de regressão no intuito de verificar o quanto cada característica é determinante neste valor. Na busca de agregar dados de mercado ao modelo estatístico, bem como com o intuito de verificação do comportamento do mercado ao longo do tempo, é comumente utilizada a variável "data do evento"

como uma variável explicativa de tendência, que pode apresentar dados em sequência ininterruptas ou com a presença de lacunas. Verifica-se também o uso desta variável com comportamento binário, quando a base de dados possui dois períodos diferentes, que podem ser dois semestres ou dois anos, por exemplo. Nos casos em que a variável data do evento apresenta uma seqüência de meses, semestres, anos, parte-se da hipótese de que o período mais antigo possui o menor valor e atribui-se 1 para a data mais antiga, 2 para a data imediatamente seqüencial à data 1, e assim sucessivamente. Em mercados recessivos, verifica-se uma inversão destes valores e é, inclusive, possível aferir quanto o mercado imobiliário apresenta de recessão.

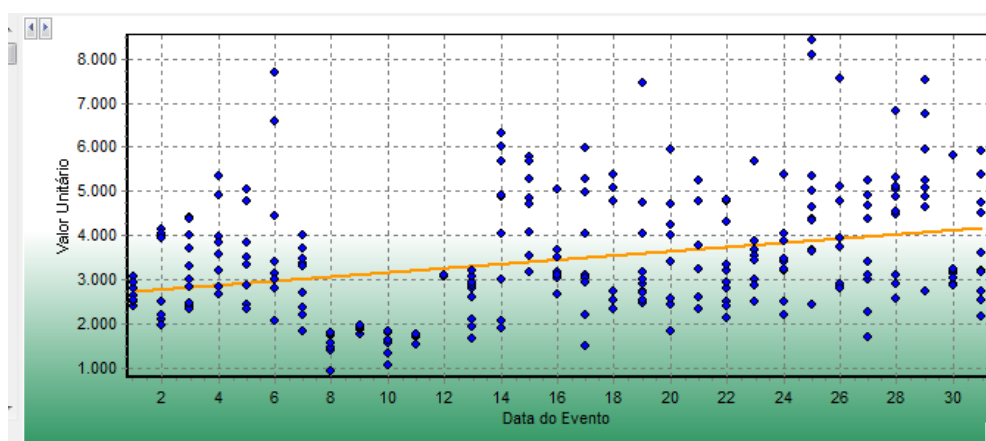


Gráfico 1 - Variável Data x Valor Unitário com crescimento positivo

A inclusão e análise dos dados inseridos nos modelos macroeconômicos que se utilizam da variável data do evento deve ser muito criteriosa. Segundo estudo de caso apresentado por Marcelo da Costa Teixeira e Vanessa Domingos Gonçalves Teixeira (2017) acerca deste assunto, o qual se valeu de análises através de média móvel e de curva polinomial ao longo dos períodos, foi constatado que esta variável, se inserida aleatoriamente pelo Engenheiro Avaliador com dados mais variados e com características assemelhadas ao seu avaliando, o que muda no decorrer dos períodos, ou seja, se em determinado mês o avaliador inserir mais dados de um imóvel padrão alto, que se aproxime de seu avaliando e no mês seguinte este inserir mais dados de um imóvel padrão baixo que se assemelhe a um segundo avaliando, o modelo passa a não refletir as informações do mercado e sim tendências sem nexos causal com o mercado imobiliário. Assim sendo, não se deve utilizar macromodelos com o emprego da variável data com a utilização de dados de terceiros e nem devem ser alteradas as características de preenchimento da pesquisa, sob o mesmo ponto de vista, os dados devem ser inseridos no modelo mantendo um equilíbrio entre a quantidade de dados para as demais variáveis explicativas.

Um dos testes mais utilizados para análise da variável data dentro de uma modelagem inferencial no universo avaliatório de imóveis é o Teste de estatística d, de Durbin Watson, que envolve o cálculo de um teste estatístico baseado nos resíduos do método de regressão de mínimos quadrados, verificando, de forma muito simplória, a presença de autocorrelação positiva no modelo. No entanto, ao se deparar com a presença da "auto-regressão positiva" no modelo inferencial, o

Engenheiro avaliador se descobre com a necessidade de descartar o modelo ou ajustá-lo, desabilitando dados ou variáveis explicativas. Opções estas que seguem em total desencontro ao caminho da econometria moderna, que vem buscando alternativas para não haver necessidade de descarte de dados e variáveis.

3 METODOLOGIA

3.1 Modelos de Regressão

3.1.1 Modelos de Regressão Linear Clássico

O modelo de regressão linear clássico busca encontrar uma relação estatística entre duas variáveis, X e Y, ou $y = f(x)$. Nestes tipos de modelos, pressupõe-se implicitamente que as relações causais, se as houver, entre as variáveis explicadas ou dependentes, só se dão em direção das variáveis explicativas ou independentes. É o modelo mais popular para estudar a relação entre duas variáveis, no qual os parâmetros de interesse são estimados a partir da minimização da soma dos quadrados dos resíduos. Estes estimadores são conhecidos como estimadores de mínimos quadrados ordinários (MQO). (MALBOUISSON; TIRYKI, 2017).

As hipóteses têm um papel importante na análise de relação de causalidade, tendo em vista o fato de não ser possível a utilização de toda população na grande maioria das vezes, razão pela qual utiliza-se de amostras aleatórias supostamente representativas do universo que se quer analisar, e para isso é necessário identificar algumas hipóteses básicas de um modelo de regressão com o intuito de garantir que as propriedades dos estimadores de MQO serão mantidas. Convém listar tais propriedades (pressupostos) do modelo:

- O modelo de regressão é linear nos parâmetros;
- O número de observações, n, deve ser superior ao número de parâmetros estimados pelo modelo;
- Os erros são variáveis aleatórias com valor esperado nulo e variância constante;
- Os termos de erros são normalmente distribuídos;
- **Os erros são não correlacionados, isto é, são independentes sob a condição de normalidade.**

Se qualquer uma dessas hipóteses estiver errada, existe um Erro de especificação. A violação dessas hipóteses pode ser causada por diferentes razões e as consequências da violação desta última hipótese é um dos objetos de estudo deste trabalho.

3.1.2 Modelos de Regressão Linear Múltipla

Amplia-se o modelo de regressão linear clássico à partir da suposição de que a variável dependente Y é uma função linear de uma série de variáveis explicativas X_1, X_2, \dots, X_k e um termo de erro. Em engenharia de avaliações geralmente se trabalha com modelos de regressão múltipla, tendo em vista a multiplicidade de fatores que interferem nos preços de um bem. (DANTAS, 1998).

Sob muitos aspectos, as premissas analisadas no modelo de regressão linear clássico, pode ser também utilizada no modelo de regressão linear múltipla. No entanto, novos testes precisam ser realizados. Pode-se citar duas novas hipóteses a serem verificadas, mas de forma alguma esse trabalho pretende adentrar em todos os testes que precisariam ser estudados num modelo de regressão múltipla:

- As variáveis independentes e o termo de erro são não correlacionados;
- Não há multicolinearidade perfeita, isto é, não há relações lineares perfeitas entre as variáveis.

Em Engenharia de Avaliações, considera-se geralmente como variável dependente os valores de mercado do bem avaliando e as variáveis independentes as respectivas características decorrente de seus aspectos físicos, de localização e de aspectos econômicos.

3.2 Propriedades Desejáveis dos Estimadores

3.2.1 Ausência de Tendenciosidade

Se diz que um estimador é não-tendencioso se a média ou o valor esperado de β^* for igual ao verdadeiro valor, ou seja, $E(\beta^*) = \beta$.

3.2.2 Eficiência

Espera-se que um estimador seja não tendencioso eficiente quando, para um dado tamanho da amostra, a variância de β^* é menor que a variância de qualquer outro estimador não-tendencioso. Algumas vezes é difícil julgar se um estimador é eficiente, então é natural descrever estimadores em termos de sua eficiência relativa, ou seja, um estimador será relativamente mais eficiente que outro.

3.2.3 Consistência

O Estimador β^* é considerado consistente quando se aproxima do verdadeiro β conforme o tamanho da amostra aumentasse. Ou, seja, espera-se que à medida que a amostra se torne muito grande, a probabilidade de que β^* seja diferente de β se torne muito pequena.

3.3 Estrutura de dados

Existe uma grande variedade de tipos de dados, mas este trabalho discorrerá apenas sobre os mais importantes encontrados nas literaturas.

3.3.1 Dados de Corte Transversal (dados em *cross-section*)

“Estes são dados em que uma ou mais variáveis foram coletadas no mesmo ponto do tempo”. (GUJARATI, 2006). “Trabalha-se com dados em *cross-section* sempre que se tem observações para várias unidades individuais (e.g. pessoas, empresas, municípios, estados, países) em um ponto no tempo”. (MALBOUISSON; TIRYAKI, 2017). “Consiste em uma amostra de indivíduos, consumidores, empresas,

idades, estados, países ou uma variedade de outras unidades, tomadas em um determinado ponto do tempo”. (WOODRIDGE, 2006). A palavra chave encontrada em todas as descrições aqui citadas é “determinado ponto do tempo”, ou seja, é o caso das Avaliações imobiliárias que se utilizam de uma amostra de dados coletada naquela data base. Sendo assim, pode-se definir que as avaliações imobiliárias utilizam-se de dados de corte transversal, como “fotos” do mercado naquele instante, com uma base de dados aleatórios, que se analisa as variáveis explicativas que têm relevância para determinação do valor do imóvel naquela data, tais como área privativa, localização, padrão construtivo, estado de conservação, idade aparente, dentre outras que são analisadas para cada tipologia de imóvel. Na estrutura de corte transversal, os dados de todas as unidades não correspondem precisamente ao mesmo período, tendo em vista que a pesquisa não acontece no mesmo dia, ou seja, por exemplo, muitos imóveis podem ser pesquisados durante diferentes semanas, mas ainda assim, ignoraria se, na coleta de dados, quaisquer diferenças de tempo não importantes na análise de cortes transversal. Segundo (WOODRIDGE, 2006):

Em uma análise pura de dados de corte transversal, ignoraríamos, na coleta de dados, quaisquer diferenças de tempo importantes. Se o conjunto de famílias fosse pesquisado durante diferentes semanas do tempo do mesmo ano, ainda veríamos isso como um conjunto de dados de corte transversal.

Um fato importante na utilização desse tipo de dados é que a ordenação dos dados, a partir da amostra aleatória, não é relevante. Ainda segundo (WOODRIDGE, 2006) "As observações estão ordenadas alfabeticamente por país, mas essa ordenação não afeta em nada qualquer análise subsequente", ainda se referindo à ausência de importância entre a ordenação dos dados na estrutura de cortes transversais. Trazendo este exemplo para a Engenharia de Avaliações, seria equivalente a dizer que ordenar os dados dentro do modelo por ordem alfabética, por bairros, por ordem crescente da área privativa, ou qualquer outra variável explicativa, não traria quaisquer prejuízo à modelagem.

3.3.2 Séries Temporais

“É um conjunto de observações dos valores que uma variável assume em diferentes momentos do tempo”. (GUJARATI, 2006). Exemplos de dados de séries temporais incluem preços de ações, oferta de moeda, índice de preços ao consumidor, Produto Interno bruto (PIB), taxas de ocupação de um resort, relatórios meteorológicos, taxa de desempregos, entre tantos outros. Esses dados são coletados em intervalos regulares, que pode ser diariamente, semanalmente, mensalmente, trimestralmente, anualmente, quinquenalmente (a cada cinco anos) ou decenalmente (censo demográfico). Existem ainda aqueles que são coletados em intervalos extremamente curtos, como exemplo das ações, que são obtidas em cotação de tempo quase real. Uma das características principais das séries temporais é a organização de diversas variáveis ao longo tempo, para uma única unidade individual.

“Um modelo de série temporal reflete o padrão de movimentos passados de uma variável e usa essa informação para prever seus movimentos futuros”. (PINDYCK; RUBINFELD, 2004). Sendo assim, um dos objetivos principais de uma

análise de série temporal é a possibilidade de previsões futuras com base no comportamento passado dessa mesma variável. Segundo William H. Greene (2003):

A característica de assinatura de um processo de séries temporais é que, empiricamente, o mecanismo gerador de dados produz exatamente uma realização da seqüência. Resultados estatísticos baseados em características de amostragem não dizem respeito à amostragem da população, mas a partir de distribuições de estatísticas construídas a partir de conjuntos de observações tiradas desta realização em uma janela de tempo, $t = 1, \dots T$.

Nos modelos de regressão com base de dados de séries temporais, umas das características mais relevantes é que as observações vizinhas são dependentes e o modelo de regressão busca modelar esta dependência. Nestes modelos, a ordem das observações é de extrema importância.

As séries Temporais podem apresentar características distintas. Podendo ser:

- Média
- Tendência Linear
- Sazonal
- Cíclica
- Aleatória

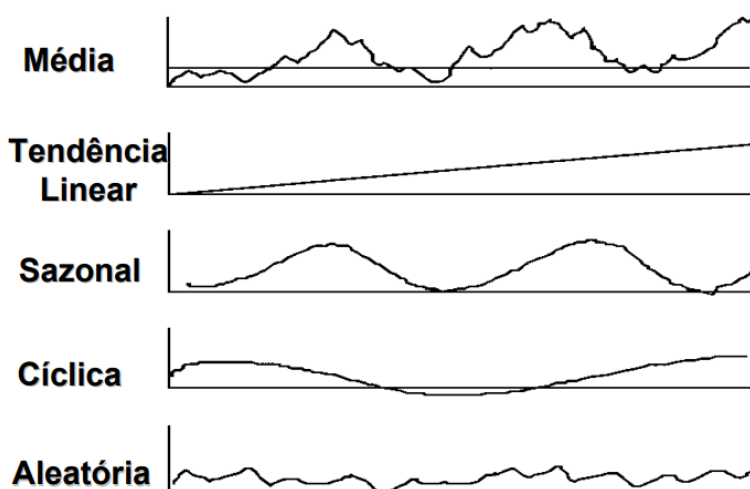


Gráfico 2 – Séries Temporais

O estudo das Séries Temporais requer uso de técnicas específicas e, embora não seja esse o objetivo deste trabalho, será mencionada de forma superficial algumas destas dificuldades encontradas nas análises de séries temporais, conforme citado por (GUJARATI, 2006):

- O trabalho empírico baseado em séries temporais pressupõe que a série temporal subjacente seja estacionária, ou seja, pode-se dizer de maneira geral, que sua média e sua variância não variam sistematicamente ao longo do tempo;
- Frequentemente verifica-se a existência de auto-correlação em séries não-estacionárias;

- Observações correlacionadas são mais difíceis de analisar e requerem técnicas específicas. Portanto é muito importante verificar se a relação entre as variáveis é espúria ou sem sentido;
- Presença de tendências, variações cíclicas ou sazonais podem ser difíceis de estimar ou remover;
- É mais difícil lidar com observações perdidas e dados discrepantes devido à natureza seqüencial;
- Algumas séries temporais financeiras, como preço das ações, exibem o que é conhecido como fenômeno do passeio aleatório. Isso significa que a melhor previsão de uma ação de amanhã, é seu valor de hoje acrescido de um termo de erro. O que não acontece na prática.

3.3.3 Agrupamento de Cortes Transversais

Alguns conjuntos de dados têm tanto características de corte transversal quanto série de tempo. Este tipo de estrutura ocorre quando um mesmo conjunto de variáveis é coletado em diferentes períodos de tempo, em distintas amostras aleatórias de uma mesma população. Por exemplo, suponha que dois estudos sobre família sejam realizados no Brasil com dados de corte transversal, um em 2013 e outro em 2017. Em 2013, uma amostra aleatória de família é pesquisada para variáveis tais como renda, poupança, tamanho da família, e assim por diante. Em 2017, uma nova amostra aleatória de família é extraída usando as mesmas questões da pesquisa.

A fim de aumentar o tamanho de amostra, pode-se formar um corte transversal agrupado ao combinar os dois anos. Sendo assim, os macromodelos utilizados em Avaliações imobiliárias, que se utilizam da variável data do evento como uma das variáveis explicativas de tendência, possuem estrutura de Agrupamento de dados de Cortes Transversais. Jeffrey M. Wooldridge (2006) trás um exemplo bem similar ao que é utilizado nos Macromodelos com a variável data do evento.

Como por exemplo, considere o seguinte conjunto de dados sobre os preços da moradia coletados em 1993 e 1995 nos EUA, quando houve uma redução nos impostos sobre a propriedade em 1994. Suponha que tenhamos dados sobre 250 residências para 1993 e sobre 270 para 1995 e vamos tabelar esses dados utilizando a uma nova variável (ano) separada. Embora a ordem na qual armazenamos os dados não se revele crucial, não se esqueça de que o ano de cada observação é, geralmente, muito importante. Essa é a razão de introduzimos ano como uma variável separada.

Com a utilização desse tipo de dados, além de aumentar o tamanho da amostra, a análise do corte transversal agrupada é importante para estimar como uma relação fundamental mudou ao longo do tempo. Podendo obter ainda, estimadores mais precisos e estatísticas de testes mais poderosas. No entanto, segundo Jeffrey M. Wooldridge (2006), o agrupamento só é útil nesse sentido, se a relação entre a variável dependente e pelo menos uma das variáveis independentes permanecerem constantes ao longo do tempo.

3.3.4 Dados de Painel ou Longitudinais

“Consiste em uma série de tempo para cada membro do corte transversal do conjunto de dados”. (WOODRIDGE, 2006). “São estruturados com observações de variáveis de diferentes unidades em cross-section ao longo de um mesmo período de tempo, sendo que se tem mais de uma observação temporal por unidade de *cross-section*”. (MALBOUISSON; TIRYAKI, 2017). A característica essencial dos dados de painel que os distingue dos dados de corte transversal agrupado é o fato de que as mesmas unidades de corte transversal são acompanhadas ao longo de um determinado período. Podemos usar como exemplo de dados de painel a criminalidade de determinados municípios ao longo dos anos, desde que a análise seja dos mesmos municípios.

Outro exemplo seria o estudo de desempenho escolar de um mesmo conjunto de indivíduos durante o período fixo. No mercado imobiliário, não se trabalha com dados de painel, tendo em vista que os imóveis coletados ao longo do tempo são dados aleatórios e não os mesmo dados. Uma vantagem dos dados de painel é que ele, freqüentemente, permite estudar a importância das defasagens do comportamento ou resultado de tomar decisões. Outra vantagem do uso de dados de painel é que podem ser respondidas questões que outro tipo de estrutura de dados não permite.

Diante das classificações das estruturas de dados supracitadas, verifica-se que os macromodelos de regressão que apresentam uso da variável data do evento se enquadram como Cortes Transversais Agrupados, visto que os dados coletados ao longo do tempo são aleatórios e que existe uma combinação entre períodos diferentes de vários cortes transversais. É bom ressaltar que nestes modelos a ordem dos dados não se revela crucial, no entanto, a determinação da seqüência exata da variável data do evento é de extrema importância.

3.4 Autocorrelação

Como já apresentado anteriormente, uma dos pressupostos que os modelos de regressão precisam atender é a ausência de correlação entre os erros, isto é, os erros são independentes sob a condição de normalidade. A autocorrelação pode ser definida como “correlações entre integrantes de séries de observações ordenadas no tempo, como as séries temporais, ou no espaço, como nos dados de corte transversais” (GUJARATI, 2006). Nesse sentido, a autocorrelação existe quando um termo de erro relacionado a qualquer das observações é influenciado pelo termo de erro de qualquer outra observação. Ainda segundo Damodar Gujarati (2006), nas estruturas de cortes transversais, os dados são coletados através de amostras aleatórias, de forma que não há razões, *à priori*, para considerar que o termo de erro de uma observação aleatória teria qualquer correlação com o termo de erro de outra observação e quando esta autocorrelação é verificada, é denominada de Autocorrelação espacial, isto é, correlação no espaço e não ao longo do tempo. Jeffrey M. Wooldridge (2006) afirma que, do ponto de vista estatístico, quando se trabalha com dados em corte transversal ou agrupamento de dados de corte transversal, as observações amostrais são independentes, de forma que elimina-se a correlação nos erros entre as diferentes observações.

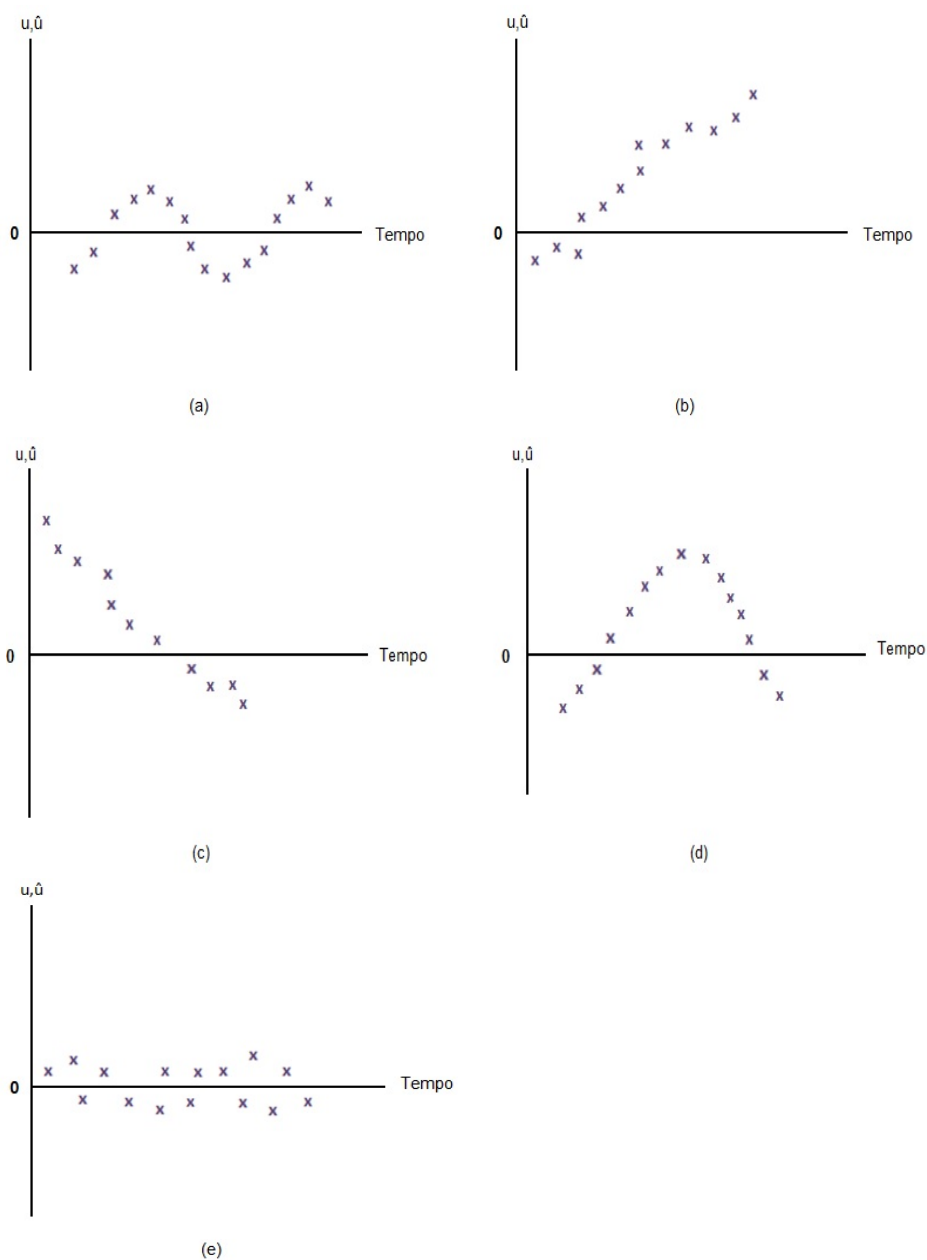


Gráfico 3 – Padrões de presença de autocorrelação (a, b, c, d) e ausência de autocorrelação (e). Fonte: Gujarati, 2000, pg 403

No entanto, quando se trabalha com séries temporais, os dados seguem um ordenamento natural ao longo do tempo e nestes casos, a pressuposição de que os erros correspondentes a observações diferentes não estão correlacionados, não se mantém. Sendo assim, nas séries temporais é muito difícil assumir que esta premissa seja atendida, pois na grande maioria das vezes as séries estão fortemente relacionadas com seus históricos anteriores. A correlação serial ocorre em estudos de séries temporais quando os erros associados com observações em um dado período de tempo se mantêm por transferência nos períodos futuros. (PINDYCK; RUBINFELD, 2004). A título de exemplo, numa análise de série temporal trimestral, se ocorre uma greve que interfere na produção de um trimestre, se a

perturbação causada pela greve permanecer no trimestre seguinte, verifica-se aí a existência de autocorrelação.

A presença da autocorrelação não afetará a ausência de tendenciosidade ou a consistência dos estimadores pelo MQO, mas interfere na eficiência, ou seja, eles deixam de ter variância mínima. Segundo Robert S. Pindyck e Daniel L. Rubinfeld (2004), no caso da autocorrelação positiva, a perda de eficiência resulta em erros-padrão da regressão com tendência para baixo, mesmo que os estimadores continuem não-tendenciosos, o que pode acabar por induzir a uma conclusão de que as estimativas dos parâmetros são mais precisas que de fato são.

Dentro da Engenharia de Avaliações, a expectativa é que cada transação se realize independentemente da outra, ou seja, que os dados de mercado sejam independentes, de forma que o preço ofertado por um imóvel não interfira no outro. Segundo (Dantas,1998) "O conceito de independência dos resíduos está ligado à independência dos dados de mercado. A situação ideal é aquela onde cada transação se realiza independentemente da outra. Isto é, o conhecimento do preço e condições de uma não interfira na outra." A NBR 14.653-2, não é clara acerca da análise da autocorrelação. Embora o assunto seja tratado dentro do Anexo A, que aborda os procedimentos para utilização de modelos de regressão linear, e ressalta a necessidade de análise dos pressupostos básicos dos modelos de regressão, esta orienta que o exame da autocorrelação deve ser precedida pelo pré-ordenamento dos elementos amostrais, ou seja, das observações coletadas no mercado, em relação aos valores ajustados e, se for o caso, às variáveis independentes possivelmente causadoras do problema. A Norma não menciona em quais condições a análise da autocorrelação deve ser verificada e, tendo em vista o fato da estrutura de dados utilizada em modelos de regressão nas avaliações imobiliárias ser de corte transversal ou agrupamento de cortes transversais, não há o que se falar em autocorrelação, a não ser que seja a autocorrelação espacial. Ademais, a Norma sugere ainda que faça o pré-ordenamento das variáveis independentes possivelmente causadoras do problema, desprezando por completo todas as demais razões que podem gerar uma autocorrelação, a exemplo da inércia, do viés de especificação no caso das variáveis excluídas e da forma funcional incorreta, defasagens, manipulação dos dados, entre outros.

3.4.1 Autocorrelação Espacial

A diferença entre a Econometria espacial e a convencional concentra-se na preocupação de explicar os efeitos espaciais na regressão. "Um modelo clássico de regressão linear convencional tem a limitação de não ser capaz de controlar esses efeitos espaciais". (ALMEIDA, 2012). De forma que se faz necessária a verificação se os pressupostos de uma modelagem pelo MQO se sustentam na presença dos efeitos espaciais.

A dependência espacial, que se trata de um efeito espacial, é "um caso especial de *Cross sectional dependence*, que aflora quando as unidades de corte transversal -indivíduos, domicílios, empresas ou regiões - não são mais independentes entre si". (ALMEIDA, 2012). Sendo assim, a dependência espacial significa que a variável dependente (i) - numa determinada região, não é explicada apenas pelas suas variáveis explicativas, mas também pelo valor dessa variável dependente (j) - nas regiões vizinhas. No entanto, numa situação análoga, verifica-

se também que a variável dependente (j), depende da variável dependente (i), numa situação de simultaneidade espacial.

Para atender o pressuposto da independência dos erros, é de suma importância manter a aleatoriedade dos dados. No entanto, quando se estuda um mercado que possui dados em cortes transversais ou cortes transversais agrupados, distribuídos em regiões distintas, se torna improvável que essa independência se mantenha. Eduardo Almeida (2012) assim descreve:

Quando se assume que os dados com os quais se trabalha são provenientes de uma amostragem aleatória, esta hipótese é satisfeita. Entretanto, quando se trabalha com unidades de corte transversal que são unidades geográficas, tais como bairros, setores censitários, distritos, regiões urbanas, setores postais, zona de tráfego, regiões de planejamento, municípios, microregiões, estados ou países, muito provavelmente, esta hipótese é violada.

Segundo artigo publicado por Francisco Aranha (1999), a autocorrelação espacial refere-se à uma redundância de informação entre duas realizações de um fenômeno quando elas ocorrem próximas uma da outra. Sua presença distorce os resultados obtidos pela aplicação de modelos estatísticos tradicionais, baseados na hipótese de independência entre as observações e a variável dependente.

A autocorrelação espacial pode ocorrer, basicamente, de duas formas, na variável dependente ou nos erros. Segundo Eduardo Almeida (2012):

Em séries de tempo, a violação desta hipótese acarreta a ineficiência do estimador MQO. Em processos espaciais, a violação desta hipótese não apenas implica ineficiência, mas dependendo da forma da dependência dos erros, pode acarretar viés e inconsistência do estimador MQO. Portanto, trata-se de um erro mais grave.

É importante destacar a principal diferença entre a autocorrelação temporal e autocorrelação espacial, sendo que os dados organizados no tempo são relativamente mais fáceis de analisar do que os dados organizados geograficamente. Ao contrário do espaço, o tempo flui em um único sentido, unidirecional, pois em séries temporais o passado pode influenciar o presente, mas o futuro, que ainda não aconteceu, não pode fazer o mesmo. Já no espaço todas as direções, e todos os sentidos, podem ser relevantes, dessa forma, a dependência espacial possui multidireções, ou seja, o valor de imóvel pode ser influenciado pelo seu vizinho a norte, a sul, a leste, a oeste, etc. "A direção da influencia no tempo é para frente, ao passo que, no espaço, a direção de influencia pode ser para frente, para trás, para cima ou para baixo". (BEENSTOCK; FELSENSTEIN, 2009, apud ALMEIDA, 2012)

Com base nos argumentos já trazidos neste estudo, o mercado imobiliário possui características de unidade geográfica, com seus dados distribuídos num mapa de acordo com a abrangência da pesquisa do avaliador, bem como apresenta estrutura de dados de cortes transversais ou cortes transversais agrupados. Sendo assim, se verifica, quando houver, a autocorrelação espacial e não a autocorrelação temporal. Pode-se citar a título de exemplo, o valor do metro quadrado de dois bairros localizados próximos um do outro em uma região do município que seja mais rica do que foi possível captar pelas variáveis incorporadas ao modelo, possivelmente apresentará uma autocorrelação positiva. Por outro lado, o contrario possivelmente ocorrerá para outros bairros localizados em regiões mais pobres do município onde o valor do metro quadrado pode não ter sido captado de forma efetiva pelas variáveis incluídas no modelo, de forma que os erros estocásticos

desse modelo econométrico possivelmente serão espacialmente correlacionados. Sendo o primeiro caso com autocorrelação positiva e o segundo caso com autocorrelação negativa.

Rodrigo de Souza Vieira (2009) informa que:

Para Anselin & Bera (1998), a autocorrelação espacial pode ser definida como a coincidência entre valores similares e similaridades locais. Assim, quando altos ou baixos valores para uma variável aleatória tendem a agrupar-se no espaço, temos o processo de autocorrelação espacial positiva. No entanto, pode acontecer também das unidades espaciais serem circundadas por unidades com valores significativamente distintos, ou seja, pode ocorrer que altos valores sejam acompanhados por vizinhos com valores baixos, ou vice-versa, processo que se denomina autocorrelação espacial negativa.

Embora os dois processos sejam igualmente importantes e dignos de consideração, a autocorrelação espacial positiva é, sobretudo, a mais intuitiva, e é encontrada, com maior frequência nos fenômenos econômicos. Na maior parte das vezes, um processo que apresenta autocorrelação espacial negativa é de difícil interpretação.

Seja positiva ou negativa, a presença de autocorrelação influencia na qualidade da informação contida em dados espacializados, afetando as possíveis interpretações dos modelos estatísticos.

Como a autocorrelação surge pela influência das realizações vizinhas sobre a realização do fenômeno na área de interesse, sua intensidade geralmente decai com a distância, tornando-se irrelevante quando as observações estão suficientemente afastadas entre si. (ARANHA, FRANCISCO, 1999).

Em outras palavras, o valor de um imóvel tende a não mais influenciar o valor do outro quando a distância entre eles aumenta.

3.4.2 Consequências dos Erros Autocorrelacionados

Segundo J. Johnston (1986), três consequências principais são apresentadas nos modelos de regressão pelo MQO com a presença de autocorrelação:

Primeira, embora se obtenha estimadores não-viesados de β , as variâncias amostrais desses estimadores podem ser excessivamente grandes, quando comparadas com aquelas que podem ser obtidas por um método de estimação ligeiramente diferente. Segunda, se for aplicada as fórmulas usuais de mínimos quadrados para as variâncias amostrais dos coeficientes de regressão, provavelmente se terá uma série subestimada dessas variâncias. De todo modo, essas fórmulas não tem mais validade, nem os testes t e F. Terceira, será obtido estimativas ineficientes, isto é, estimativas com variâncias amostrais desnecessariamente grandes.

3.5 Teste de Durbin-Watson

Já foram apresentadas neste trabalho todas as consequências negativas da presença da autocorrelação em um modelo de regressão e por esta razão, é muito importante testar a sua presença e, se necessário, adotar processos alternativos de estimação. Um dos testes mais conhecidos para verificação de autocorrelação temporal é a estatística d, de Durbin Watson, que envolve o cálculo de um teste estatístico baseado nos resíduos do método de regressão de mínimos quadrados.

Sejam \hat{e} ($t = 1, 2, 3, \dots, n$), os resíduos da regressão ajustada por Mínimos Quadrados, então se tem a razão entre a soma das diferenças, elevadas ao quadrado, entre sucessivos resíduos e a soma do quadrado dos resíduos (SQR):

$$\text{Estatística } d = \frac{\sum_{t=2}^n (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^n \hat{e}_t^2}$$

Durbin e Watson tabularam os limites inferiores, d_L , e os limites superiores, d_S , para vários valores de n (número de dados) e k (número de variáveis explicativas), tais que, se o d calculado estiver fora desses valores críticos, é possível a verificação da autocorrelação.

Para efetuar um teste de autocorrelação positiva, calcula-se d .

- Quando valores sucessivos de \hat{e}_t estão próximos um dos outros, a Estatística d será baixa, indicando a presença de correlação serial positiva, ou seja, se $d < d_L$, rejeita-se a hipótese de erros aleatórios e aceita a autocorrelação positiva;
- Se $d > d_U$, não se rejeita a hipótese e não existe autocorrelação ($\rho = 0$);
- Se $d_L < d < d_U$, o teste é inconclusivo e idealmente seriam necessárias observações adicionais;

A interpretação exata da estatística de Durbin-Watson é difícil porque a seqüência dos termos de erro depende não só das seqüências dos \hat{e} , mas também da seqüência de todos os valores de X . (PINDYCK; RUBINFELD, 2004)

3.5.1 Análise do Teste de Durbin-Watson

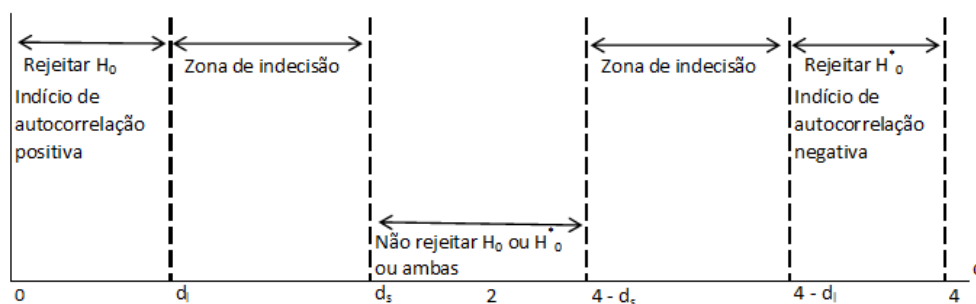


Gráfico 4 – Teste de Durbin Watson para autocorrelação dos resíduos

Na análise do Gráfico 4, verifica-se que a Estatística de Durbin – Watson está na faixa entre 0 a 4, sendo que um valor próximo de 2 indica que não há autocorrelação. Através de várias aproximações é possível mostrar que $DW = 2(1 - \rho^*)$. Assim, quando não há autocorrelação ($\rho = 0$), a estatística de Durbin – Watson será próxima de 2. Uma autocorrelação positiva está associada com valores de DW abaixo de 2, e uma autocorrelação negativa, com valores de DW acima de 2. Se o propósito é investigar a possibilidade de autocorrelação positiva, um valor da estatística de DW abaixo de d_L permite

rejeitar a hipótese nula de que não há autocorrelação. Se a estatística de DW é superior a d_s , a hipótese nula se mantém. A faixa entre d_L e d_s apresenta resultados inconclusivos. Para análise a autocorrelação negativa, a hipótese nula é rejeitada se a estatística de DW é maior que $4 - d_L$, e a hipótese é aceita se DW é menor que $4 - d_s$. No interior da faixa entre d_L e d_s o teste é inconclusivo.

A distribuição de d para amostras finitas depende dos Regressores. Dessa forma, é comum utilizar tabelas que fornecem apenas *intervalos* para os valores críticos determinados pelas distribuições limites inferior e superior d_L e d_U , respectivamente. Quando o valor calculado de d pertencer aos intervalos $[d_{LC}, d_{UC}]$ ou $[4 - d_{UC}, 4 - d_{LC}]$ torna-se impossível decidir sobre a rejeição ou não da hipótese nula de não-autocorrelação dos erros (zona de indecisão (CABRAL; LINS, 2011)).

Segundo Damodar N. Gujarati (2006), diante da facilidade de se encontrar o teste de estatística “d” de Durbin Watson, se passou a ser usual a informação deste resultado, quando da informação de outras medidas sintéticas, tais como o R^2 , o R^2 ajustado, os testes “t” e “F” nas análises de regressão. No entanto, ele é taxativo ao elencar as premissas que embasam a estatística d. Sendo elas:

1. O modelo de regressão inclui o termo de intercepto. Se este não tiver presente, como no caso do modelo que passa pela origem, é necessário refazer a regressão para incluir o intercepto antes de obter a SQR.;

2. As variáveis explanatórias, os X, são não estocásticas ou fixadas em amostras repetidas;

3. Os termos de erro, u, são gerados pelo esquema auto-regressivo de primeira ordem: $u = \rho u_{t-1} + \varepsilon_t$. Portanto não podem ser empregados para detectar esquemas auto-regressivos de ordens mais elevadas;

4. Pressupõe-se que o termo de erro u, seja normalmente distribuído;

5. O modelo de regressão não inclui os valores defasados da variável dependente como uma das variáveis explanatórias. Portanto, o teste não pode ser aplicado a modelos do seguinte tipo: $Y_t = \beta_1 + \beta_2 X_{2T} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t$

Onde Y_{t-1} é o valor de Y com defasagem de um período. Esses modelos são conhecidos como modelos auto-regressivos.

6. Não há falta de observações nos dados. Assim, na regressão dos salários contra a produtividade no período de 1959-1998, se, por exemplo, faltasse por alguma razão os dados relativos a 1978 e 1982, a estatística d não levaria em conta a falta dessas observações.

Embora alguns dos itens transcritos anteriormente não sejam objeto deste estudo, os itens 2 e 6 corroboram a tese que se quer provar neste trabalho, em virtude dos modelos de regressão utilizados nas avaliações de imóveis pelo Método Comparativo com uso da variável data do evento não trabalham com a mesma amostra de dados, e sim com amostras aleatórias coletadas em períodos diferentes do tempo, em um processo estocástico e com uma estrutura de dados em Agrupamento de Cortes Transversais, bem como nos modelos de regressão utilizados em Avaliações, a ordem das observações é irrelevante para a análise. Ademais, a grande maioria dos modelos de regressão com uso desta variável não possuem uma seqüência ininterrupta na variável data do evento, com a presença de várias lacunas de tempo.

O quanto apontado no item 6 de Damodar N. Gujarati (2006), foi objeto do artigo o qual afirma:

Outro pressuposto importante é a presença de observações para todos os períodos considerados. Por exemplo, se os dados são bimestrais para um

período de 3 anos, a base de dados deve apresentar um total de 18 observações, uma para cada bimestre. Caso haja a omissão de algum dado, o teste não é capaz de detectar autocorrelação serial de primeira ordem de maneira confiável. (CABRAL; LINS, 2011).

Násser Júnior Radegáz (2011), afirma que o teste de Durbin Watson é utilizado “quando estamos trabalhando com série temporal – elementos coletados ao longo de um determinado tempo, meses, anos, etc”.

Jeffrey M. Wooldridge (2006), aborda o teste de Durbin-Watson dentro do capítulo de Correlação Serial e Heteroscedasticidade em Regressões de Séries Temporais, de forma que deixa subtendido que este teste se aplica no caso de séries temporais concluindo que:

“O fato de que uma distribuição amostral exata de Durbin Watson (DW) pode ser tabelada é a única vantagem que DW tem sobre o teste t (da Correlação Serial AR). Dado que os valores críticos tabulados são válidos somente sob o conjunto total das hipóteses do MLC (Modelo Linear Clássico) e que eles podem levar a uma ampla região inconclusiva, as desvantagens práticas da estatística DW são substanciais.”

Em toda literatura aqui estudada, sob uma análise rigorosa, se verificou a afirmativa que o Teste de Durbin-Watson se aplica em casos de séries temporais de primeira ordem, não sendo aplicado em outras estruturas de dados. Verifica-se ainda a afirmativa de Jeffrey M. Wooldridge (2006) que a única vantagem deste teste é possuir a distribuição amostral tabelada, pois por apresentar uma grande região inconclusiva, as desvantagens se apresentam de forma elevada.

4. APLICAÇÃO PRÁTICA

4.1 Modelo Padrão

4.1.1 Banco de Dados

Objetivando estudar de forma prática todo o conteúdo bibliográfico até então apresentado, elaborou-se um macromodelo inferencial, estimado através do MQO, com uma pesquisa de 292 (duzentos e noventa e dois) dados coletados no Município do Salvador, nos mais diversos bairros da cidade, com a presença da variável "data do evento", entre junho/2016 a dezembro/2018, totalizando 31 meses.

Para explicar o valor de mercado do imóvel, foram consideradas as seguintes variáveis explicativas:

Variável	Formato	Descrição
Área privativa	Quantitativa Contínua	Área privativa do imóvel em m ²
Setor Urbano 14	Proxy	Refere-se ao índice fiscal dos logradouros dos imóveis
Nível	Quantitativa Discreta	Pavimento em que se encontra o imóvel
Oferta/Venda	Dicotômica	1 valor de oferta, 0 for valor de venda
Nº de Sanitários	Quantitativa Discreta	Número de sanitário do imóvel

Idade Aparente	Quantitativa Discreta	Idade aparente do imóvel
Quadra	Dicotômica	1 se o imóvel possui quadra, 0 caso contrário
Padrão de Fachada Normal	Dummy	1 se o imóvel possui padrão de fachada normal, 0 caso contrário
Padrão de Fachada Normal Alto	Dummy	1 se o imóvel possui padrão de fachada normal alto, 0 caso contrário
Padrão de Fachada Alto	Dummy	1 se o imóvel possui padrão de fachada alto, 0 caso contrário
Conservação de Fachada Boa	Dummy	1 se o imóvel possui conservação de fachada boa, 0 caso contrário
Imóvel Novo	Dummy	1 se o imóvel for novo, 0 caso contrário
Elevador	Dicotômica	1 se o imóvel possui elevador, 0 caso contrário
WC de Serviço	Dicotômica	1 se o imóvel possui sanitário de serviço, 0 caso contrário
Total de Vagas	Quantitativa Discreta	Número de vagas para veículos na garagem
Data do Evento	Variável de tendência	Assume valores de 1 a 31, referentes aos meses dos dados coletados
Piscina	Dicotômica	1 se o imóvel possui piscina, 0 caso contrário
Salão de Ginástica	Dicotômica	1 se o imóvel possui salão de ginástica, 0 caso contrário

4.1.2 Da Presença da Auto-Regressão Positiva

4.1.2.1 - Hipótese Inicial

Da análise da modelagem, verificou-se que a hipótese inicialmente formulada acerca do mercado pesquisado, as variáveis explicativas "Quadra", "Elevador" e "Salão de Ginástica", não interferem na formação do valor, sendo desta forma desabilitadas. Verificou-se ainda a presença de Auto-regressão positiva para a variável data do evento, com d calculado de 1,43.

<input checked="" type="checkbox"/>	Variáveis	Equação	Relação	t Calcul...	Signific...
<input checked="" type="checkbox"/>	Área Privativa	ln(x)	-11,87	-10,56	0,01
<input checked="" type="checkbox"/>	Setor Urbano 14	x	6,71	8,34	0,01
<input checked="" type="checkbox"/>	Nível	x	2,92	3,38	0,08
<input checked="" type="checkbox"/>	Oferta/Venda	x	5,67	2,24	2,55
<input checked="" type="checkbox"/>	Nº de Sanitários	x	3,59	4,17	0,01
<input checked="" type="checkbox"/>	Quadra	x	1,30	0,44	65,82
<input checked="" type="checkbox"/>	Idade Aparente	x	-1,04	-1,30	19,46
<input checked="" type="checkbox"/>	Pdr. Fach. Normal	x	10,38	2,82	0,51
<input checked="" type="checkbox"/>	Pdr. Fach. Normal Alto	x	20,55	3,58	0,04
<input checked="" type="checkbox"/>	Pdr. Fach. Alto	x	67,43	6,89	0,01
<input checked="" type="checkbox"/>	Cons. Fach. Boa	x	13,72	4,59	0,01
<input checked="" type="checkbox"/>	Imóvel Novo	x	14,79	2,86	0,45
<input checked="" type="checkbox"/>	Elevador	x	2,95	0,76	44,33
<input checked="" type="checkbox"/>	WC Serviço	x	10,31	3,28	0,11
<input checked="" type="checkbox"/>	Total de Vagas	x	3,86	3,81	0,01
<input checked="" type="checkbox"/>	Data do Evento	x	1,25	3,08	0,22
<input checked="" type="checkbox"/>	Piscina	x	12,93	3,01	0,28
<input checked="" type="checkbox"/>	Salão de Ginástica	x	2,94	0,82	41,22
<input type="checkbox"/>	Longitude				
<input type="checkbox"/>	Latitude				
<input checked="" type="checkbox"/>	Valor Unitário	ln(x)			

Dados e Variáveis	
Redes Neurais Artificiais	
Erro Quadrático	
Erro Quadrático Validação	
Interações	
Tolerância	
Taxa de Aprendizagem	
Coefficientes	
Correlação	1 - 0,9149247 / 0,9119418
Determinação	1 - 0,8370872 / 0,8316378
R2 Ajustado	1 - 0,8263457 / 0,8205370
Testes de Hipóteses	
F Calculado	77,93
Significância do Modelo	0,01
Durbin Watson	
d Calculado	1,43
Durbin Watson	Data do Evento
Normalidade dos Resíduos	
-1 e +1 desvios padrões	75%
-1,64 e +1,64 desvios padrões	90%
-1,96 e +1,96 desvios padrões	94%
Diversos	
Desvio Padrão	0,18852
Outliers do Modelo	17 (5,82%)
Opções de Cálculo	Simplificado

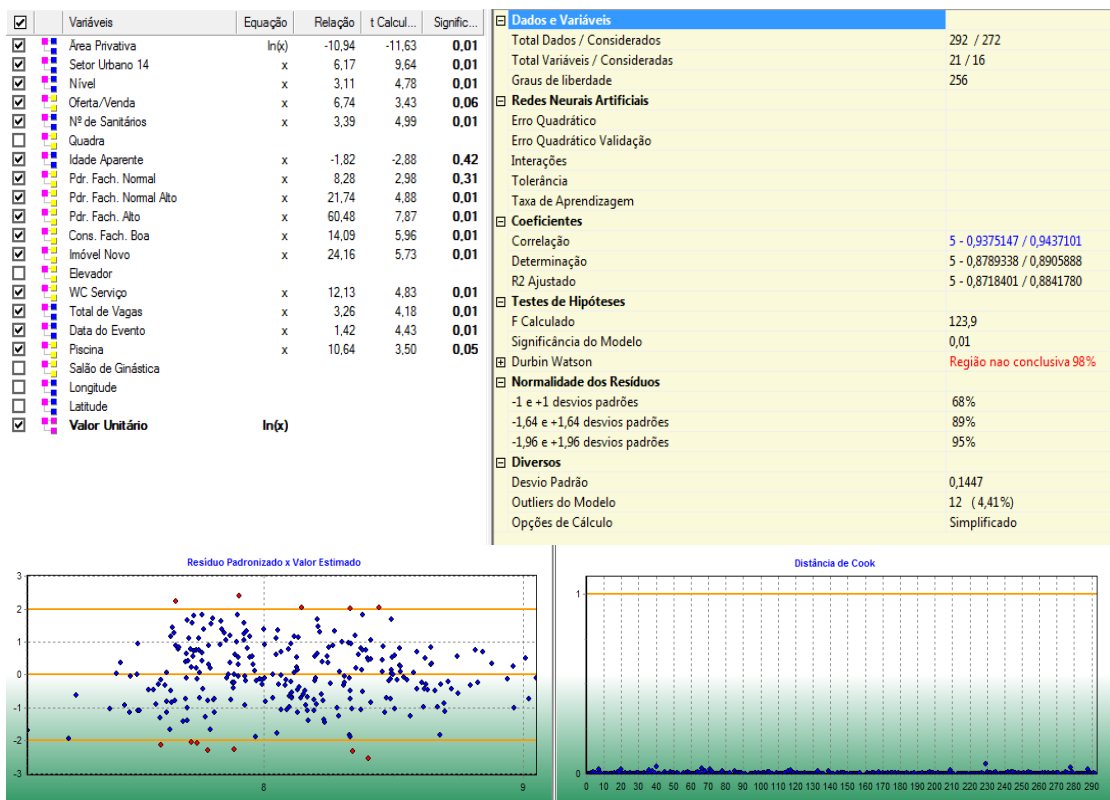
4.1.2.2 Tratamento de Dados

A) Após desabilitar as 3 (três) variáveis que apresentaram elevada significância dos regressores, foi encontrada uma modelagem coerente, com atendimentos a todos os parâmetros estatísticos, no entanto, com 5,82% dos dados sendo outliers, se fez necessário novos ajustes no modelo. Verifica-se ainda a presença da auto-regressão positiva.

<input checked="" type="checkbox"/>	Variáveis	Equação	Relação	t Calcul...	Signific...
<input checked="" type="checkbox"/>	Área Privativa	ln(x)	-12,09	-10,93	0,01
<input checked="" type="checkbox"/>	Setor Urbano 14	x	6,86	8,97	0,01
<input checked="" type="checkbox"/>	Nível	x	3,25	4,01	0,01
<input checked="" type="checkbox"/>	Oferta/Venda	x	5,79	2,35	1,91
<input checked="" type="checkbox"/>	Nº de Sanitários	x	3,60	4,21	0,01
<input type="checkbox"/>	Quadra				
<input checked="" type="checkbox"/>	Idade Aparente	x	-1,11	-1,39	16,32
<input checked="" type="checkbox"/>	Pdr. Fach. Normal	x	10,82	3,04	0,25
<input checked="" type="checkbox"/>	Pdr. Fach. Normal Alto	x	21,38	3,75	0,02
<input checked="" type="checkbox"/>	Pdr. Fach. Alto	x	67,76	7,01	0,01
<input checked="" type="checkbox"/>	Cons. Fach. Boa	x	13,96	4,69	0,01
<input checked="" type="checkbox"/>	Imóvel Novo	x	16,02	3,16	0,17
<input type="checkbox"/>	Elevador				
<input checked="" type="checkbox"/>	WC Serviço	x	10,93	3,54	0,04
<input checked="" type="checkbox"/>	Total de Vagas	x	4,17	4,26	0,01
<input checked="" type="checkbox"/>	Data do Evento	x	1,30	3,24	0,13
<input checked="" type="checkbox"/>	Piscina	x	15,34	3,94	0,01
<input type="checkbox"/>	Salão de Ginástica				
<input type="checkbox"/>	Longitude				
<input type="checkbox"/>	Latitude				
<input checked="" type="checkbox"/>	Valor Unitário	ln(x)			

Dados e Variáveis		
Total Dados / Considerados		292 / 292
Total Variáveis / Consideradas		21 / 16
Graus de liberdade		276
Redes Neurais Artificiais		
Erro Quadrático		
Erro Quadrático Validação		
Interações		
Tolerância		
Taxa de Aprendizagem		
Coefficientes		
Correlação	1 - 0,9143321 / 0,9100136	
Determinação	1 - 0,8360033 / 0,8281248	
R2 Ajustado	1 - 0,8270904 / 0,8187838	
Testes de Hipóteses		
F Calculado	93,8	
Significância do Modelo	0,01	
Durbin Watson		
d Calculado	1,42	
Durbin Watson	Data do Evento	
Normalidade dos Resíduos		
-1 e +1 desvios padrões	75%	
-1,64 e +1,64 desvios padrões	90%	
-1,96 e +1,96 desvios padrões	93%	
Diversos		
Desvio Padrão	0,18812	
Outliers do Modelo	17 (5,82%)	
Opções de Cálculo	Simplificado	

B) Após a retirada dos pontos influenciadores/ outliers que apresentaram resíduo padronizado superior a 2,5, encontra-se um modelo homocedástico, coerente, aderente, com resíduos apresentando dispersão aleatória, Normalidade dos resíduos, alto coeficiente de determinação e correlação, no entanto, ainda com a presença do teste de Durbin Watson acusando "Região não conclusiva".



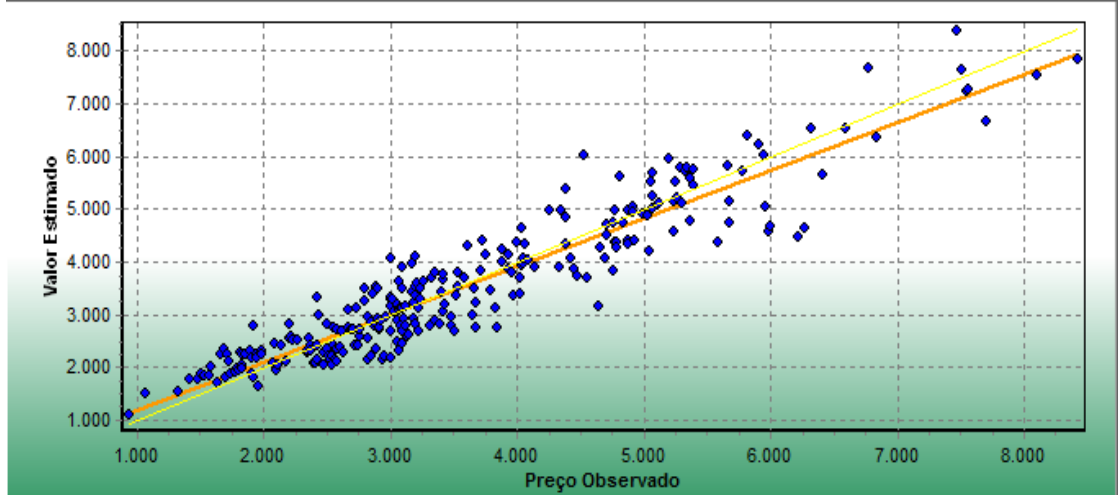
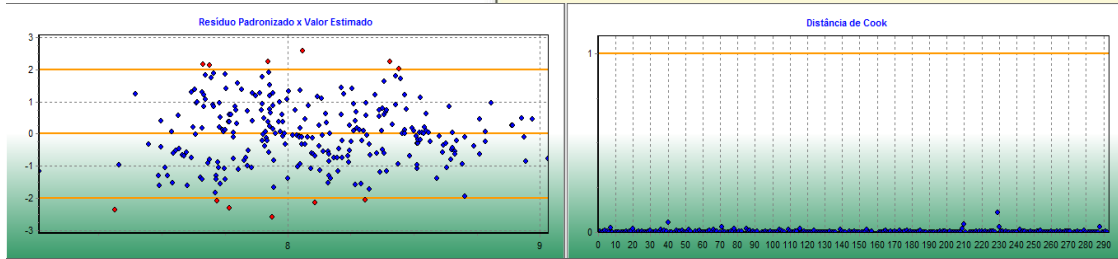
Após análise da pesquisa realizada e ajustes necessários, se verificou que mesmo o modelo tendo atendido a todos os parâmetros estatísticos, bem como acatado aos requisitos do Anexo A, da NBR 14.653-2, seria descartado em razão do teste de Durbin Watson ter apresentado resultado inconclusivo, muito embora não exista nenhuma menção a este teste na referida Norma. Outra opção para solucionar o resultado apresentado seria mais ajustes no modelo, com a retirada de mais dados e/ou mais variáveis.

4.1.3 Modelo Padrão Ajustado

Após aplicação dos ajustes necessários para que o resultado do teste de Durbin Watson com resultado de "região não conclusiva" fosse alterado, chegou ao Modelo padrão Ajustado. O tratamento de dados foi realizado através da inferência estatística, com calculo dos estimadores através do MQO, com o descarte de 22 (vinte e dois) dados, além de serem desabilitadas as variáveis Quadra, Elevador, e Salão de Ginástica por apresentarem significância acima do permitido normativamente, o Modelo Padrão apresentou resultados estatísticos adequados, podendo afirmar com 99% de Confiança que pode-se rejeitar a hipótese do modelo nulo, com também que 87% da variabilidade dos preços pode ser explicada pelo modelo. Os resíduos estão distribuídos de forma aleatória, sem nenhuma tendência e a distribuição de frequência próxima à Normalidade (69%, 91%, 95%), além do teste de Durbin Watson que apresentou o resultado de "não auto-regressão 95%", atendendo assim a todos os pressupostos Normativos e tradições utilizadas na Engenharia de Avaliações.

☑	Variáveis	Equação	Relação	t Calcul...	Signific...		
☑	Área Privativa	x	-9,10	-10,72			
☑	Setor Urbano 14	ln(x)	4,21	9,04	0,01		
☑	Nível	x	3,42	5,09	0,01		
☑	Oferta/Venda	x	7,03	3,49	0,05		
☑	Nº de Sanitários	x	3,96	5,50	0,01		
☐	Quadra						
☑	Idade Aparente	x	-1,72	-2,64	0,87		
☑	Pdr. Fach. Normal	x	5,26	1,83	6,77		
☑	Pdr. Fach. Normal Alto	x	18,13	3,98	0,01		
☑	Pdr. Fach. Alto	x	59,87	7,53	0,01		
☑	Cons. Fach. Boa	x	13,91	5,74	0,01		
☑	Imóvel Novo	x	26,29	5,93	0,01		
☑	Elevador						
☑	WC Serviço	x	8,58	3,47	0,06		
☑	Total de Vagas	x	2,44	4,06	0,01		
☑	Data do Evento	x	1,75	5,23	0,01		
☑	Piscina	x	9,13	2,93	0,36		
☐	Sala de Ginástica						
☐	Longitude						
☐	Latitude						
☑	Valor Unitário	ln(x)					

Dados e Variáveis		
Classificar: Significância	Considerados	292 / 270
Total Variáveis / Consideradas		21 / 16
Graus de liberdade		254
Redes Neurais Artificiais		
Erro Quadrático		
Erro Quadrático Validação		
Interações		
Tolerância		
Taxa de Aprendizagem		
Coefficientes		
Correlação	8 - 0,9334694 / 0,9395204	
Determinação	8 - 0,8713651 / 0,8826985	
R2 Ajustado	8 - 0,8637686 / 0,8757713	
Testes de Hipóteses		
F Calculado		114,7
Significância do Modelo		0,01
Durbin Watson		Não auto-regressão 95%
Normalidade dos Resíduos		
-1 e +1 desvios padrões		69%
-1,64 e +1,64 desvios padrões		91%
-1,96 e +1,96 desvios padrões		95%
Diversos		
Desvio Padrão		0,14808
Outliers do Modelo		12 (4,44%)
Opções de Cálculo		Simplificado



Equação de Regressão:

$$\text{Valor Unitário} = e^{(+7,359257153 - 0,00461699; [7833 * \text{Área Privativa} + 0,1462413374 * \ln(\text{Setor Urbano 14}) + 0,01121223439 * \text{Nível} + 0,0679544503 * \text{Oferta/Venda} + 0,1295313868 * \text{Nº de Sanitários} - 0,003626781206 * \text{Idade Aparente} + 0,05132215574 * \text{Pdr. Fach. Normal} + 0,1666879669 * \text{Pdr. Fach. Normal Alto} + 0,469244907 * \text{Pdr. Fach. Alto} + 0,1302447509 * \text{Cons. Fach. Boa} + 0,2334592517 * \text{Imóvel Novo} + 0,08234672161 * \text{WC Serviço} + 0,0805017907 * \text{Total de Vagas} + 0,005793220414 * \text{Data do Evento} + 0,0874263679 * \text{Piscina})}$$

Valor unitário (R\$/m²): 1.570,67 x (0,9954)Área Privativa x (Setor urbano)0,15 x (1,0112) Nível x (1,070) oferta/ venda x (1,138)Sanitário x (0,996) Idade Aparente x (1,05) Fachada Normal x (1,18) Fachada Normal Alto x (1,60) Alto x (1,139)Conservação Boa x (1,26) Imóvel Novo x (1,08) Wc de Serviço x (1,08) Quantidade de Vagas x (1,0058) Data do Evento x (1,091) Piscina

Portanto, como pode ser observado no Modelo Padrão Ajustado, todos os parâmetros foram atendidos, inclusive o teste de Durbin Watson que não verificou mais a presença de auto-regressão (autocorrelação). Ainda no que se refere ao Modelo Padrão Ajustado, observa-se que não existe uma busca pelos parâmetros mais perfeitos, com os maiores coeficientes de determinação e correlação, afinal este não é o objetivo deste trabalho, mas sim preservar o máximo de dados e variáveis possíveis no modelo. Eduardo Almeida (2012) afirma que se o pesquisador estimar a sua regressão e obtiver resultados que atendam a todos estes pressupostos, ele pode ser considerado uma pessoa de muita sorte. Na prática, surgem com frequência, os chamados problemas econométricos, que são as violações ou desvios dos pressupostos do Modelo de Regressão linear clássico. E o autor conclui afirmando que a Econometria moderna segue desenvolvendo novos testes e soluções para lidar com essas violações.

5. RESULTADOS

5.1 Modelo com a Presença de Auto-Regressão Positiva (Autocorrelação)

De acordo com o cenário apresentado, verifica-se que as soluções encontradas atualmente dentro da Engenharia de avaliações são limitadas, e quando o Engenheiro Avaliador se depara com um modelo estatisticamente correto e dentro dos parâmetros Normativos, mas que acusa um teste de Durbin-Watson com resultados positivos para autocorrelação, como "Auto-Regressão positiva" ou inconclusivas, "Região não-conclusiva", se faz necessário novos ajustes, com o sacrifício de mais dados e/ou variáveis, mesmo esse teste não sendo citado na NBR 14.653-2, quando não se vê obrigado a abandonar todo o trabalho já realizado e descartar aquela modelagem.

5.2 Modelo com a Presença de Autocorrelação Espacial

5.2.1 Testes para Detecção da Autocorrelação

Tomando como base as argumentações supracitadas, se buscou alternativas ao uso do teste de Durbin Watson, bem como outra solução à aquela do descarte de mais dados e/ou variáveis quando do seu resultado positivo ou inconclusivo. Segundo Eduardo Almeida (2012):

A evolução da econometria está intimamente vinculada ao desenvolvimento de diversos testes de diagnósticos para detectar as violações aos pressupostos, bem como métodos de estimação alternativos ao MQO para lidar com os problemas advindos dessas violações, tentando contorná-los.

Não é o objetivo deste trabalho adentrar nas demonstrações matemáticas dos testes existentes para a verificação de autocorrelação espacial, de forma que as opções aqui elencadas serão tratadas de forma superficial, apenas com o propósito de demonstrar que existem testes mais modernos e que se adéquam mais a estrutura de dados dos modelos de regressão utilizados na Engenharia de Avaliações.

Rodrigo de Souza Vieira (2009) afirma que segundo Anselin & Florax (1995), uma das formas mais amplamente utilizadas de se medir a autocorrelação espacial é o Índice global de Moran (I). Eduardo Almeida (2012) informa que segundo Anselin (2011), trata-se de um teste simples sobre autocorrelação da análise dos resíduos da regressão e que sua vantagem é a simplicidade computacional, vez que apenas os resíduos da regressão estimada por MQO são necessários para calculá-lo. Essa estatística varia entre -1 e 1, fornecendo uma medida geral da associação linear (espacial) entre os vetores Z_t no tempo t e a média ponderada dos valores da vizinhança, ou *lags* espaciais (WZ_t). Valores próximos de zero indicam inexistência de autocorrelação espacial significativa, ao tempo que quanto mais próximo de 1, mais autocorrelacionado está. A estatística I de Moran tem sido utilizada para verificação da autocorrelação espacial residual, assim como a estatística d de Durbin-Watson para verificação de autocorrelação para séries de tempo. Assim:

Neste caso, o teste I de Moran é aplicado sobre as estimativas dos erros de uma regressão feita por MQO, com a estatística I observada, comparada com uma distribuição aleatória aproximada por seus momentos, sob a hipótese nula de nenhuma correlação residual. “Tiefelsdorf & Boots (1995) fornecem os momentos exatos”. (VIEIRA, 2009).

Eduardo Almeida (2012) sugere que a estimação do modelo econométrico espacial pelo MQO não é uma boa ideia, e informa que os estimadores mais adotados em aplicações são os baseados em máxima verossimilhança e os baseados no princípio do método generalizado de momento.

Além da estatística I de Moran, pode-se utilizar também o teste de Wald, bem como os testes de Razão de Verossimilhança (RV), que assim como o teste de Wald, pode ser utilizado para verificar a autocorrelação espacial tanto na forma de defasagem quanto de erro. Existe ainda uma família de testes baseadas no Multiplicador de Lagrange (ML), que utilizado para detectar a defasagem espacial da variável dependente. Pode-se citar ainda modelos mais gerais de dependência espacial que é o Modelo SAR e o modelo SEM que a escolha dentre eles depende do resultado do Multiplicador de Lagrange (ML).Rodrigo de Souza Vieira (2009):

Apesar de o caminho percorrido pela econometria espacial, em termos de testes de especificação, ter sido muito parecido com o caminho da econometria tradicional, a implementação desses testes se mostrou bastante distinta entre os dois campos de pesquisa. Anselin & Bera (1998) chamam atenção para o fato de que os testes para os modelos de econometria espacial não seguem a forma padrão da maioria dos testes de econometria tradicional, na forma “ NR^2 ” – em que N é o tamanho da amostra, e R^2 é o coeficiente de determinação. Além disso, a possibilidade de defasagem espacial tanto na variável dependente quanto no termo de erro tornam os testes dos modelos espaciais mais complexos.

6. CONCLUSÕES

Remetendo-se a tudo que foi dito, fica bastante evidente que a estrutura de dados utilizada nos macromodelos de regressão dentro da Engenharia de

avaliações, com o uso da variável “data do evento”, se trata de agrupamento de cortes transversais e não de séries temporais, inegavelmente em razão dos dados serem aleatórios, a ordem das observações serem irrelevante e existirem lacunas de tempo. Verificou-se também que os dados são distribuídos geograficamente, podendo estar localizados em bairros, microregiões, municípios ou macroregiões, a depender da abrangência da pesquisa. Em síntese, nos modelos econométricos utilizados em Avaliações de Imóveis não há o que se falar em autocorrelação temporal e sim, quando houver, em autocorrelação espacial. Por esta razão, embora admita-se opiniões contrárias, entende-se que o teste de Durbin-Watson não é o teste mais adequado para a análise da autocorrelação dentro desses modelos, pois este teste, embora validado e amplamente utilizado no mundo estatístico, se aplica, apenas, para verificação de autocorrelação temporal. Na análise prática do teste, verificou-se, contudo, que o Teste de DW acusa alguns dados influenciadores ou com alto resíduo relativo, porém, mesmo quando da exclusão desses dados, o teste continuou apresentando resultado inconclusivo. Em suma, para conseguir um teste negativo de DW seria necessária a exclusão de mais dados e/ou variáveis explicativas. Ademais, conforme amplamente abordado, se basear no resultado deste teste como justificativa para eliminar um modelo estatístico que atenda a todos os outros pressupostos básicos de uma regressão e que seja representativo do mercado imobiliário, por este apresentar resultado positivo para autocorrelação ou região inconclusiva, é assumir o risco de inferências equivocadas e sem razoabilidade.

Este trabalho apresentou ainda as diferenças existentes entre a autocorrelação temporal e a autocorrelação espacial, sendo esta última a verificada nos modelos de regressão utilizados na Engenharia de Avaliações, e todos os prejuízos que a sua presença pode trazer a um modelo de regressão e por fim, foram indicados testes mais adequados para a verificação de sua presença que não o teste de Durbin-Watson.

BIBLIOGRAFIA

ABNT, ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. Avaliação de bens. Parte 1: Procedimentos Gerais: NBR 14653-1.

ABNT, ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. Avaliação de bens. Parte 2: Imóveis Urbanos: NBR 14653-2.

ABUNAHMAN, Sérgio Antonio. **Curso básico de engenharia legal e de avaliações**. São Paulo, Edidora Pini, 2008

ALMEIDA, Eduardo, **Econometria Espacial Aplicada**. Campinas, SP, Editora Alínea, 2012.

ARANHA, Francisco. **Autocorrelação espacial na área de loja de supermercados nos municípios paulistas: mensuração por meio do Índice de Geary**. *Rev. adm. empres.* [online]. 1999, vol.39, n.4, pp.38-45.

CABRAL, Alexandra Maria Rios; LINS, Isis Didier. **Aplicação de Bootstrap para teste de Durbin-Watson** - Uma contribuição para cenários na economia. Economia política do desenvolvimento. Maceió, vol. 4, n. 12, p. 69-90, 2011

DANTAS, Rubens Alves. **Engenharia de Avaliações**: uma introdução à metodologia científica. São Paulo: Pini, 1998.

GREENE, William H., **Econometric Analysis**, Fifth Edition, Editora: Prentice Hall Press, 2003.

GUJARATI, Damodar N., **Econometria Básica**, Elsevier Editora Ltda, 2006

JOHNSTON, John, **Métodos Econométricos**, São Paulo, Editada Atlas S.A. 1986

MALBOUISSON, Cláudia; TIRYAKI, Gisele F. **Econometria na Prática**, Alta Books Editora, 2017.

MATTA, Túlio Alves. **Avaliação do valor de imóveis por análise de regressão**: estudo de caso para a cidade de Juiz de Fora. Juiz de Fora: Universidade Federal de Juiz de Fora, 2007.

PINDYCK, Robert S.; RUBINFELD, D.L. **Econometria**: Modelos & Previsões, 4 ed. Rio de Janeiro: Campus: 2004

RADEGAZ, Násser Júnior. **Avaliação de bens**: princípios básicos e aplicações. São Paulo: Liv. e Ed. Universitária de Direito, 2011

TEIXEIRA, Marcelo da Costa; TEIXEIRA, Vanessa Domingos Gonçalves. **Proposta de exclusão da variável data (ou tempo) em macromodelos estatísticos (inferência) com análise de estudo de Caso. Congresso Brasileiro de Engenharia de Avaliações e Perícias**. Foz do Iguaçu. 2017.

VIEIRA, Rodrigo de S. Crescimento econômico no estado de São Paulo: uma análise espacial [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009.

WOOLDRIDGE, Jeffrey M., **Introdução a Econometria, Uma abordagem Moderna**. Editora Cengage Learning, 4ª Edição, 2006