

تحيز المتغير المحذوف (Omitted Variable Bias)

Foundations • The Valuation Engineer

William Bert Craytor • October 1, 2026

Version 1.2.0

DOI: 10.67330/460smg10

تعريف صوري. انحراف المتغير المستبعد هو التشويه المنهجي الذي يحدث في المعاملات المُقدَّرة عندما يتم استبعاد سمة تؤثر على السعر من النموذج ، وتتغير هذه السمة المستبعدة بالتوازي مع إحدى السمات المُدرجة --- أي تتغير معها بعد أن تم السماح ، وهو ما يقيسه الانحدار المساعد (وهو انحدار ثانٍ للميزة المستبعدة على الميزات المضمنة). ثم يتسرب تأثير الميزة المستبعدة إلى المعاملات الخاصة بالميزات المضمنة (الدولارات لكل وحدة التي يخصصها النموذج لكل ميزة).

تقريب حدسي. قد تكون القيمة الخاطئة ناتجة عن احتوائها على قيمة سمة أخرى. فقد بيعت قيمة H Comp في العدد 1 بمبلغ \$46,455 أكثر مما توقعه الانحدار بناءً على السمات الأربع المسجلة. والسبب، الذي تم الكشف عنه في المتغيرات للعدد 1، هو عملية تجديد حديثة للمطبخ، والتي تشير تلك المعلومة إلى أنها تضيف حوالي \$60,000 إلى قيمة المنزل. كانت «الحالة» مدرجة في قائمة السمات، لكن لم يُسجل أي شيء عن عملية التجديد، وأيًا كان ما أضافته عملية التجديد فقد كان مفقودًا من البيانات. لم تختف تلك القيمة. بل تم توزيعها عبر معدلات السمات التي تم تضمينها، لأن تلك السمات تصادف أنها ارتبطت بها في عمليات البيع. يجب أن يتحقق أمران لكي يحدث هذا: يجب أن تكون القيمة المفقودة ذات أهمية بالنسبة للسعر، ويجب أن ترتبط بميزة موجودة في القائمة. عندما تكون ذات أهمية ولكنها لا ترتبط بأي شيء، تكون المعدلات صحيحة ويكون التباين أوسع فقط. وينطبق هذا سواء حصلت الميزة على معدل واحد أو عدة معدلات: تمنح طريقة MARS (الخطوط المنحنية للانحدار التكميلي متعدد المتغيرات، فريدمان 1991)، وهي طريقة الانحدار الرئيسية التي تستخدمها المجلة، السمة معدلات مختلفة على نطاقات مختلفة من قيمها، وهناك تنتشر القيمة المستعارة عبر تلك النطاقات بدلاً من أن تقتصر على رقم واحد، لكنها موجودة، لنفس السبب.

للقارئ الذي يرغب في معرفة الآلية. في الانحدار العادي، يُعد «المعامل» معاملاً --- أي المبلغ بالدولار الذي يخصصه النموذج لكل وحدة إضافية من السمة --- وتكون القيمة المستعارة ذات حجم محدد بدقة: معامل السمة المستبعدة نفسه مضروباً في مدى قوة ارتباطها بالسمة المُدرجة، حيث يمثل العامل الثاني معامل الانحدار المساعد للسمة المستبعدة على السمات المُدرجة. يشرح المقال هذه الصيغة، التي تمثل الحالة الخطية، ثم يوضح نفس التسرب في الشبكة، حيث تنتقله المبيعات المزدوجة للمقيمين تماماً كما تفعل المعاملات. وقد كُتب هذا المقال للمقيمين الذين يحصلون على تعديلات من بيانات السوق.

ما يمثلته المثال الوارد في هذا العدد، وما لا يمثلته. إن سوق «باسيفيكا» الافتراضي المستخدم في المقالات من رقم 007 (التوقع الشرطي) إلى رقم 010 (تحديد المعلمات) يعامل الحالة كسمة ذات سعر بالدولار، كما يفعل التقييم التقليدي وكما فعل العدد الأول، ويوضح المثال ما الذي يحدث من خطأ في هذا النهج. منهجية التقييم التقليدية ليست هي المنهجية التي تتبعها المجلة، بل منهج القيد المتبقي (RCA) --- وهو منهج القيد المتبقي الوارد في العدد 1 Craytor (2026) --- الذي يستبعد الحالة والجودة من معادلة الانحدار؛ ويوضح المقال رقم 007 (التوقع الشرطي) كيفية ذلك. ينبغي على القارئ أن يستفيد من الدرس المستخلص من المثال، وليس من مواصفاته.

الحالة، ودرجة الحالة، وإعادة التقييم. هذه الأمور الثلاثة غالباً ما تُخلط بينها. الحالة هي حالة المنزل: مدى تآكل التشطيبات، وعمر الأنظمة، والصيانة المؤجلة. أما درجة الحالة فهي تقدير لتلك الحالة، مقارنةً بحالة المنازل المحيطة في منطقة السوق المحددة. وهي تستند إلى فهم المقيمين لتلك المنطقة السوقية، ولا ينبغي أن تتعارض بشكل جوهري مع فهم المقيمين الأقران العاملين في نفس المنطقة السوقية. في هذه المجلة، تتراوح الدرجة بين 00.0 و00.10: وهي تمثل النسبة المئوية للمنازل في المنطقة السوقية التي تكون في حالة أسوأ، مقسومة على عشرة. الدرجة 00.5 تعني أن نصف المنازل في منطقة السوق في حالة أسوأ؛ و00.9 تعني أن تسعة من كل عشرة منازل في حالة أسوأ. هذه الدرجة هي تقدير، ويمكن أن يختلف منزلان يحملان نفس الدرجة عن بعضهما. أما عملية التجديد فهي ليست حالة ولا درجة، بل حدث: أعمال تم إنجازها خلال فترة حديثة، وأدت إلى تغيير المنزل. وتتمثل قيمتها في قيمة ما أحدثته من تغيير --- الحالة، وعادةً الجودة، والمواد والتشطيبات --- وفقاً للأسعار التي يدفعها السوق مقابلها. وعادةً ما يكون تأثير عملية التجديد في الغالب على الحالة وأقل على الجودة، ويرجع ذلك إلى حد كبير إلى تكاليف العمالة، على الرغم من أن القيمة المضافة من خلال الجودة في بعض عمليات التجديد تكون هي الأعلى بين الاثنين. ينبغي أن تكون قيمة المنزل الذي خضع لعملية إعادة تصميم والمنزل الذي لم يخضع لها قط، وهما متشابهان في كل السمة الحالية التي تؤثر على القيمة، متساوية؛ فالسوق يدفع مقابل الحالة والجودة التي نتجت عن عملية إعادة التصميم، وليس مقابل حقيقة حدوث عملية إعادة التصميم بحد ذاتها. لذا، فإن عملية إعادة التصميم ليست سمة لها سعر خاص بها. عندما تحتوي مجموعة البيانات على مؤشر «إعادة التصميم» بجانب CQA، فإن هذا المؤشر يمثل كل ما جلبته عملية إعادة التصميم ولم تسجله الميزات الأخرى في النموذج (المساحة الإجمالية القابلة للاستخدام، قطعة الأرض، الإطلالة، وCQA): الحالة التي تتجاوز ما تظهره CQA والجودة التي لا تسجلها مجموعة البيانات. وهذه

هي الطريقة التي يجب بها قراءة العبارة الواردة في المتغير الكامن (Latent Variable) في العدد 1، والتي تنص على أن عملية إعادة التصميم في H Comp «تضيف ما يقارب \$60,000 إلى قيمة الحزمة الخاصة بها»، وهذا هو ما يركز عليه المثال الوارد في هذا المقال. لكي يحدث التحيز، يجب أن يتحقق شرطان معاً: يجب أن يكون للسمة التي تم استبعادها تأثير على السعر، ويجب أن تتغير مع إحدى السمات التي تم تضمينها. وإذا لم يتحقق أي من هذين الشرطين، فلا يوجد تحيز.

إن هذا الشرط المكون من جزأين مهم لأن الصيغة الشائعة للقاعدة --- وهي أن «النموذج الذي يغفل متغيراً مهماً يعطي إجابات خاطئة» --- هي صيغة خاطئة كما هي مذكورة. إذا كانت إحدى السمات تؤثر على السعر ولكنها لا تتحرك مع أي من السمات المضمنة، فإنها توسع نطاق تباين الأسعار حول النموذج، وبالتالي توسع نطاق الأخطاء المعيارية (وهي مقياس لمدى ابتعاد المعامل المقدر عن الرقم الذي قد تعطيها عينة كبيرة جداً)، ولكنها لا تغير المعاملات نفسها؛ فالمقطع (المبلغ الثابت بالنقد للنموذج) يستوعب متوسط مساهمة السمة المستبعدة. في هذه الحالة، تُفقد الدقة، وليس الصحة. من ناحية أخرى، فإن السمة التي تتغير مع السمات المدرجة ولكنها لا تؤثر على السعر لا تكلف شيئاً على الإطلاق. ولا يحدث التحيز إلا عندما يكون كلا الأمرين صحيحين.

وتعمل هذه الطريقة عن طريق الاستبدال. فإذا كانت إحدى السمات التي تؤثر على السعر مفقودة من المتغيرات المستقلة (السمات التي يستخدمها الانحدار لتفسير السعر)، فإن انحدار OLS (طريقة المربعات الصغرى العادية) لا يترك مساهمة تلك السمة دون تفسير ببساطة. فالسمات المضمنة التي تتغير بالتزامن مع السمة المفقودة تعمل كبدايل جزئية لها، وتستوعب معاملاتها أي جزء من تأثير السمة المفقودة يمكنها التنبؤ به. ولم يعد معامل السمة المضمنة يمثل تأثير تلك السمة فحسب؛ بل هو ذلك التأثير مقترناً بحصة من تأثير السمة المفقودة التي استعارها منها.

التحيز دائماً ما يكون مرتبطاً بسؤال معين. تعتبر المعاملات في النموذج الذي لا يتضمن مؤشر «إعادة النمذجة» الإجابة الصحيحة لسؤال واحد --- وهو كيف يتحرك السعر مع درجة CQA عبر هذه المبيعات، مع مراعاة كل العوامل المرتبطة بالدرجة --- وتكون الإجابة الخاطئة عندما تُفسر على أنها قيمة السوق بالدولار لكل نقطة من نقاط CQA مع بقاء كل العوامل الأخرى ثابتة، وهو ما يفعله التعديل عند تفسيرها. النموذج ليس خاطئاً بسبب وجود سمة أخرى؛ بل إن معاملاته خاطئة بالنسبة للاستخدام المقصود منها.

للمحصة المستعارة حجم محدد، يُعطى بواسطة معادلة. بالنسبة لسمة مستبعدة w ذات المعامل الحقيقي β_w ، فإن الانحراف في معامل السمة المدرجة z_j يساوي

$$\text{bias}(\hat{\beta}_j) = \beta_w \cdot \delta_j,$$

حيث $\hat{\beta}_j$ --- حيث تشير علامة القبعة إلى تقدير --- هو المعامل الذي يقدّر الانحدار z_j ، و δ_j هو معامل z_j في الانحدار المساعد لـ w على جميع السمات المدرجة. هذه هي صيغة ثيل¹. تُحسب من الانحدارات التي تتمتع بالملاءمة، مع المعاملات المُقدّرة للسمة المستبعدة، وهي تنطبق تماماً في كل عينة؛ أما مع المعاملات الحقيقية، فإنها تنطبق في المتوسط.

الصيغة الضريبية هي الحالة الخطية. في نموذج MARS، يتم امتصاص تأثير السمة المستبعدة من قبل أي من تلك المقاطع التي تترافق معها، وبالتالي يتم توزيع التحيز عبرها بدلاً من أن يتركز في معاملات واحدة، لكنه موجود، وللسبب نفسه.

هذه الصيغة أكثر فائدة مما تبدو عليه، حيث يمكن في كثير من الأحيان تحديد إشارة كل من العاملين استناداً إلى المعرفة بالسوق، حتى لو تعذر قياس أي منهما. فعندما تؤدي السمة المستبعدة إلى ارتفاع السعر ($\beta_w > 0$) وترتبط بزيادة في إحدى السمات المضمنة، مع بقاء السمات المضمنة الأخرى ثابتة ($\delta_j > 0$)، فإن المعاملات الخاصة بتلك السمة تكون مرتفعة للغاية. والمُقيّمين الذين يعرفون اتجاه كلتا العلاقتين فقط، يعرفون اتجاه الخطأ. ملاحظة واحدة: δ_j هي علاقة تُبقى فيها السمات الأخرى المدرجة ثابتة، ومع وجود أكثر من سمة واحدة مدرجة، فإن الارتباط البسيط (رقم من -1 إلى +1 يقيس كيف ترتفع وتنخفض سمتان معاً، متجاهلاً جميع السمات الأخرى) يمكن أن يسير في الاتجاه المعاكس. في المثال أدناه، ترتبط مساحة المعيشة بمؤشر إعادة التصميم (الارتباط +0.09)، ومع ذلك فإن قيمة δ الخاصة بها سالبة، وبالتالي فإن معاملها منخفض جداً.

أين يصادفه المقيم. يمكن كتابة شبكة مقارنة المبيعات كنموذج تجميعي له نفس شكل نموذج الانحدار --- أي عدد قليل من السمات المختارة يدوياً، مع تعديل بالدولار لكل منها --- وعندما تُستخلص تعديلاته من بيانات السوق، فإنه يصبح عرضةً لتحيز المتغيرات المهمة بنفس الطريقة التي يتعرض لها أي نموذج انحدار. تمثل خطوط الشبكة السمات المدرجة. وأي سمة من سمات المنزل تؤثر على سعره ولكنها ليست خطأ في الشبكة يتم استبعادها.

ومن بين العوامل التي غالباً ما يتم تجاهلها: الجوانب المتعلقة بحالة العقار التي لا تعكسها CQA، والجودة في الحالات التي لا يتم فيها تسجيل CQA، ووظائف المخطط الأرضي، والضوء الطبيعي، والتعرض للضوضاء، والموقع المحدد داخل المجمع السكني، ودوافع البائع، والجهود التسويقية المبذولة في الإعلان. ويمكن أن يؤثر كل عنصر من هذه العناصر على السعر، وكثير منها يرتبط بعوامل تأخذها الشبكة في الاعتبار.

تعد عملية إعادة التصميم مثلاً توضيحياً، لأنها تُسجّل ومع ذلك لا تزال هناك ثغرات. يسجل تقرير التقييم السكني الموحد (URAR) ما إذا كان المطبخ والحمامات قد تم تحديثها أو إعادة تصميمها، وخلال أي فترة، كما أن تعريفات الحالة في مجموعة بيانات التقييم الموحدة

¹Theil, H. (1957). Specification Errors and the Estimation of Economic Relationships. *Revue de l'Institut International de Statistique*, 25(1/3), 41-51.

(UAD) نفسها تعتمد على مدى التحديث. ومع ذلك، فإن الشبكة تحتوي على أعمدة للحالة والجودة، ولا يوجد عمود لإعادة التصميم --- وهذا صحيح عندما يكون السؤال هو سعر السوق للحالة، حيث إن عمودي الحالة والجودة يعكسان ما أضافته عملية إعادة التصميم، لأن وجود عمود ثالث سيؤدي إلى احتسابها مرتين. وعندما يكون السؤال بدلاً من ذلك هو: ما تأثير عملية إعادة التصميم المقترحة على سعر هذا المنزل؟ فإن إبقاء «الحالة» و«الجودة» ثابتتين يزيل القناة ذاتها التي تعمل من خلالها عملية إعادة التصميم؛ ويتناول البند 009 (الداخلية) هذا السؤال. ويحدث التسرب من خلال ما تغفل عنه CQA في سوق تُنسب فيه أعلى درجات CQA في الغالب إلى عمليات التجديد الحديثة، وتضيف عمليات التجديد قيمة لا تسجلها الدرجات، فإن تعديل الحالة المستند من تلك المبيعات في التقييم التقليدي --- عن طريق المبيعات المقترنة، أو عن طريق انحدار يتضمن الحالة --- يحمل تلك القيمة معه. ووفقاً لصيغة ثيل، يكون هذا التعديل عندئذٍ مرتفعاً جداً؛ فهو لا يشمل تأثير الحالة بعد ذاتها فحسب، بل يشمل أيضاً جزءاً مما لم تسجله CQA. والمُقيّمين الذين يأخذون هذا التعديل ويطبقونه على منزل حصل على أعلى CQA بفضل الصيانة التي خضع لها، سيقومون بتعديل مفرط. وبموجب «منهج القيد المتبقي» الذي تتبعه المجلة، لا يتم استخلاص أي معدل خاص بالحالة. يقسم المُقيّمين المتبقي لكل عقار مقارن بين الحالة والجودة والميزات الأخرى التي لا تقيسها البيانات، وفقاً لتقدير المُقيّمين، في ظل قيد واحد: يجب أن يساوي مجموع مساهمات قيمة المكونات المتبقية للعقار المقارن. ثم يتم تقسيم قيمة عملية إعادة التصميم حسب تقدير المُقيّمين، ولا يتم إدراجها في بند واحد.

التسرب الذي وُصف للتو هو نسخة أكثر دقة من الاعتراض الذي يثيره المُقيّمون بشأن التعديلات المستندة إلى السوق، وهو أن التعديلات المستندة من مجموعة بيانات معينة لا يمكن نقلها بشكل جيد. والسبب في ذلك هو أن التعديل المستخرج يتألف من كل من التأثير المحدد والتأثيرات غير المحددة المصاحبة له، وتعتمد تركيبته على العلاقة بين العناصر المختلفة في مجموعة البيانات التي استُمد منها. ومن الأمثلة الأخرى الحالة التي تسفر فيها «المبيعات المزدوجة» في التقييم التقليدي --- أي عمليتين بيع متشابهتان في كل شيء باستثناء السمة المعنية، حيث يُفسر الفرق في السعر على أنه قيمة تلك السمة --- عن رقم مرتفع بشكل غير معقول. قد لا يكون السبب خطأ في القياس بل «تسرباً»: فالصفتان تختلفان في السمة المذكورة وفي شيء آخر مرتبط بها، ويُعزى فرق السعر إلى كليهما.

لماذا يهم في تماسك الرأي. إن أي تقرير لا يتناول التحيز الناجم عن المتغيرات المهمة يجعل تعديلات «الهيدونيك» و«شبكة» فيه عرضة لهذا الانتقاد.

لا يمكن استخلاص التحيز من إحصائيات الملاءمة. يمكن أن يتمتع النموذج بنسبة تفسير عالية (R^2) (أي نسبة التباين في الأسعار التي يعيد النموذج إنتاجها)، وبقياس صغيرة (الاختلافات بين أسعار البيع وأسعار النموذج)، ومعاملات تجتاز الاختبار الإحصائي المعتاد (اختبار "t"، الذي يبحث فيما إذا كان من المعقول أن يكون المعامل صفراً)، ومع ذلك قد يكون النموذج متحيزاً بشكل كبير، لأن التحيز ينشأ من العلاقة بين السمات (ما الذي يتغير مع ماذا)، وليس من حجم القيم المتبقية. ولا يُعد استخدام " R^2 " كدليل على موثوقية النموذج إجابة على السؤال. في المثال أدناه، تبلغ قيمة " R^2 " 97.0، بينما يخطئ معامل الشرط بنسبة 74 في المائة.

زيادة البيانات لن تحل المشكلة. فالتحيز لا يقلص مع زيادة حجم العينة. فعندما يتم استبعاد نفس السمة من ألف عملية بيع من نفس السوق، ينتج عن ذلك تقدير يتمحور حول نفس الرقم الخاطئ، ولكن بتقدير أكثر دقة: حيث يقلص الخطأ المعياري حول التقدير المتحيز، وهو ما يعد أسوأ من وجود نطاق واسع حوله، لأن التقرير سيبدو أكثر ثقة. وينطبق الأمر نفسه على نموذج MARS: فزيادة عدد المبيعات تجعل السطح الملائم أكثر استقراراً، ولكن حول نفس النتيجة المتحيزة؛ ويقاس نطاق «varmod» (الانحراف المعياري) انتشار الأسعار، وليس التحيز. وهذا يصح المتغيرات في العدد 1، التي ذكرت أن التحيز سيتقلص مع زيادة عدد المقارنات وأن المعاملات الأخرى كانت متحيزة نحو الأعلى. لا يقلص التحيز مع زيادة عدد المبيعات، وفي المبيعات الثمانية الواردة في العدد 1 نفسه، كما في المبيعات الستة عشر أدناه، فإن معاملات «المنظر» و«المساحة الإجمالية القابلة للتأجير» و«المساحة» منحازة نحو الانخفاض؛ أما معامل «الحالة» فهو الوحيد المنحاز نحو الارتفاع.

غالباً ما يكون من الممكن تبرير اتجاه التحيز حتى عندما يتعذر معرفة حجمه. في التقييم التقليدي، حيث يتم استخلاص تعديل الحالة من المبيعات --- إما عن طريق المبيعات المزدوجة أو عن طريق انحدار يتضمن الحالة --- فإن النهج العملي لا يتمثل في التأكيد على أن التعديل غير متحيز، بل في تحديد السمات المعقولة التي تم استبعادها، واستخدام المعرفة بالسوق لإعطاء كل منها إشارة لـ β_w و δ_j ، وبيان اتجاه التحيز الناتج. من المحتمل أن يكون تعديل الحالة مبالغاً فيه. فمعظم المبيعات التي حصلت على أعلى الدرجات هي عبارة عن عقارات تم تجديدها مؤخراً، وقد أحدثت عمليات التجديد تحسينات لا تسجلها البيانات، مثل المطابخ الجديدة والتشطيبات. وترفع هذه التحسينات السعر وتحدث في الغالب في المبيعات ذات الدرجات العالية، لذا تم احتساب جزء من قيمتها على أنها حالة" هو بيان معقول. تم استيفاء الشرطين معاً، وكلاهما يشير إلى نفس الاتجاه: التحسينات التي تم استبعادها ترفع السعر (β_w موجب) وترتبط بالدرجات العالية (δ_j موجب)، لذا فإن التحيز صاعد. يحول مثل هذا البيان مشكلة غير مقاسة إلى مشكلة معترف بها ومحددة الاتجاه، وهو بالضبط ما يحتاجه القارئ لتقييم الاستنتاج. في منهج القيد المتبقي، لا يتم استخلاص معدل الحالة، ولا يُطرح السؤال بشأن الحالة.

مثال تقييمي محلول. تسجل مجموعة بيانات «باسيفيكا» الموسعة التي تضم ستة عشر عقاراً، بالنسبة لكل عقار، المساحة الإجمالية للمعيشة (GLA)، ومساحة الأرض، والإطلالة، ودرجة الحالة، ومؤشر إعادة التصميم الملحوظ، وسعر البيع. ونظراً لأن كل سعر تم حسابه باستخدام معادلة معروفة خالية من العشوائية (البيانات خالية من الضوضاء)، فإن الانحدار الذي يستخدم السمات الصحيحة يستعيد تلك المعادلة بدقة، وأي انحراف عنها يمثل انحرافاً خالياً تماماً من مكون الضوضاء.

Comp	GLA (sf)	مساحة الأرض (قدم مربع)	عرض	الشروط	تم تجديده	السعر (\$)
أ	1٠650	9٠000	نعم	00.7	لا	1٠425٠000
ب	1٠650	7٠200	لا	00.7	لا	1٠280٠000
C	1٠850	7٠500	لا	00.7	لا	1٠407٠500
د	1٠450	7٠000	لا	00.7	لا	1٠155٠000
E	1٠700	8٠000	نعم	00.9	لا	1٠470٠000
F	1٠550	7٠100	لا	00.5	لا	1٠177٠500
G	1٠700	10٠500	لا	00.7	لا	1٠392٠500
H	1٠600	7٠300	لا	00.7	نعم	1٠312٠500
أنا	1٠750	8٠200	لا	00.9	نعم	1٠465٠000
J	1٠600	7٠600	لا	00.7	نعم	1٠320٠000
ك	1٠900	9٠500	نعم	00.9	نعم	1٠687٠500
L	1٠500	6٠800	لا	00.5	لا	1٠140٠000
م	1٠800	8٠800	لا	00.7	لا	1٠410٠000
ن	1٠650	7٠400	نعم	00.7	لا	1٠385٠000
O	1٠550	7٠900	لا	00.9	نعم	1٠337٠500
P	1٠720	9٠200	لا	00.5	لا	1٠332٠000

العدد A--H هي نتائج المقارنة مع الإصدار الأول، مع بقاء خصائصها وأسعارها المنشورة دون تغيير؛ وقد وصف الإصدار 1 الحالة بالكلمات ورمزها ب 1-3 لدرجة تدهورها؛ وفي CQA لهذه المجلة، تُقابل الكلمات بمتوسط 00.5، وجيد 00.7، وممتاز 00.9، لذا فإن خطوة واحدة في الإصدار 1 تعادل نقطتين. تم الآن تسجيل عملية إعادة التصميم الخاصة بالوحدة H Comp --- التي تم الكشف عنها في الإصدار 1 باعتبارها مصدر الفرق المتبقي، ولكن لم يتم تسجيلها كمتغير ---. يدفع السوق \$20,000 لكل نقطة من نقاط الحالة. في البيانات المُعدّة، يُباع المنزل المُجدد بمبلغ \$60,000 أكثر من منزل له نفس المساحة الإجمالية القابلة للتأجير (GLA) وقطعة الأرض، والإطلالة، ودرجة CQA. هذا المبلغ \$60,000 ليس سعرًا لعملية التجديد بحد ذاتها؛ بل يمثل كل ما أضافه التجديد ولم تسجله الميزات الأخرى. كما تحصل المنازل التي خضعت للتجديد على CQA أعلى في المتوسط (20.8 مقابل 64.6)، لذا فإن مؤشر التجديد يتحرك مع CQA المضمنة، وهو الشرط الثاني الذي يتطلبه التحيز: $\text{corr}(\text{remodel}, \text{cond}) = 0.55$. عمليًا انحدار على نفس العقارات الستة عشر:

تم حذف جزء من النص	شملت أعمال التجديد	التحيز
اعتراض	-83,923.57	-53,923.57
GLA	64.586	-13.36
لوت	32.24	-0.68
عرض	46.79٠178	-20,821.54
الشرط (لكل نقطة)	56.34٠815	+14,815.56
إعادة التصميم	---	60٠000
R^2	9733.0	0000.1

القيمة المضافة (\$60,000) في خط «إعادة التصميم» لا تمثل سعر عملية إعادة التصميم نفسها. ففي مجموعة البيانات المُعدة هذه، تمثل هذه القيمة الفرق السعري المرتبط بمؤشر «إعادة التصميم» عند تساوي كل من المساحة الإجمالية القابلة للتأجير (GLA) والمساحة الأرضية والإطلالة والـ CQA؛ وهذا هو ما «تستعيده» «دالة التوليد».

معامل الحالة مرتفع جدًا بمقدار \$14,816 نقطة --- حيث يُشير إلى \$34,816 لكل نقطة من الحالة، في حين أن الرقم السائد في السوق هو \$20,000، وهو ما يمثل مبالغة بنسبة 74 في المائة. معامل العرض منخفض جدًا بمقدار \$20,822. كلا الخطأين غير مرتبين في الملاءمة: النموذج بدون مؤشر إعادة التصميم يحتوي على $R^2 = 0.9733$ وبقيًا تبدو عادية، باستثناء ثلاث مبيعات قريبة من \$35,000. ولا يوجد في R^2 أو في الاختبارات التي أجريت على المعاملات ما يشير إلى وجود مشكلة (لو كانت هذه المتبقيات مجرد ضوضاء، لكان كل معامل قد اجتاز ذلك الاختبار عند مستوى الـ 5 في المائة المعتاد).

تفسر صيغة ثيل الأرقام بدقة. وعند إجراء تحليل انحدار لمؤشر «remodel» على السمات الأربع المُدرجة، نحصل على $\delta_{\text{cond}} = 0.24692602$ و $\delta_{\text{view}} = -0.34702564$. ومن ثم

$$\beta_w \cdot \delta_{\text{cond}} = 60,000 \times 0.24692602 = 14,815.56, \quad \beta_w \cdot \delta_{\text{view}} = 60,000 \times (-0.34702564) = -20,821.54,$$

مما يطابق التحولات الملحوظة بدقة. ولا يُفسر الانحراف تقريباً بواسطة الصيغة؛ بل هو الصيغة نفسها. الضوضاء ستؤدي إلى تغيير أقل مما قد يتوقعه المرء: مع وجود ضوضاء في البيانات، تظل الصيغة صحيحة تماماً بين الانحدارين الملائمين (بدون مؤشر إعادة البناء ومعه)، باستخدام المعاملات المقدرة للمؤشر $\hat{\beta}_w$ بدلاً من β_w ؛ وبالمقارنة مع المعاملات الحقيقية، فإنها تظل صحيحة في المتوسط عبر العديد من العينات.

يستحق معامل العرض أن نلقي عليه نظرة ثانية، لأن انحيازه يسير في الاتجاه المعاكس، والسبب وراء ذلك مفيد. مع بقاء الخصائص الأخرى ثابتة، تقل احتمالية إعادة تصميم العقارات ذات الإطالة في هذه العينة ($\delta_{view} < 0$). لذا ينخفض معامل الإطالة بدلاً من أن يرتفع: فهو \$20,822 منخفض جداً، في حين أن معامل الحالة \$14,816 مرتفع جداً. هذان الخطان هما تأثيران منفصلان لنفس الإغفال، وليس انتقالاً بينهما. والمُقيّمين الذين يعتمدون في استنتاجهم على الحالة فقط ما كانوا ليتوقعوا وجود خطأ في تعديل الإطالة على الإطلاق --- ومع ذلك، فإن تعديل الإطالة الذي تم التقليل من قيمته بنسبة 21 في المائة هو النتيجة المباشرة لنفس الإغفال.

من الناحية العملية، يتركز الضرر في الأماكن التي يقل احتمال أن يبحث فيها المُقيّمين عنه. لننظر أولاً إلى دور النموذج المتحيز كمتنبئ. في حالة موضوع لم يخضع لأي تجديد ومطابق لـ N Comp --- 1,650 قدم مربع من المساحة الإجمالية القابلة للتأجير (GLA) على قطعة أرض مساحتها 7,400 قدم مربع مع إطالة و CQA تبلغ 00.7 --- فإنه يشير إلى \$1,386,914 مقابل السعر المتوسط الحقيقي البالغ \$1,385,000 للمنازل التي تتمتع بهذه الميزات بالضبط --- وفقاً لصيغة الخمس ميزات الواردة في الإدخال رقم 007 (التوقع الشرطي)، مع إعادة التصميم $= 0$ --- وهو خطأ يبلغ \$1,914، أو 14.0 في المائة. وعلى مستوى جميع عمليات البيع الست عشرة، يبلغ الخطأ النموذجي حوالي \$21,000، لذا فإن النموذج متحيز بشدة ومع ذلك يقدم تنبؤات مقبولة، وهذا ليس من قبيل الصدفة: حيث إن طريقة OLS تتكيف مع العينة المعطاة، والمعاملات المتحيزة هي بالضبط تلك التي تقترب أكثر من غيرها من أسعار العينة. دقة التنبؤ ليست دليلاً ضد انحياز المتغيرات المهمة، كما أن التحقق المتقاطع --- اختبار النموذج على المبيعات التي تم استبعادها من عملية الملاءمة --- لا يستبعده أيضاً: فنتائج التحقق المتقاطع الجيدة لا تدل على خلو المعاملات من التحيز، لأن النموذج المُحدّد بشكل خاطئ يمكنه التنبؤ بشكل جيد حيثما تتوافق السمات معاً كما حدث في عملية الملاءمة. وبترك كل عقار على حدة، فإن أكبر خطأين --- حوالي \$56,000 في K و \$48,000 في G --- يقعان على العقارات التي تتبع نمط هذه العينة، حيث تنتمي درجات CQA العالية في الغالب إلى العقارات التي خضعت للتجديد؛ أما E، التي تخالف هذا النمط، فتحتل المرتبة الثالثة بحوالي \$45,000، و H و J، اللتان تخالفان هذا النمط أيضاً، تحتلان المرتبتين الرابعة والخامسة بحوالي \$41,000. لا تحتوي البيانات هنا على أي تشويش، لذا فإن كل خطأ من هذه الأخطاء ينشأ عن مؤشر التجديد الذي تم استبعاده، ولا ينشأ أي منها عن التشويش؛ وفي المبيعات الفعلية، حيث يوجد تشويش أيضاً، لا يشير حجم الخطأ بأي شكل إلى ما إذا كان ناجماً عن سمة تم استبعادها أم عن التباين العادي في القيم المتبقية.

يظهر الضرر في اللحظة التي تُستخدم فيها المعاملات كت عديلات بدلاً من كونها قاعدة للتنبؤ. فالمُقيّمين الذين يستخلصون تعديلاً متعلقاً بالحالة من هذا النموذج يطبقون \$34,816 لكل نقطة، في حين أن الرقم السائد في السوق هو \$20,000 --- \$29,631 زيادة كبيرة جداً على كل عقار مقارنة بـ تعديل عبر النقطتين من «جيد» إلى «ممتاز» --- وتعديلاً متعلقاً بالإطالة بقيمة \$79,178 في حين أن الرقم السائد في السوق هو \$100,000. الشبكة هي المكان الذي يتم فيه قراءة معاملات الانحدار واحدة تلو الأخرى، وهذا هو بالضبط القراءة التي يدمرها التحيز.

كما أن التنبؤ يفشل فشلاً ذريعاً في حالة العقارات التي تخرج عن النمط الذي بُني عليه التحيز. منزل مُعتنى به جيداً مثل المنزل موضوع الدراسة (1,650 قدمًا مربعًا من المساحة الإجمالية القابلة للتأجير، و 7,400 قدم مربع من مساحة الأرض) ولكنه يفتقر إلى الإطالة، وحصل على CQA ممتاز بلغ 00.9، ولم يخضع لأي عملية إعادة تصميم من قبل، وهو مذكور على الرابط \$1,377,367 مقابل السعر الفعلي البالغ \$1,325,000، مما يمثل تقييماً زائداً بنسبة \$52,367 (95.3 في المائة)، لأن CQA 00.9 في هذه المبيعات كانت في الغالب مصحوبة بتجديد، ومعدل حالة النموذج يشمل تكلفة التجديد البالغة \$60,000 إلى جانب CQA. يُشار إلى منزل تم تجديده وهو مطابق تماماً للمنزل موضوع الدراسة (إطالة، CQA 00.7) في \$1,386,914 مقابل \$1,445,000، وهو أقل من قيمته الحقيقية بمقدار \$58,086 (02.4 في المائة). هذه هي العقارات غير النمطية --- تلك التي غالباً ما يُستعان بالمُقيّمين خصباً لتقييمها، وتلك التي سيقراً التقرير الخاص بها بأكثر قدر من الدقة.

References.

- Craytor, W. B. (2026). Residual Constraint Approach (RCA): Framework & Protocol. *The Valuation Engineer*, 1(1). <https://doi.org/10.67330/kgebf46>
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Theil, H. (1957). Specification Errors and the Estimation of Economic Relationships. *Revue de l'Institut International de Statistique*, 25(1/3), 41–51.

endogeneity ;variable latent ;price implicit ;regression hedonic ;expectation conditional Cross-references:

Formalization Prolog Addendum:

Prolog journal's the in entry this formalizes (008-omitted-variable-bias.pl) companion machine-readable following The article. this alongside galley browsable a as published also is It base. knowledge

```
later-or-0.3-AGPL :Identifier-License-SPDX %
Engineer Valuation The by Published .Craytor Bert William 2026 (C) Copyright %
can you :software free is program This .(Net Vista Pacific) Journal %
General Affero GNU the of terms the under it modify or/and it redistribute %
3 version either ,Foundation Software Free the by published as License Public %
distributed is It .version later any (option your at) or ,License the of %
or MERCHANTABILITY of warranty implied the even without ,WARRANTY ANY WITHOUT %
or it beside file COPYING the see ;PURPOSE PARTICULAR A FOR FITNESS %
.<html.0.3-agspl/licenses/org.gnu.www//:https> %

,2 Issue ,1 .Vol Foundations to companion — pl.bias-variable-omitted-008 %%
.0.2.v1 ,(omitted_variable_bias) Bias Variable Omitted ,008 Entry %%
the to delta_j * beta_omitted = bias each) 0/theil_formula_holds :Checks %%
.(000,60 s'indicator remodel the) 0/remodel_reading_holds ;(cent %%

so ,time run at loaded not is (dictionary/tools/Issues) dictionary The %
The .Sryer and Prolog-SWI ,Trealla in unchanged runs file the that %
.linter dictionary the by read ,facts plain are below declarations %
.('1.2026-dictionary')dictionary_release_pinned

,conditional_expectation ,omitted_variable_bias ])uses_terms
,latent_variable ,implicit_price ,hedonic_regression
,residual ,ordinary_least_squares ,multicollinearity
,paired_sales_analysis ,comp_selection ,adjustment
.([ uspap ,sales_comparison_approach
.([2/introduced_by ,2/related ,2/depends_on])uses_relations

.(omitted_variable_bias)article_concept
.((2026)year ,(2)number ,(1)volume)article_issue
.('008.2026.foundations')article_id

.('pacific_comps_extended/kb/../../')consult)initialization -:
;request on predicates list the and 2/forall loads Prolog Sryer %
below loads guarded The .in built them have Trealla and Prolog-SWI %
.three all in unchanged run file the let %
.((true ,_ ,((lists)library)use_module)catch)initialization -:
.((true ,_ ,((iso_ext)library)use_module)catch)initialization -:

.(state ,condition)kind
.((condition)measurement_of ,condition_score)kind
.(state ,quality)kind
.(event ,remodel)kind
.(condition ,remodel)changes
.(quality ,remodel)changes

.(unrecorded_condition ,remodel_numeric)stand_in
.(unrecorded_quality ,remodel_numeric)stand_in

,affects_outcome)ovb_condition
.("outcome the on effect nonzero a have must variable omitted The"
,correlated_with_included)ovb_condition
not ,partial) regression auxiliary the in regressor included one least at on coefficient nonzero a have must variable omitted The"
.(".(correlation ,marginal merely ⇔
```

```

-: (Delta ,BetaOmitted)biased
  ,0 == BetaOmitted
  .0 == Delta

,affects_but_uncorrelated)consequence
omitted the absorbs intercept the) unbiased remain coefficients slope ;errors standard widens and variance residual Inflates"
  .(" (mean s'variable ↔
  ,correlated_but_no_effect)consequence
  .(" .all at cost No"

-: (Bias ,Delta ,BetaOmitted)theil_bias
  .Delta * BetaOmitted is Bias

  .(remodel_numeric)omitted_variable
  .(60000 ,remodel_numeric)omitted_variable_effect

  .(0.10 ,0.0)condition_scale
  .(0.5 ,average)condition_score
  .(0.7 ,good)condition_score
  .(0.9 ,excellent)condition_score
  .(20000)condition_rate

  .(571.83923- ,intercept ,remodel_omitted)coefficient
  .(645.586 ,gla_sqft ,remodel_omitted)coefficient
  .(323.24 ,lot_sqft ,remodel_omitted)coefficient
  .(461.79178 ,view_numeric ,remodel_omitted)coefficient
  .(561.34815 ,cond_score ,remodel_omitted)coefficient

  .(30000- ,intercept ,remodel_included)coefficient
  .(600 ,gla_sqft ,remodel_included)coefficient
  .(25 ,lot_sqft ,remodel_included)coefficient
  .(100000 ,view_numeric ,remodel_included)coefficient
  .(20000 ,cond_score ,remodel_included)coefficient
  .(60000 ,remodel_numeric ,remodel_included)coefficient

  .(89872625.0- ,intercept)aux_slope
  .(00022259.0- ,gla_sqft)aux_slope
  .(00001129.0- ,lot_sqft)aux_slope
  .(34702564.0- ,view_numeric)aux_slope
  .(24692602.0 ,cond_score)aux_slope

  .(9733.0 ,r_squared ,remodel_omitted)model_fit
  .(0000.1 ,r_squared ,remodel_included)model_fit

-: (Bias ,R)observed_bias
  ,(Omitted ,R ,remodel_omitted)coefficient
  ,(True ,R ,remodel_included)coefficient
  .True - Omitted is Bias

-: theil_formula_holds
  ,(BetaOmitted ,_)omitted_variable_effect
  ,( (Delta ,R)aux_slope ,(Observed ,R)observed_bias ))forall
  ,(Predicted ,Delta ,BetaOmitted)theil_bias )
  .(( 01.0 > (Predicted - Observed)abs

-: remodel_reading_holds
  ,(Beta ,remodel_numeric)omitted_variable_effect
  ,(Coef ,remodel_numeric ,remodel_included)coefficient

```

.01.0 > (Beta - Coef)abs

.(56.14815 ,cond_score)reported_bias
.(54.20821- ,view_numeric)reported_bias

—: (Pct ,R)relative_bias
,(Bias ,R)observed_bias
,(True ,R ,remodel_included)coefficient
,0 =\= True
.True / Bias * 100 is Pct

,prediction)use_of_coefficients
.("offset can coefficients individual in biases ;point a at surface fitted whole the Evaluate"
,adjustment)use_of_coefficients
.("offset not does and directly enters bias ;price implicit an as time a at coefficient one Read"

.(1386914 ,remodel_omitted ,subject_typical)prediction_check
.(1385000 ,remodel_included ,subject_typical)prediction_check

.(1377367 ,remodel_omitted ,excellent_unremodeled)prediction_check
.(1325000 ,remodel_included ,excellent_unremodeled)prediction_check
.(1386914 ,remodel_omitted ,remodeled_with_view)prediction_check
.(1445000 ,remodel_included ,remodeled_with_view)prediction_check

—: (S)atypical_property_fails_worse
,(A ,remodel_omitted ,S)prediction_check
,(B ,remodel_included ,S)prediction_check
.50000 < (B - A)abs

,fit_statistics)defensibility_consequence
.("residuals the not ,structure correlation the in lives it ;size residual or squared-R from undetectable is Bias"
,sample_size)defensibility_consequence
.("number wrong estimated tightly a gives data more ;n with shrink not does Bias"
,direction_disclosure)defensibility_consequence
is magnitude when even statement directional defensible a yields knowledge market from delta_j and beta_omitted Signing"
.("unknown ⇔

,what_the_scores_miss)practice_encounter
RCA .value unrecorded their carries remodels with go scores top where extracted adjustment condition a :appraisal Traditional"
.("rate condition no extracts ⇔
,no_remodel_line)practice_encounter
a :rightly ,line remodel no and lines quality and condition has grid the ;updates bathroom and kitchen records URAR The"
.("twice value same the count would line remodel ⇔
,adjustments_do_not_travel)practice_encounter
.("specific-dataset is it so ,it with correlated effects unnamed and effect named the of composite a is adjustment extracted An"
,implausible_paired_sale)practice_encounter
.("attribute omitted correlated a from leakage reflect can cost than larger far adjustment sales-paired A"