Crítica ao Campo de Arbítrio do Avaliador

Luiz Fernando Palin Droubi, Willian Zonato, Norberto Hochheim

2019-11-21

Introdução

Recentemente, durante o XX Cobreap, foi apresentado o artigo intitulado "Crítica à Avaliação Intervalar na NBR14.653-02".

Eventualmente, numa próxima revisão da parte 2 da NBR 14.653-02, recomendou-se que a avaliação intervalar, tal como está normatizada no momento, não deve ser mantida numa próxima revisão.

Visando aprofundar o assunto da avaliação intervalar, assim como acrescentar alguns pontos no que tange ao campo de arbítrio, pretende-se com este artigo deixar detalhes de uma proposta de como poder-se-ia modificar o atual consenso da avaliação intervalar numa futura revisão normativa.

Revisão Bibliográfica

Campo de Arbítrio

Para isto, começo com o Campo de Arbítrio. Infelizmente este tópico não foi discutido o suficiente no nosso artigo e do debate que houve após a apresentação do mesmo, onde estavam presentes alguns dos membros do comitê normatizador, surgiram novais idéias.

A NBR 14.653-02 define que "o Campo de Arbítrio pode ser utilizado quando variáveis relevantes para a avaliação do imóvel não tiverem sido contempladas no modelo, por escassez de dados de mercado, por inexistência de fatores de homogeinização aplicáveis ou porque essas variáveis não se apresentaram estatisticamente significantes em modelos de regressão, *desde que a amplitude de até mais ou menos 15% seja suficiente para absorver as influências não consideradas e que os ajustes sejam justificados*".

Um dos pontos importantes levantandos na nossa apresentação é a de que este intervalo fixo, de $\pm$ 15%, pode até ser razoável em algumas situações, mas pode levar a valores de baixíssima probabilidade de ocorrência prática em mercados de baixa variabilidade, ou ainda, num mercado de grande variância, levar a valores que sequer chegam próximos dos valores extremos observados no mercado.

Outro ponto por nós levantado é que é absurdo comparar este valor arbitrado ao intervalo de confiança da regressão linear, pois *o intervalo de confiança é para a média*, ou seja, o intervalo de confiança é apenas para aferir a precisão do cálculo dos valores ajustados pela reta de regressão (média + variância explicada) e não serve para contemplar outros efeitos não calculados pelo modelo para a avaliação do bem-avaliando.

Para isto, é necessário que a comparação seja com o intervalo de predição, ou seja, o intervalo onde se encontram os valores reais observados do mercado e não apenas os valores médios (a soma da média e da variância explicada e não-explicada).

Efeitos da “falta” de uma variável relevante no modelo

A não-inclusão de uma variável relevante no modelo de regressão linear causa o problema de viés devido à omissão dessa variável (*omitted variable bias*).

Ou seja, uma variável relevante não inclusa no modelo tem o efeito de viesar as estimativas dos coeficientes das outras variáveis presentes no modelo.

Estudo de Casos

Exemplo 1

Por exemplo, seja o caso de se avaliar um lote urbano de esquina com 480m^2, baseado na amostra ao lado, onde `situacao` é uma variável dicotômica que diferencia os imóveis de meio de quadra, em que esta variável assume o valor zero, e imóveis de esquina, em que a variável assume o valor 1.

id	VU	area	situacao
1	3104	360	0
2	3171	360	0
3	3226	360	0
4	3085	360	0
5	2870	480	1
6	2853	480	1
7	3209	360	0
8	3312	360	0
9	2478	480	0
10	2727	480	0

Os dados da amostra acima foram criados randomicamente através da seguinte expressão:

$$VU = 5000 - 5 \cdot area + 250 \cdot situacao + \epsilon$$

Onde $\epsilon \sim \mathcal{N}(0,\ 100^2)$

De acordo com a NBR 14.653-02, como foram encontrados apenas 2 dados de mercado em situação de esquina, não seria possível, de acordo com a NBR 14.653-02 a utilização da variável `situacao`, devido à micronumerosidade. Neste tipo de situação (escassez de dados de mercado), a norma permite a utilização do campo de arbítrio.

Ou seja, poderia ser elaborado um modelo com todos os dados amostrais, com a inclusão apenas da variável `area`, e utilizar o campo de arbítrio para majorar o valor estimado do lote pelo modelo, por este estar em situação de esquina, correto?

A resposta é negativa. Pela análise da tabela abaixo, onde ajustamos três modelos para os dados da amostra, isso fica claro: no primeiro modelo (coluna 1), foram utilizados todos os dados e apenas a variável `area`; no segundo

	Dependent variable:		
	VU		
	(1)	(2)	(3)
Constant	4.541,37	4.929,67	4.929,67
	(4.174,75, 4.907,99)	(4.569,82, 5.289,52)	(4.596,17, 5.263,17)
	t = 15,87	t = 17,56	t = 18,94
	p = 0,0000***	p = 0,0000***	p = 0,0000***
area	−3,77	−4,85	−4,85
	(−4,66, −2,88)	(−5,76, −3,93)	(−5,70, −4,00)
	t = −5,43	t = −6,79	t = −7,33
	p = 0,001***	p = 0,0005***	p = 0,0002***
situacao			258,87
			(134,29, 383,45)
			t = 2,66
			p = 0,04**
Observations	10	8	10
R^2	0,79	0,88	0,89
Adjusted R^2	0,76	0,87	0,86
Residual Std. Error	129,02 (df = 8)	104,89 (df = 6)	97,21 (df = 7)
F Statistic	29,51*** (df = 1; 8)	46,15*** (df = 1; 6)	29,53*** (df = 2; 7)
Note:			*p<0,1; **p<0,05; ***p<0,01

modelo (coluna 2), foram utilizados apenas os dados em situação de meio-de-quadra; e no terceiro modelo (coluna 3) foram utilizados todos os dados e foi incluída a variável `situacao`, apesar da micronumerosidade.

O que se percebe é que o primeiro modelo viesou para cima o coeficiente da variável `area`. Ou seja, o efeito da `situacao` do lote, que não foi incluso no modelo, foi "absorvido" pela outra variável restante, ou seja, a variável `area`. Os outros dois modelos se aproximaram melhor dos coeficientes "reais" de regressão, ou seja, o valor dos coeficientes para a população (-5,0 e 250).

Vamos ver o que ocorre então com as estimativas.

Utilizando-se o primeiro modelo, a estimava encontrada é de R$2.731,86. Utilizando-se o segundo modelo, a estimava encontrada é de R$2.602,43. Utilizando-se o terceiro modelo, a estimava encontrada é de R$2.861,30.

O valor utilizando-se os coeficientes de regressão para a população seria de R$2.850,00.

Percebe-se desta maneira que para o primeiro modelo o valor adicional do lote em esquina já foi absorvido e já se encontra, em grande parte, na estimativa efetuada com o mesmo.

No segundo modelo, que é um modelo feito apenas para os dados de lotes em situação de meio de quadra, não há qualquer influência de lotes em situação de esquina, ficando a estimação comprometida.

O terceiro modelo, apesar da micronumerosidade, foi o que melhor estimou o lote em situação de esquina.

Os intervalos de predição @80% para os três modelos, bem como os limites do campo de arbítrio podem ser vistos na tabela abaixo:

Modelo	Estimativa central	IP_{inf}	IP_{sup}	Amplitude	CA_{inf}	CA_{sup}
1	2.731,86	2.530,38	2.933,34	14.75	2.322,08	3.141,64
2	2.602,43	2.417,47	2.787,38	14.21	2.212,06	2.992,79
3	2.861,30	2.692,84	3.029,75	11.77	2.432,10	3.290,49

Exemplo 2

Neste segundo exemplo, foram seguidos exatamente os mesmos passos do exemplo anterior. Contudo, para a geração dos valores, foi utilizado um peso maior para os lotes de esquinam, conforme a expressão abaixo:

$$VU = 5000 - 5 \cdot area + 390 \cdot situacao + \epsilon$$

Onde $\epsilon \sim \mathcal{N}(0,\, 100^2)$

Deve-se reparar que, aqui, erros à parte,

id	VU	area	situacao
1	3104	360	0
2	3171	360	0
3	3226	360	0
4	3085	360	0
5	3010	480	1
6	2993	480	1
7	3209	360	0
8	3312	360	0
9	2478	480	0
10	2727	480	0

foi adicionado R$390,00 ao valor dos lotes de esquina, o que significa um adicional de exatamente 15% ao valor do lote normal (centro de quadra).

Novamente, então, foram ajustados três modelos, como os descritos no exemplo anterior.

Utilizando-se o primeiro modelo, a estimava encontrada é de R$2.801,86. Utilizando-se o segundo modelo, a estimava encontrada é de R$2.602,43. Utilizando-se o terceiro modelo, a estimava encontrada é de R$3.001,30.

O valor utilizando-se os coeficientes de regressão para a população seria de R$2.990,00.

	Dependent variable:		
	VU		
	(1)	(2)	(3)
Constant	4.331,37	4.929,67	4.929,67
	(3.854,55, 4.808,19)	(4.569,82, 5.289,52)	(4.596,17, 5.263,17)
	t = 11,64	t = 17,56	t = 18,94
	p = 0,0000***	p = 0,0000***	p = 0,0000***
area	−3,19	−4,85	−4,85
	(−4,34, −2,03)	(−5,76, −3,93)	(−5,70, −4,00)
	t = −3,53	t = −6,79	t = −7,33
	p = 0,01***	p = 0,0005***	p = 0,0002***
situacao			398,87
			(274,29, 523,45)
			t = 4,10
			p = 0,005***
Observations	10	8	10
R^2	0,61	0,88	0,89
Adjusted R^2	0,56	0,87	0,85
Residual Std. Error	167,80 (df = 8)	104,89 (df = 6)	97,21 (df = 7)
F Statistic	12,46*** (df = 1; 8)	46,15*** (df = 1; 6)	26,99*** (df = 2; 7)
Note:			*p<0,1; **p<0,05; ***p<0,01

Os intervalos de predição @80% para os três modelos, bem como os limites do campo de arbítrio podem ser vistos na tabela abaixo:

Modelo	Estimativa central	IP_{inf}	IP_{sup}	Amplitude	CA_{inf}	CA_{sup}
1	2.801,86	2.539,82	3.063,91	18.71	2.381,58	3.222,14
2	2.602,43	2.417,47	2.787,38	14.21	2.212,06	2.992,79

Modelo	Estimativa central	IP_{inf}	IP_{sup}	Amplitude	CA_{inf}	CA_{sup}
3	3.001,30	2.832,84	3.169,75	11.23	2.551,10	3.451,49

Exemplo 3

Neste segundo exemplo, foram seguidos exatamente os mesmos passos do exemplo anterior. Contudo, para a geração dos valores, foi utilizado um peso maior para os lotes de esquinam, conforme a expressão abaixo:

id	VU	area	situacao
1	3104	360	0
2	3171	360	0
3	3226	360	0
4	3085	360	0
5	3120	480	1
6	3103	480	1
7	3209	360	0
8	3312	360	0
9	2478	480	0
10	2727	480	0

$$VU = 5000 - 5 \cdot area + 500 \cdot situacao + \epsilon$$

Onde $\epsilon \sim \mathcal{N}(0,\ 100^2)$

Deve-se reparar que, aqui, erros à parte, foi adicionado R$500,00 ao valor dos lotes de esquina, o que significa um adicional maior do que 15% ao valor do lote normal (centro de quadra).

Novamente, então, foram ajustados três modelos, como os descritos no exemplo anterior.

Utilizando-se o primeiro modelo, a estimava encontrada é de R$2.856,86. Utilizando-se o segundo modelo, a estimava encontrada é de R$2.602,43. Utilizando-se o terceiro modelo, a estimava encontrada é de R$3.111,30.

O valor utilizando-se os coeficientes de regressão para a população seria de R$3.100,00.

Os intervalos de predição @80% para os três modelos, bem como os limites do campo de arbítrio podem ser vistos na tabela abaixo:

Modelo	Estimativa central	IP_{inf}	IP_{sup}	Amplitude	CA_{inf}	CA_{sup}
1	2.856,86	2.542,05	3.171,67	22.04	2.428,33	3.285,39
2	2.602,43	2.417,47	2.787,38	14.21	2.212,06	2.992,79
3	3.111,30	2.942,84	3.279,75	10.83	2.644,60	3.577,99

Resumo

Em suma, para três mercados diferentes (3 bairros diferentes, digamos), foram testados três tipos de abordagens diferentes:

1. A abordagem de se considerar todos os dados, porém excluindo a variável `situacao`, por conta da micronumerosidade;

2. A abordagem de se desconsiderar, além das variáveis, também os dados de

	Dependent variable:		
	VU		
	(1)	(2)	(3)
Constant	4.166,37 (3.593,53, 4.739,20) t = 9,32 p = 0,0001***	4.929,67 (4.569,82, 5.289,52) t = 17,56 p = 0,0000***	4.929,67 (4.596,17, 5.263,17) t = 18,94 p = 0,0000***
area	−2,73 (−4,12, −1,34) t = −2,52 p = 0,04**	−4,85 (−5,76, −3,93) t = −6,79 p = 0,0005***	−4,85 (−5,70, −4,00) t = −7,33 p = 0,0002***
situacao			508,87 (384,29, 633,45) t = 5,23 p = 0,002***
Observations	10	8	10
R^2	0,44	0,88	0,89
Adjusted R^2	0,37	0,87	0,85
Residual Std. Error	201,59 (df = 8)	104,89 (df = 6)	97,21 (df = 7)
F Statistic	6,33** (df = 1; 8)	46,15*** (df = 1; 6)	27,31*** (df = 2; 7)
Note:			*p<0,1; **p<0,05; ***p<0,01

esquina, chegando-se assim a um modelo apenas para os lotes em situação de meio de quadra;

3. A abordagem de utilizar-se todos os dados e todas as variáveis, apesar de ir de encontro com a norma vigente, por conta da micronumerosidade.

Em todos os mercados (bairros) a terceira abordagem se mostrou superior, obtendo-se uma valor de estimativa central mais próxima do "real", e um intervalo de predição mais estreito.

Mercado	Modelo	Valor "Real"	Estimativa central	IP_{sup}	CA_{sup}
	1		2.731,86	2.933,34	3.141,64
A	2	R$2.850,00	2.602,43	2.787,38	2.992,79
	3		2.861,30	3.029,75	3.290,49
	1		2.801,86	3.063,91	3.222,14
B	2	R$2.990,00	2.602,43	2.787,38	2.992,79
	3		3.001,30	3.169,75	3.451,49
	1		2.856,86	3.171,67	3.285,39
C	2	R$3.100,00	2.602,43	2.787,38	2.992,79
	3		3.111,30	3.279,75	3.577,99

Com a primeira abordagem, obtem-se um modelo razoavelmente bom para fazer previsões, porém com coeficientes "errados", por conta do efeito da variável omitida.

Finalmente, com a segunda abordagem, obtem-se sempre o mesmo modelo, com mesmos intervalos de predição e mesmos campos de arbitrio, haja vista que o que se altera nos três mercados é apenas o efeito de majoração devido à situação de esquina. Apesar deste modelo obter coeficientes "corretos", ele não é bom para se efetuar previsões. E o campo de arbítrio só se mostra eficaz no segundo exemplo, onde forçamos que o valor do lote de esquina tenha majoração de exatos 15% em relação ao lote de meio de quadra.

A adoção da primeira abordagem é preferível à adoção da segunda, haja vista que ela leva a estimativas centrais sempre aquém dos valores "reais", contudo, menos aquém do que com os valores obtidos com a segunda abordagem. Além do mais, com a adoção do limite superior do intervalo de predição, o erro cometido com a primeira abordagem é muito pequeno, da ordem de 2,5 a 3,0%.

Mercado	Erro $\hat{Y}$	Erro $\hat{Y}_{sup}$	Erro CA_{sup}
A	-4.15%	2.92%	10.2%
B	-6.29%	2.47%	7.76%
C	-7.84%	2.31%	5.98%

Mercado	Erro $\hat{Y}$	Erro $\hat{Y}_{sup}$	Erro CA_{sup}
A	-8.69%	-2.2%	5.01%
B	-13%	-6.78%	0.0933%
C	-16.1%	-10.1%	-3.46%

Conclusão

Primeiramente, deve-se concluir que, pelo menos para este caso, melhor seria utilizar o modelo com micronumerosidade (terceiro modelo). o assunto da micronumerosidade, porém, fica para outro *post*. A estimativa feita com tal modelo é a mais próxima da estimativa feita com os coeficientes de regressão “reais”.

A opção de se retirar os dados em situação de esquina e ajustar um modelo apenas para os dados de meio de quadra (segundo modelo) não parece uma boa solução: no final o que se tem é um modelo apenas para lotes em meio de quadra, sem qualquer parâmetro para se inferir o valor dos lotes de esquina. Além do mais, dados estão sendo jogados fora (20% deles).

A opção de se utilizar o primeiro modelo, sem a variável `situacao` deve ser feita com cuidado: o valor do coeficiente da variável presente (`area`) deve ser visto com ressalvas, pois nele estão embutidos os efeitos da situação dos lotes, ainda que a variável esteja ausente. A estimativa calculada com este modelo encontra-se sempre um pouco abaixo da “real”. A diferença entre a estimativa central obtida com este modelo do valor “real” do lote de esquina aumenta com o aumento do efeito de majoração da variável `situacao`. É mais ou menos como se o modelo de regressão estivesse “errado” (todos os modelos estão errados), mas ainda fosse útil (as previsões com ele obtidas são razoáveis). O modelo, porém, avalia lotes em meio de quadra e lotes de esquina com os mesmos valores. Isto significa que as estimativas estarão apenas um pouco maiores para os lotes de meio de quadra e um pouco menores para os lotes de esquina. Mas quão maiores/menores? O intervalo de predição, neste caso, é um bom parâmetro.

Resumindo: no terceiro modelo não seria necessário utilizar o campo de arbítrio, adotando-se a estimativa de tendência central; no segundo modelo, poderíamos utilizar o campo de arbítrio para majorar o efeito da variável `situacao`, porém deve-se salientar que, aqui, estaríamos atirando no escuro, pois não há qualquer base teórica que justifique o coeficiente de majoração adotado; finalmente, no primeiro modelo, a adoção do campo de arbítrio estaria incrementando um efeito que, ao menos em parte, já está embutido no modelo. Desta forma, a utilização de todo o campo de arbítrio para majorar o valor estimado, levaria o valor do lote a valores extremos; seria mais indicado a utilização do intervalo de predição, onde sabe-se que se encontram os valores “razoáveis” para o mercado em análise.

Inexistência de fatores de homogeneização

Ficou claro na discussão que a grande maioria dos avaliadores se utiliza do campo de arbítrio para resolver um dos maiores problemas da prática atual da Engenharia de Avaliações, que é a falta da determinação precisa de um fator oferta. A imensa maioria dos Avaliadores se valem de uma pesquisa de dados ofertados, chegando assim a um modelo para os preços de oferta, que infelizmente não corresponde a um modelo de preços de venda dos imóveis.

A atual normativa não permite a aplicação de um fator oferta aos dados pesquisados, a menos que este fator seja deduzido com a utilização de metodologia científica.

Pretendo mostrar que se trata de um problema diferente do de estabelecer o valor de um imóvel com variáveis relevantes ausentes do modelo.

Primeiramente vou fazer uma consideração a respeito do cálculo do fator oferta.

Para fazer um cálculo correto do fator oferta seria necessário ter acesso aos dados de oferta e dados de venda dos **mesmos imóveis**. Um artigo onde isto fica claro pode ser encontrado neste link. Na verdade, neste artigo se mostra que os preços de venda dos imóveis podem ser melhor previstas pelo modelo econométrico proposto, que prevê os preços de venda em função do preço anunciado, do que por um simples modelo hedônico de preços de venda.

Usualmente, no entanto, se dispomos de dados de oferta e dados de venda, o que se faz é acrescentar uma variável dicotômica com a informação se o dado se trata de um dado de oferta ou se se trata de um dado de venda. Não há valores de venda e de oferta para os mesmos dados.

Assim, caso resulte significante, o valor do coeficiente calculado para essa variável dicotômica é considerado para o cálculo do fator oferta.

Isto é questionável. Por exemplo, imagine-se dois apartamentos semelhantes, ou seja, com as mesmas características, pelas variávies do modelo. O primeiro anunciado por R$1.000.000,00 e outro por R$945.000,00. O último não foi ainda negociado, e o primeiro apartamento foi negociado por R$950.000,00. Chegaría-se à conclusão que o fator oferta foi 1,005. Como na Engenharia de Avaliações despreza-se as variáveis que não se comportam como o esperado, mesmo que significante, esta variável seria descartada.

Por outro lado, imagine-se que o segundo apartamento tenha sido negociado por R$900.000,00. Chegaría-se à conclusão que o fator oferta foi de 0,9, enquanto o fator real de oferta foi de 0,95. Provavelmente o apartamento anunciado por R$1.000.000,00 possui características não contempladas no modelo que elevam o seu valor, e o proprietário não aceitaria os R$900.000,00 ofertados no segundo apartamento.

E ao final o que temos é um modelo que não é nem de dados de oferta, nem de dados de venda.

Pode-se argumentar que se trata apenas de uma aproximação e que para o cálculo do fator oferta não são utilizados apenas dois dados, mas no mínimo três dados para cada situação, o que muda um pouco as coisas, mas em essência, tem-se o mesmo problema.

Mas onde eu quero mesmo chegar é que a mudança de um modelo de ofertas para um modelo de preços de venda é uma mudança que ocorre na forma de um deslocamento de todo o modelo em uma mesma direção. Diferentemente do que ocorre ao suprimir uma variável hedônica do modelo, onde ocorre o deslocamento de um ponto dentro de uma distribuição de probabilidade, a falta de uma homogeneização dos dados de oferta desloca todo o modelo para cima.

Propostas de mudanças na normatização

O que seria razoável, então, em termos de normatização?

Ter um grau de fundamentação maior para os modelos onde o fator oferta foi calculado com precisão. Permitir, num grau de fundamentação menor, que se aplique um fator oferta de acordo com a experiência do avaliador.

Além disto, definir um campo de arbítrio de maneira diversa da que está definida atualmente na norma, negando a possibilidade de uso do campo de arbítrio para a falta de fator oferta e permitindo ao avaliador a utilização do campo de arbítrio para que o avaliador avalie o imóvel-avaliando dentro do intervalo de predição do modelo, que é mais adequado do que o campo de arbítrio fixo de mais ou menos 15%, como já demonstramos.

Tudo isto, claro, devidamente justificado.

Sobre a avaliação intervalar

Resta ainda abordar o problema da avaliação intervalar.

Escolhido um valor diferente da média, como avaliar a precisão desta estimativa?

Atualmente a NBR 14.653-02 faz uma verdadeira salada com Campo de Arbítrio e intervalos de confiança e predição para se chegar a um intervalo de valores admissíveis.

Na minha opinião o mais razoável seria adotar a amplitude do próprio intervalo de confiança.

Assim, se o valor médio ajusta $\hat{y}_0$ de um imóvel com características x_0 for de R$1.000.000,00, com intervalo de confiança entre R$900.000,00 e R$1.100.000,00, e intervalo de predição entre R$750.000,00 e R$1.250.000,00, seria possível ao avaliador escolher para o valor do imóvel qualquer valor dentro do IP, com a amplitude do intervalo de valores admssíveis iguais ao do IC.

Por exemplo, caso o avaliador opte pelo valor mínimo do IP, R$750.000,00, o intervalo de valores admissíveis ficaria entre R$650.000,00 e R$850.000,00.

Isto é estatisticamente consistente. Na verdade o que se estaria fazendo nada mais é do que deslocar toda a distribuição de probabilidades do mercado (intervalo de predição), dentro do IC.

No exemplo acima, o valor central da distribuição dos dados do mercado pode ser considerado de R$1.000.000,00, mas também pode ser considerado de R$900.000,00, ou R$1.100.000,00. Deslocando a média distribuição para estes extremos, também estamos deslocando os limites inferior e superior do intervalo de predição.