أخطاء توصيف النموذج: توصيف النموذج واختبارات التشخيص

لا يمكن تطبيق الاقتصاد القياسي بشكل آلي، فلابد من فهم تداعيات الموقف والمهارات المتاحة.

عندما نعبر جســرًا ما لا يكون لدينا أيّ قلق من انهياره، نظرًا للثقة الشــديدة في المهندســين الذين قاموا بتصــميم وبناء هذا الجســر. على الاقتصــاديين أن يفعلوا الشــيء ذاته في النماذج الاقتصــادية، وإلا عليهم أن يلصقوا التحذير التالي بالنماذج المختارة "لا تستخدمه فقد ينهار بك!".

الاقتصاديون يبحثون عن "الحقيقة" على مرِّ السنوات السابقة، على أساس أنَّ الاقتصاديين أشخاصٌ يبحثون عن قطة سوداء غير موجودة أصلاً في حجرة مظلمة، وبالتالي مستحيل أن يجدوها.

أحد فروض نموذج الانحدار الخطي التقليدي (CLRM)، الفرض 9. إن نموذج الانحدار المســتخدم في التحليل "صحيح" التوصيف، وإذا كان النموذج غير صحيح التوصيف، فإننا نتعرض إلى المشكلة المعروفة باسم "خطأ توصيف النموذج" أو "تحيز توصيف النموذج". في هذا الفصل، سنلقي الضوء على مثل هذا الفرض، لأن البحث عن النموذج الصــحيح يشــبه البحث عن Grail المقدس. بالتحديد ســنناقش الأســئلة التالية (جوجاراتي، 2015 صفحة 649):

1. كيف يمكن للباحث تحديد النموذج "الصــحيح"؟ أو بمعنى آخر، ما هي الأدوات التي يمكن اســتخدامها لتحديد النموذج المناسب للتحليل التطبيقي؟
2. ما هي أخطاء توصيف النموذج المحتملة عند التطبيق؟
3. ما هي عواقب أخطاء التوصيف؟
4. كيف يمكن للباحث أن يكتشــف خطأ التوصــيف؟ أو بمعنى آخر ما هي بعض أدوات التشــخيص التي يمكن للباحث أن يستخدمها؟
5. إذا كانت هناك أخطاء في التوصيف، ما هي الأساليب العلاجية الممكن استخدامها وما هي فوائدها؟
6. كيف يمكن تقييم أداء النماذج المنافسة؟

موضــوع توصــيف النموذج وتقييمه يعتبر موضــوعًا متعدّد الأبعاد، وله العديد من التطبيقات. ليس ذلك فقط ولكن هناك فروق فلســفية وراء النماذج المختلفة في هذا الموضــوع. بالطبع لن نســتطيع أن نغطي كل ذلك في فصلٍ واحد، ولكن نتمنى أن نلقي الضَّوء على بعض المواضــيع الأساسية الخاصــة بتوصــيف النموذج وتقييمه.

1-معيار اختيار النموذج: MODEL SELECTION CRITERIA

بناءً على دراســـة Richard و Hendry فإن المعايير التي يجب اســـتيفاؤها لاختيار نموذج ما للتحليل التطبيقي هي كالتالي:

1. أن يكون في إطـار البيانـات المتاحـة: أي لابـد أن تكون التنبؤات المتوقعـة من النموذج يمكن فعليًا ومنطقيا الحصول عليها.
2. الاتساق مع النظرية: فلابد أن يكون هناك معنى منطقي اقتصـادي، فمثلاً إذا تحقق فرض الدخل الدائم لـــ Milton Friedman's فإن قيمة الجزء الثابت المقطوع من المحور الصادي في الانحدار الخاص بالاستهلاك الدائم على الدخل الدائم يجب أن يساوي الصفر.
3. المتغيرات المنحدرة تكون غير مترابطة: أي أن المتغيرات المفســـرة أو المنحدرة يجب ألا تكون مرتبطة مع حد الخطأ.
4. اتســاق المعالم: بمعنى أن قيم المعالم يجب أن تكون مســتقرة. وإلا ســيكون التنبؤ صــعبًا. فكما قال Friedman: "اختبار صــلاحية النموذج الافتراضــي الأفضــل هو مقارن تنبؤات النموذج بالواقع". في غياب اتساق المعالم لا يكون مثل هذا التنبؤ موجودًا.
5. ترابط البيانات: أي أن البواقي المقدرة من النموذج، يجب أن تكون عشـوائية تمامًا. بمعنى آخر إذا كان النموذج المختار مناسـبًا فإن بواقي هذا النموذج يجب أن تكون عشـوائية تمامًا white noise وإذا لم تكن كذلك، فإن هناك خطأ توصيف في النموذج. وسنناقش لاحقًا طبيعة أخطاء التوصيف.
6. النموذج الشـامل: فيكون النموذج شـاملاً عندما تكون قادرًا من خلاله على توضـيح نتائج باقي النماذج المحتملة. أي باختصار النماذج الأخرى لا تقدم أي جديد بعد اختيارك للنموذج الشامل.

أحيانا نختار معيارًا ما للنموذج " الجيد" ونهمل معيارًا آخر حتى نســـتطيع تكوين النموذج فعليًا، وبالتالي غالبًا ما يوجد خطأ توصـيف في النموذج المختار عمليًا، وسنناقش ذلك بالتفصـيل في الفقرة القادمة (جوجاراتي، 2015 صفحة 650).

2.13-أنواع أخطاء التوصيف TYPES OF SPECIFICATION ERRORS:

افترض أنه وفقًا للمعايير السابق ذكرها، اخترنا النموذج التالي كنموذج جيد. وافترض أن هذا النموذج يمكن كتابته كالتالي:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + B_4 X_i^3 + u_{1i} \quad \ldots\ldots\ldots (1.2.13)$$

حيث Y : إجمالي تكلفة الإنتاج و X : الناتج. المعادلة (1) تعتبر مثالاً شهيرًا في كتب الاقتصاد، وتسمى دالة التكلفة الكلية التكعيبية.

ولكن افترض أنه لأسباب ما (مثلاً شكل الانتشار) اختار باحث آخر النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \dots \dots . (2.2.13)$$

لاحظ أننا غيرنا الرموز حتى نستطيع التفرقة بين هذا النموذج والنموذج الصحيح.

بما أن (1.2.13) يفترض صحته، فإن (2.2.13) سيكون قيمة خطأ توصيف، والخطأ الموجود هنا هو حذف متغير مهم في النموذج وهو (X_i^3) وبالتالي فإن حد الخطأ u_{2i} في (2.2.13) سيكون في الحقيقة كالتالي:

$$u_{2i} = u_{1i} + \beta_4 X_i^3 \ \dots \dots \dots \dots \dots (3.2.13)$$

سنوضح لاحقا أهمية هذه العلاقة.

والآن افترض أن باحثًا آخر اختار النموذج التالي:

$$Y_i = \lambda_1 + \lambda_2 X_i + \lambda_3 X_i^2 + \lambda_4 X_i^3 + \lambda_5 X_i^4 + u_{3i} \dots \dots \dots . (4.2.13)$$

إذا كان (1.2.13) صحيحًا فإن (4.2.13) يوجد فيه أيضًا خطأ توصيف. هذا الخطأ هو إضافة متغير غير ضروري وغير مهم للنموذج، وذلك حيث إنه في النموذج الصحيح يفترض أن تكون 15 تساوي الصفر. وبناءً على ذلك يكون حد الخطأ في الحقيقة كالتالي:

$$u_{3i} = u_{1i} - \lambda_5 X_i^4 \ \dots \dots \dots \dots \dots (5.2.13)$$

$$= u_i \ \text{ since } 15 \ = \ 0 \text{ in the true model } \ (\text{لماذا؟})$$

والآن افترض أن باحثًا آخر اختار النموذج التالي:

$$\ln Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 X_i^2 + \gamma_4 X_i^3 + u_{4i} \dots \dots \dots (6.2.13)$$

;بالتالي بالنسبة للنموذج الصحيح، فإن هناك تحيزًا في توصيف النموذج السابق، هذا التحيز يرجع إلى استخدام شكل دالة غير صحيح: ففي (1.2.13) تكون Y خطية، بينما في (13.6.2) تظهر في صورة لوغاريتمية خطية.

وأخيراً افترض أن باحثًا آخر اختار النموذج التالي:

$$Y_i^\star = \beta_1^\star + \beta_2^\star X_i^\star + \beta_3^\star X_i^{\star 2} + \beta_4^\star X_i^{\star 3} + u_i^\star \ \dots \dots \dots \dots . (7.2.13)$$

حيث $Y_i^\star = Y_i + \varepsilon_i$ و $X_i^\star = X_i + w_i$ وبالتالي فإن ε_i و w_i فإن الباحث استخدم Y_i و X_i الصحيحة فإن الباحث استخدم $Y_i^\star$ و $X_i^\star$ والتي تشتمل على حدود أخطاء. وبالتالي في (7.2.13) يوجد خطأ توصيف.

في الواقع العملي تكون البيانات بها أخطاء راجعة للتقريب أو أخطاء راجعة إلى عدم التغطية الكاملة للدراسة أو ببساطة أخطاء لحذف بعض المفردات. العلوم الاجتماعية تعتمد عادة على البيانات الثانوية، ولا تكون لدينا وسيلة لمعرفة أنواع الأخطاء المحتملة، والتي تقوم بها المؤسّسة التي جمعت البيانات الأولية.

أحد أنواع أخطاء التوصيف يعتمد على طريقة إدخال الخطأ العشوائي u_i (أو u_t) إلى النموذج. فعلى سبيل المثال، اعتبر نموذج الانحدار ثنائي المتغيرات غير المشتمل على جزء ثابت كالتالي:

$$Y_i = \beta X_i u_i \quad \ldots\ldots\ldots\ldots\ldots\ldots (8.2.13)$$

حيث إن الخطأ العشوائي موجود في صورة مضروب، ويكون $\ln u_i$ مستوفيًا فروض الـCLRM، وافترض أن لدينا أيضًا النموذج التالي: $Y_i = \alpha X_i + u_i \quad \ldots\ldots\ldots\ldots\ldots\ldots (9.2.13)$

وهنا يضاف حد الخطأ، على الرغم من أن المتغيرات لم تختلف في النموذجين إلا أننا نلاحظ أن معامل الميل في (8.2.13) هو النموذج الصحيح" أو "الحقيقي" فهل المقدرة α تعتبر مقدرًا غير متحيز لـβ الحقيقية؟ أي هل ستكون $E(\hat{\alpha}) = \beta$؟ إذا لم يتحقق ذلك، يكون هناك خطأ في توصيف حد الخطأ العشوائي، والذي يعتبر أحد مصادر أخطاء التوصيف، وبالتالي لتجميع كل الملاحظات السابقة والخاصة بتكوين نموذج تطبيقي لتوصيف البيانات، فمن الممكن أن يقع الباحث في أحد أخطاء التوصيف التالية (جوجاراتي، 2015 صفحة 654):

1. حذف واحد أو أكثر من المتغيرات المهمة.
2. إضافة واحد أو أكثر من المتغيرات غير المهمة.
3. استخدام شكل دالة خاطئ.
4. أخطاء القياس.
5. توصيف غير سليم لحد الخطأ العشوائي.

قبل البدء في اختيار هذه الأنواع من أخطاء التوصيف بالتفصيل، يجب التفرقة بين أخطاء توصيف النموذج وأخطاء التوصيف غير الصحيح للنموذج. فالأنواع الأربعة الأولى السابقة أساسية في تحديد طبيعة أخطاء توصيف النموذج، حيث إن لدينا بشكل ما تحديدًا للنموذج، الصحيح، ولكن لأسباب ما لا نقدره. في أخطاء التوصيف غير الصحيح للنموذج، لا نعرف طبيعة النموذج الصحيح حتى نأخذه كخطوة مبدئية. وهنا يمكن أن

يسترجع القارئ التعارض بين الكنزيين والنقديين. فالنقديون يعطون أولوية أساسية للمال في تفسير التغير في GDP في حين أن الكينزيين يهتمون أكثر بدور الحكومة في تفسير التغير في GDP، وبالتالي هناك نموذجان متنافسان.

في المثال التالي، سنعتبر مبدئيًا أخطاء توصيف النموذج، ثم اختيار أخطاء التوصيف غير السليم للنموذج.

3-عواقب أخطاء توصيف النموذج: CONSEQUENCES OF MODEL SPECIFICATION ERRORS

بغض النظر عن مصدر أخطاء التوصيف ماهي العواقب؟ لتبسيط المناقشة، ستتم الإجابة على السؤال السابق في إطار نموذج الانحدار ثلاثي المتغيرات، وسنهتم بأول مصدرين للأخطاء التوصيفة السابق مناقشتها وهي بالتحديد: (1) توصيف النموذج بعدد من المتغيرات أقل من العدد الصحيح، أي أن هناك حذفٌ لمتغير أو أكثر من المتغيرات المهمة بالنموذج و(2) توصيف النموذج بعدد من المتغيرات أكثر من العدد الصحيح، أي أن هناك إضافة لمتغير أو أكثر غير مهم بالنموذج. مناقشتنا الحالية من الممكن بسهولة أن يتوسع نطاقها لتشمل أكثر من ثلاثة متغيرات بنموذج الانحدار، ولكنها تحتاج إلى مهارات جبرية، واستخدام جبر المصفوفات يصبح ضرورة ملحة إذا وجد أكثر من ثلاثة متغيرات في نموذج الانحدار.

توصيف النموذج بأقل من الصحيح (حذف متغير مهم): Underfitting a Model (Omitting a Relevant Variable)

افترض أن النموذج الصحيح كالتالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad \dots\dots\dots (1.3.13)$$

ولكن لأسباب ما استخدمنا النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i \dots\dots\dots\dots\dots (2.3.13)$$

عواقب حذف المتغير X_3 تكون كالتالي:

1. إذا كان المتغير المحذوف مرتبطًا مع المتغير الموجود بالنموذج X_2، وكان معامل الارتباط $r_{2\,3}$ بين المتغيرين لا يساوي الصفر، فإن $\hat{\alpha}_1$ و $\hat{\alpha}_2$ ستكون مقدرات متحيزة وغير متسقة أيضًا. وبالتالي يكون $E(\hat{\alpha}_1) \neq \beta_1$ و $E(\hat{\alpha}_2) \neq \beta_2$ وهذا التحيز لا يختفي بزيادة حجم العينة.

2. حتى إذا لم يكن X_2، و X_3، مرتبطين فإن $\hat{\alpha}_1$ سيكون متحيزًا على الرغم من أن $\hat{\alpha}_2$ الآن سيكون غير متحيز.
3. مقياس التباين التقليدي لـ $\left(=\sigma^2/\sum x_{2i}^2\right)\hat{\alpha}_2$ سيكون مقدرًا متحيزًا لتباين المعلمة الحقيقية $\hat{\beta}_2$.
4. من توابع ذلك أيضًا أن تكون فترات الثقة اختبارات الفروض غير سليمة، وتؤدي إلى استنتاجات خاصة بمعنوية المعالم المقدرة خاطئة.
5. أحد العواقب المهمة أيضًا أن التنبؤ بناءً على النموذج غير الصحيح ستكون مضللة وغير سليمة، وأيضًا فترات الثقة المتنبأ بها.

بالطبع إثبات كل نقطة من النقاط السابقة يأخذنا بعيدًا عن إطار هذا الكتاب. إلا أننا استعرضنا في الملحق A13، الفقرة 1.A13 التالي:

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 b_{3\ 2} \quad \ldots\ldots\ldots\ldots (3.3.13)$$

حيث $b_{3\ 2}$ هو ميل الانحدار الخاص بالمتغير المحذوف X_3 على المتغير الموجود في النموذج $\left(b_{3\ 2} = \sum x_{3i}x_{2i}/\sum x_{2i}^2\right)X_2$ كما يتضح من (13.3.3) فإن $\hat{\alpha}_2$ مقدر متحيز إلا إذا كان β_3 أو $b_{3\ 2}$ أو كلاهما يساوي الصفر.

وقد أشرنا من قبل إلى أنه إذا كانت β_3 تساوي الصفر، فإنه لا يوجد في الأساس خطأ توصيف. المعامل $b_{3\ 2}$ سيساوي الصفر إذا كان X_2 وX_3غير مرتبطتين. وهذا نادر الحدوث في أغلب البيانات الاقتصادية.

عموماً مقدار التحيز سيعتمد على حد التحيز $\beta_3 b_{3\ 2}$، فإذا كان على سبيل المثال، β_3 موجبة (أي أن X_3 له تأثير طردي على Y) و $b_{3\ 2}$ موجب (أي أن X_2 و X_3 مرتبطتين طرديًا) فإن $\hat{\alpha}_2$، في المتوسط ستقدر بأعلى من قيمتها الحقيقية (أي تحيز موجب). وهذه النتيجة متوقعة، حيث إن X_2 لا تمثل فقط التأثير المباشر على Y ، ولكن أيضًا التأثير غير المباشر (عن طريق X_3) علىY . باختصار X_2، و يظهر لها دورٌ مؤثرٌ في النموذج من خلال تأثير X_3، والذي لا يظهر له أثرٌ مباشر لأنه غير "مسموح" له بالظهور في النموذج. كمثال تطبيقي دعنا نسترجع المثال السابق مناقشته في الفصل (7).

مثال توضيحي: ILLUSTRATIVE EXAMPLE

وفيات الأطفال (مرة أخرى): Child Mortality Revisited

يعمل انحدار وفيات الأطفال (CM) على GNP للفرد (PGNP) ومعدل الأمية لدى السـيدات (FLR) حصـلنا على نتائج الانحدار الموضـحة في المعادلة (2.6.7). وفي هذه النتائج، نجد معاملات الميل الجزيئية للمتغيرين هي $0.0056-$ و $2.2346-$ بالترتيب. ولكن إذا حذفنا متغير FLR، نحصل على النتائج المعطاة في المعادلة (2.7.7) إذا اعتبرنا أن (2.6.7) هو النموذج الصـحيح، إذًا (2.7.7) يعاني من خطأ التوصـيف، حيث إنه تم حذف متغير مهم وهو FLR. نرى الآن أنه في النموذج الصحيح معامل PGNP كان $0.0056-$ في حين أنه في النموذج "غير السليم " (2.7.7) كان $0.0114-$.

باستخدام القيم المطلقة، نرى أن PGNP الآن له تأثير أكبر على CM مقارنة مع النموذج الصحيح. ولكن إذا قمنا بعمل انحدار لـــــ FLR على PGNP (انحدار للمتغير المحذوف على المتغير الباقي في النموذج). معامل الميل في هذا الانحدار ($b_{3\ 2}$ في المعادلة (3.3.13)) يســـاوي 0.00256 وهذا يعني أنه كلما زاد PGNP بوحدة واحدة فإنه في المتوســـط تزداد FLIR بحوالي 0.00256 وحدة. ولكن إذا زادت FLR بثلاث وحدات فإن تأثير ذلك على CM يكون كالتالي (جوجاراتي، 2015 صفحة 659):

$$(-2.2316)(0.00256) = \hat{\beta}_3 b_{3\ 2} = -0.00543$$

وبالتالي من (3.3.13) لدينا في النهاية:

$$\left(\hat{\beta}_2 + \hat{\beta}_3 b_{2\ 3}\right) = -0.0056 + (-2.2316)(0.00256) \approx -0.0111$$

وهذه نفس القيمة تقريبًا الخاصة بمعامل PGNP التي نحصل عليها من النموذج غير الصحيح (2.7.7).

كما يتضح من هذا المثال، التأثير الحقيقي لــ PGNP على CM أقل بــ $(0.0056-)$ من الأثر المقترح من النموذج غير الصحيح (2.7.7) والمساوي $(0.0114-)$

دعنا الآن نختبر تباينات $\hat{\alpha}_2$ و $\hat{\beta}_2$:

$$var(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots (4.3.13)$$

$$var\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum x_{2i}^2\ (1 - r_{23}^2)} = \frac{\sigma^2}{\sum x_{2i}^2} VIF \quad \ldots\ldots\ldots\ldots\ldots\ldots (5.3.13)$$

حيث VIF (مقياس للتعدد الخطي) هو تباين معامل التضخم ($= 1/(1 - r_{2\ 3}^2)$) والذي سبق وناقشناه في الفصـــل (10) و $r_{2\ 3}^2$ هو معامل الارتباط بين X_2 و X_3. المعادلات (4.3.13) و (5.3.13) ســـبق وتعرضنا لها في الفصلين 3 و7.

بما أن المعادلات (4.3.13) و (5.3.13) غير متطابقة، فإنه في العموم تباين $(\hat{\alpha}_2)$ سيكون مختلفًا عن تباين $(\hat{\beta}_2)$. ولكننا نعلم أن $\text{var}(\hat{\beta}_2)$ غير متحيز (لماذا؟).

وبالتالي فإن $\text{var}(\hat{\alpha}_2)$ متحيز، وهذا يدعم ما سبق وذكرناه في النقطة (4) سابقا.

بما أن $0 < r_{23}^2 < 1$ فإنه $\text{var}(\hat{\alpha}_2) < \text{var}(\hat{\beta}_2)$. وهنا لدينا مشكلة وهي كالتالي: على الرغم من أن $\hat{\alpha}_2$ مقدر متحيز، إلا أن تباينه أقل من تباين المقدر غير المتحيز (بالطبع تم استبعاد الحالة التي يكون فيها $r_{23}^2 = 0$ ، بما أنه في الواقع يكون هناك بعض الارتباط بين المتغيرات المنحدرة) وبالتالي لابد من المفاضلة بين الأمرين السابقين.

الموضوع لم ينته عند هذا الحد، فالقيمة المقدرة لـσ^2 من النموذج (2.3.13) والأخرى المقدرة من النموذج الصحيح (1.3.13) غير متساويين حيث إن RSS للنموذجين وكذلك درجات الحرية الخاصة بهما مختلفة. تذكر أننا حصلنا على تقدير لـσ^2 كالتالي $\hat{\sigma}^2 \neq \text{RSS/df}$ وبالتالي فالتقدير يعتمد على عدد المتغيرات المنحدرة الموجودة في النموذج، بالإضافة إلى درجات الحرية $(= n$، عدد المعالم المقدرة). والآن إذا أضفنا متغيرات جديدة للنموذج، فإن RSS بوجه عام يقل (تذكر أنه كلما زاد عدد المتغيرات المضافة إلى النموذج، فإن R^2 يزداد ولكن درجات الحرية أيضًا تقل، حيث عدد المعالم المقدرة يزيد. المحصلة النهائية تعتمد على ما إذا كان RSS يقل بشكل كفء أم لا مما يعوض مشكلة تقليل درجات الحرية الناتجة عن إضافة متغيرات منحدرة جديدة.

من المحتمل أنه مثلاً إذا كان لمتغير منحدر تأثير كبير على المتغير المنحدر عليه، فإنه يؤدي إلى تقليل RSS أقل من الخسارة في درجات الحرية كنتيجة لإضافة المتغيرات إلى النموذج. إضافة مثل هذا المتغير إلى النموذج لن يؤدي إلى تقليل التحيز فقط، وإنما سيزيد أيضًا من الدقة في التقدير (أي سيقلل الأخطاء المعيارية).

على الجانب الآخر، إذا كان المتغير له تأثير محدود على المتغير المنحدر، وإذا كانوا مرتبطين بدرجة عالية (أي أن VIF كبير يمكننا تقليل التحيز في معاملات المتغيرات الموجودة فعليا في النموذج، ولكن يزداد الخطأ المعياري لهم (أي سيكون أقل كفاءة).

بالطبع المفاضلة في هذه الحالة بين التحيز والدقة تكون مهمة وحيوية. وكما ترى من المناقشة السابقة، هذه المفاضلة ستعتمد على الأهمية النسبية للمتغيرات المنحدرة المختلفة.

يمكن أن نستنتج التالي من المناقشة السابقة: دعنا نعتبر الحالة الخاصة التي يكون فيها $r_{23} = 0$ أي أن X_2 و X_3 غير مرتبطتين. سيؤدي ذلك إلى أن يكون b_{23} يساوي الصفر (لماذا؟). وبالتالي نرى من

(3.3.13) أن $\hat{\alpha}_2$ الآن مقدر غير متحيز. ويبدو أيضًـــا من (4.3.13) و (5.3.13) أن تباينات $\hat{\alpha}_2$ و$\hat{\beta}_2$ متساوية.

هل لا يوجد أي ضـــرر من حذف المتغير X_3 من النموذج حتى وإذا كان مهمًا من الناحية النظرية؟ الإجابة بوجه عام هي لا. ولكن في هذه الحالة كما سبق وذكرنا، فإن $var\ (\hat{\alpha}_2)$ المقدر من (4.3.13) مازال متحيزًا وبالتالي اختبارات الفروض ســتكون نتائجها غير موثوق فيها بالإضـــافة إلى أنه في العديد من الأبحاث الاقتصادية يكون X_2 و X_3 مرتبطين مما يؤدي إلى المشكلة السابق ذكرها.

وبالتالي، فالنقطة المهمة هنا هي: إذا تم تكوين النموذج على أسـاس نظري سـليم، يكون حذف أي متغير من النموذج أمرًا لا ينصح به.

إدخال متغير غير ضروري في النموذج (توفيق النموذج بعدد من المتغيرات أكثر من المطلوب)

Inclusion of an Irrelevant Variable (Overfitting a Model)

دعنا الآن نفترض التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \dots\dots\dots\dots (6.3.13)$$

هو النموذج الصحيح، ولكننا استخدمنا النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \ \dots\dots\dots\dots\dots (7.3.13)$$

وهذا يمثل خطأ توصيف ناتج عن إضافة متغير غير مهم إلى النموذج.

عواقب خطأ التوصيف السابق كالتالي:

1. مقـدرات OLS لمعـالم النموذج "غير الصـــحيح" تكون جميعهـا غير متحيزة ومتســـقـة. أي أن $E(\hat{\alpha}_3) = \beta_3 = 0$ و $E(\hat{\alpha}_2) = \beta_2$ ، $E(\alpha_1) = \beta_1$.
2. تباين الخطأ σ^2 مقدر بشكل سليم.
3. فترات الثقة العادية واختبارات الفروض سليمة.
4. عموما مقدرات αs ستكون غير كف، أي أن تباينها عمومًا أكبر من نظيرها الخاص بـــ $\hat{\beta}'_s$ للنموذج الصحيح.

إثبات بعض هذه النقاط موجود في ملحق A13 الفقرة $2A13$. النقطة المهمة هنا هي الكفاءة النســـبية لـ $\hat{\alpha}'_s$ والتي يمكن توضيحها بسهولة كالتالي:

من الـ OLS التقليدية تعلم أن:

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \dots\dots\dots (8.3.13)$$

و:

$$var(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1-r_{23}^2)} \dots\dots\dots (9.3.13)$$

وبالتالي:

$$\frac{var(\hat{\alpha}_2)}{var(\hat{\beta}_2)} = \frac{1}{1-r_{23}^2}$$

بما أن $0 \leq r_{23}^2 \leq 1$، يكون $var(\hat{\alpha}_2) \geq var(\hat{\beta}_2)$ أي أن تباين $\hat{\alpha}_2$ أكبر من تباين $\hat{\beta}_2$ وإن كان في المتوسط $\hat{\alpha}_2 = \hat{\beta}_2$ (أي أن $E(\hat{\alpha}_2) = \beta_2$) معنى هذه النتيجة أن إدخال متغير غير مهم مثل X_3 إلى النموذج جعل تباين $\hat{\alpha}_2$ أكبر من المفروض، وبالتالي قلت دقة $\hat{\alpha}_2$. وهذا ينطبق أيضا على$\hat{\alpha}_2$.

لاحظ عدم التماثل بين نوعي الخطأ السابق ذكرهما من أخطاء التوصيف. فإذا حذفنا متغيرًا مهمًا وضروريًا، فإن معاملات المتغيرات الباقية في النموذج ستكون بوجه عام متحيزة وغير متسقة، وتباين الخطأ سيقدر بشكل غير سليم، واختبارات الفروض تكون غير سليمة على الجانب الآخر، إضافة متغير غير مهم أو غير ضروري إلى النموذج ستظل التقديرات غير متحيزة ومنسقة، وتباين الخطأ سيقدر بشكل سليم وكل خطوات ونتائج اختبارات الفرض التقليدية تكون سليمة. المشكلة الوحيدة لإدخال نموذج غير مهم إلى النموذج أن التباين المقدر للمعاملات سيكون أكبر من المفروض، وبالتالي كنتيجة لذلك استبدلنا الإحصائي عن المعالم سيكون له دقة أقل (جوجاراتي، 2015 صفحة 663).

وليس من المطلوب الآن أن تستنتج عزيزي القارئ أنه من الأفضل إضافة متغير غير مهم إلى النموذج عن حذف متغير مهم، حيث إن إضافة هذا المتغير غير المهم ستؤدي إلى خسارة في كفاءة المقدرات، وقد تؤدي إلى مشكلة التعدد الخطي (لماذا؟) بالإضافة إلى مشكلة تقليل درجات الحرية وبالتالي:

عمومًا، الأسلوب الأفضل هو إضافة المتغيرات المفسرة إلى النموذج على أساس نظري، وتكون هذه المتغيرات مؤشرة بشكل مباشر على المتغير التابع وتأثيرها لا يكفي تضمينه في المتغيرات الأخرى الموجودة فعلاً في النموذج.

4.13 اختبارات لأخطاء التوصيف: TESTS OF SPECIFICATION ERRORS

معرفة عواقب أخطاء التوصيف تعتبر موضوعًا قائمًا بذاته ومختلفًا عن اكتشاف خطأ التوصيف. في الأساس غالبًا ما يظهر تحيز التوصيف بدون قصد أو بدون عمد. ويكون بسبب رغبتنا في تكوين نموذج دقيق

بأكبر قدر ممكن معتمدين على أن النظرية ضـــعيفة أو ليس لدينا البيانات الكافية لاختبار مثل هذا النموذج. وكما لاحظ Davidson نظرا للطبيعة غير القابلة للتجربة للاقتصـاد، لا نسـتطيع الجزم بكيفية تجميع البيانات. فاختبار أي فرض في الاقتصـــاد يعتمد في الأســـاس على المزيد من الفروض التي تجعلنا قادرين على تكوين نموذج مناسب مع العلم بأن هذه الفروض قد تكون منطقية أو لا.

السؤال العملي الآن ليس عن تحليل وجود أخطاء التوصـيف، ولكن الأهم هو كيف يمكن اكتشـافها. فمجرد التأكد من حدوث خطأ في التوصـــيف، فإن طرق العلاج يمكن اتباعها لتجنب العواقب. فمثلاً العلاج الواضـــح لمشكلة خطأ التوصـيف الناتجة عن حذف متغير مهم من النموذج هو إضـافة هذا المتغير إلى النموذج مباشـرة بعد التأكد من وجود بيانات خاصة به.

في الفقرة التالية، سنناقش بعض الاختبارات التي يمكن استخدامها لاكتشاف وجود خطأ التوصيف.

اكتشاف وجود متغير غير مهم في النموذج (توفيق النموذج بعدد من المتغيرات أكثر من المطلوب)

Detecting The Presence of Unnecessary Variables (Overfitting a Model)

افترض أن لدينا نموذجًا له متغير لتفسير ظاهرة ما كالتالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \dots\dots\dots (1.4.13)$$

عمومًا، نحن غير متأكدين من أن مثلاً، المتغير X_k فعلاً يجب أن يتواجد في هذا النموذج. أحد الطرق البسيطة لاكتشاف ذلك هو اختبار معنوية القيمة المقدرة لـــ β_k باختبار t التقليدي $t = \hat{\beta}_k / se(\hat{\beta}_k)$. ولكن افترض أننا غير متأكدين من أهمية وجود X_3 و X_4 معًا في النموذج. يمكن اختبار ذلك مباشـــرة باختبار F السابق مناقشته في الفصل (8). وبالتالي اكتشاف وجود متغير غير مهم في النموذج ليس عملية صعبة.

ولكن من المهم تذكر أنه حتى يمكنك القيام بمثل هذه الاختبارات، يجب أن يكون هناك في الأساس نموذج للتعامل معه، فنحن نقبل هذا النموذج أولاً ثم نتعامل معه. وبالتالي، فإنه وفقا لهذا النموذج يمكن تحديد ما إذا كان واحدًا أو أكثر من المتغيرات المنحدرة غير مهم في النموذج باستخدام اختبارات t و F التقليدية.

ولكن لاحظ أننا لا نستخدم اختبارات t و F لبناء النموذج، فلا يجب أن نستنتج أن Y مرتبطة بـــ X_2 مثلاً لأن $\hat{\beta}_3$ لها معنوية إحصائية ثم تبدأ في توسيع النطاق وإدخال X_3 للنموذج وإبقائه فيه إذا كانت $\hat{\beta}_3$ لها معنوية إحصائية وهكذا (جوجاراتي، 2015 صفحة 667).

فهذه الطريقة في بناء النموذج تسـمى أسـلوب الأعلى–الأسـفل (حيث نبدأ بنموذج أصـغر ويتسـع نطاقه تدريجيًا) أو تسمى أيضًا التنقيب عن البيانات (مسميات أخرى مثل انحدار الأسماك، تجميع البيانات).

الهدف الأساس للتنقيب عن البيانات هو تكوين النموذج "الأفضل" بعد عمل العديد من اختبار التشخيص، وبالتالي يكون النموذج النهائي المختار هو نموذج "جيد" بمعنى أن كل المعاملات المقدرة لها الإشارة "الصحيحة" ولها معنوية إحصـــائية بناء على اختبارات t و F، و، وقيمة R^2 تكون كبيرة بدرجة معقولة و Durbin Watsond لها قيمة مقبولة (قريبة من 2) وهكذا. أما الجانب الفني الخاص بتطبيق تنقيب البيانات فنجده في مقولة william pool التالية "... الاعتماد على التطبيق العملي بدلاً من تطبيق نظرية اقتصادية سيكون دائما خطرًا". أحد أسباب إدانة تنقيب البيانات كالتالي:

مستوى المعنوية الحقيقي والأسمى في حالة استخدام تنقيب البيانات

Nominal versus True Level of Significance in the Presence of Data Mining

خطر اســـتخدام تنقيب البيانات والذي قد يواجه الباحث هو مســـتوى المعنوية التقليدي (α) مثل 1، 5 أو 10% فلن يكون هو مستوى المعنوية الحقيقي. وقد اقترح Lovell أنه إذا كان هناك من المتغيرات المنحدرة وتم اختيار k كعدد نهائي ($k \leq c$) بناءً على أسلوب تنقيب البيانات، فإن مستوى المعنوية الحقيقي (α^*) مرتبط بمستوى المعنوية الاسمي (α) كالتالي:

$$\alpha^* = 1 - (1 - \alpha)^{c/k} \quad \ldots\ldots (2.4.13)$$

أو تقريبًا كالتالي: $\alpha^* = (c/k)\alpha$

فمثلاً إذا كان $c = 15$، $k = 5$ و $\alpha = 5\%$ فمن (3.4.13) يكون مســـتوى المعنوية الحقيقي هو $15\% = (5)\ (15/5)$. وبالتالي إذا اســـتخدم الباحث تنقيب البيانات واختار 5 من 15 متغير منحدر واستعرض نتائجه على أساس أنه استخدم مستوى معنوية 5%، واستعرض بعد ذلك نتائج المعنوية الإحصائية، يجب علينا أن ندرك أن هذه النتائج غير سليمة بالمرة، حيث إننا نعلم أن مستوى المعنوية (الحقيقي) هو 15%. ويجب ملاحظة أنه إذا كان $c = k$ فلا يوجد تنقيب للبيانات، ومستوى المعنوية الحقيقي والاسمي متساويان. بالطبع في الواقع معظم الباحثين يسجلون نتائجهم الخاصة بالانحدار " النهائي". بدون توضيح استخدم أسلوبًا لتنقيب البيانات أو الخطوات المختلفة التي قام بها للوصول إلى هذا النموذج.

وعلى الرغم من العيوب الواضحة لهذه الطريقة، فإن هناك زيادة عملية تم استخدامها خصوصًا في الاقتصاد القياسي التطبيقي. وبالتالي يصبح أسلوب الوصول إلى نموذج (بدون استخدام تنقيب البيانات من الصعب الحفاظ عليه.

وكما لاحظ Zaman: للأسف التجارب المختلفة بيانات حقيقية أوضحت أن (الأسلوب الذي لا يعتمد على تنقيب البيانات) غير متاح وغير مرغوب فيه أيضًا. غير متاح لأنه من النادر في نظرية الاقتصاد أن تجد ما يشير إلى نموذج فريد لتوصيف البيانات. وغير مرغوب فيه، لأن بعض المفاهيم التي يتعلم الباحث من البيانات لاستخدام أو عدم استخدام نماذج معينة لا يمكن الاعتماد عليها في هذا الأسلوب، حتى ولو على سبيل الصدفة. كان النموذج المبدئي نموذجًا جيدًا. يكون من المفيد البحث عن أنواع من النماذج الأخرى التي قد تنفق أو تختلف مع البيانات محل الدراسة.

وقد اتفق Kery Patlerson مع ذلك ووضح التالي: أسلوب (تنقيب البيانات) يقترح أن يتم الفصل بين النظرية الاقتصادية من جهة، وبين التطبيق العلمي للبيانات من جهة أخرى.

وبدلاً من الاستمرار في المقارنة بين الأسلوب التقليدي الذي (لا يوجد فيه تنقيب للبيانات) وأسلوب التنقيب في البيانات، يمكن للباحث أن يستخدم وجهة نظر Peter Kennedy والتي تنص على التالي: (توصيف النموذج) يجب أن يتم وفقا للمزج بين النظرية والبيانات، والاختبارات المختلفة المستخدمة للتوصيف، يجب أن تستخدم لتقليل تكلفة تنقيب البيانات. أمثلة على تلك الأساليب هي: تجنب البيانات التي لا ينطبق عليها اختبارات التنبؤ، تعديل مستويات المعنوية [a la lovell] وتجنب الطرق المشكوك فيها مثل تكبير R^2.

إذا اقتصرنا النظر لتنقيب البيانات على أنه أسلوب يعتمد على بيانات عملية قد يكون بها أخطاء، وقد يكون بها فائدة كبيرة لتوصيف النموذج يكون تنقيب البيانات في ذلك الحين أسلوبًا فعالاً جدًا وينصح باستخدامه. وفقا لـ kennedy مرة أخرى، "فن الاقتصاد القياسي التطبيقي يعتمد على اختيار البيانات وفقا للنظرية، ومحاولة تفادي الأخطار المحتملة للتنقيب في البيانات" (جوجاراتي، 2015 صفحة 669).

اختبارات للمتغيرات المحذوفة وشكل الدالة غير الصحيح: Tests for Omitted Variables and Incorrect Functional Form

في الواقع لا نستطيع الجزم بأن النموذج المختار وفقًا لاختبارات تطبيقه هو "النموذج السليم ولا شيء غير ذلكّ. فبناء على النظرية والجانب التطبيقي السابق في نفس المجال، نصل إلى نموذج نعتقد أنه يتناول كافة جوانب الموضوع محل الدراسة، ثم نطبق هذا النموذج باستخدام اختبارات تطبيقية. بعد أن نحصل على النتائج، نبدأ في تحليل البيانات واضعين في الاعتبار الأساليب التي تم بناءً على اختيار النموذج الجيد والسابق ذكرها مباشرة عند هذه المرحلة، نحن نحاول معرفة درجة دقة النموذج المختار، لتحديد دقة النموذج نستعرض بعض النتائج مثل قيمة R^2، قيم t المقدرة، إشارات المعاملات المقدرة وعلاقتها بتوقعاتنا السابقة، إحصاء Durbin-Watson صغير جدًا، هنا يجب أن نقلق على جودة ودقة النموذج ونفكر في أساليب علاجية: فقد نكون

حذفنا متغيرًا مهمًا، أو استخدمنا شكل دالة غير سليم، أو لم تستخدم الفروق وكانت لدينا بيانات سلسلة زمنية لإزالة أثر الارتباط التسلسلي) وهكذا. لمعرفة ما إذا كان النموذج المختار يعاني من واحد أو أكثر من هذه المشاكل، يمكن أن نستخدم بعض الأساليب التالية.

اختبار البواقي: Examination of Residuals

كما سبق وأشرنا في الفصل (12)، اختبار البواقي يعتبر أداة تشخيص بصرية مهمة لاكتشاف الارتباط الذاتي أو اختلاف التباين، ولكن هذه البواقي يمكن اختبارها أيضًا، خصوصًا في البيانات المقطعية، وبالنسبة لأخطاء التوصيف المتعلقة بحذف أحد المتغيرات المهمة أو استخدام شكل دالي غير سليم. إذا كان هناك أحد هذه الأخطاء، فإن الرسم البياني الخاص بالبواقي سيكون له نمطٌ محدد.

للتوضيح، دعنا نعتبر دالة التكلفة التكعيبية والتي سبق وناقشناها في الفصل (7).

افترض أن دالة التكلفة الكلية الحقيقية كالتالي حيث Y = الكلية التكلفة و الناتج = x:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_3 X_i^3 + u_i \dots\dots.. (4.4.13)$$

ولكن إذا قام البحث باستخدام الدالة التربيعية التالية:

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_i^2 \dots\dots\dots (5.4.13)$$

وإذا افترضنا أن باحثا آخر اختار الدالة الخطية التالية:

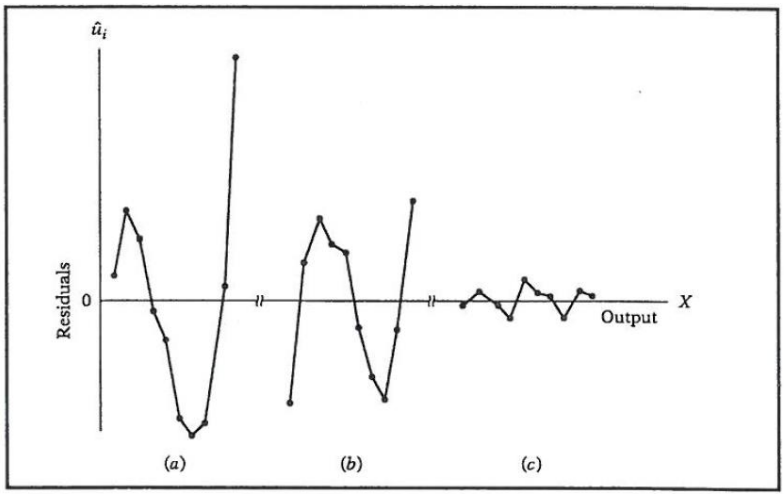
$$Y_i + \lambda_1 + \lambda_2 X_i + u_{3i} \dots\dots\dots\dots\dots (6.4.13)$$

على الرغم من أننا نعلم أن كلاً من الباحثين السابقين قد وقعا في خطأ التوصيف، فدعنا نستعرض شكل البواقي المقدرة في كل من النماذج الثلاثة السابقة. [بيانات التكلفة – النتائج معطاة في جدول (47)]. شكل (1.13) يتضح منه التالي: كلما انتقلنا من اليسار إلى اليمين، أي نقترب من الحقيقة، ليس فقط تقل البواقي (بصورة مطلقة) ولكن أيضًا يختفي الشكل أو النمط الدائري المصاحب للنماذج غير الصّحيحة.

وبالتالي فائدة استخدام شكل البواقي واضحة تمامًا. فإذا كان هناك خطأ توصيف، فإن البواقي سيظهر فيها نمط محددً وواضح.

اختبار d لـ Durbin–Watson مرة أخرى The Durbin-Watson d Statistic Once Again

إذا اخترنا إحصاء Durbin Watson d المحسوب من جدول (1.13) نرى أنه يساوي 0.716 بالنسبة لدالة التكلفة الخطية، ممَّا يعني أن هناك "ارتباط طردي في البواقي المقدرة: بالنسبة لـ $n = 10$ و $k = 1$ فإن الـ 5% قيمة حرجة لـ d هي $d_L = 0.879$ و $d_u = 1.320$



شكل (1.13) البواقي $\hat{u}_i$ محسوبة بناء على (a) دالة تكلفة خطية، (b) دالة تكلفة تربيعية و(c) دالة تكلفة تكعيبية

جدول (113) البواقي المقدرة من دالة التكلفة الكلية الخطية، التربيعية والتكعيبية

Observation Number	$\hat{u}_{in}$ *Inear Model**	$\hat{u}_{in}$ *Quadratic Model*†	$\hat{u}_{in}$ *Cubic Model***
1	6.600	−23.900	−0.222
2	19.667	9.500	1.607
3	13.733	18.817	−0.915
4	−2.200	13.050	−4.426
5	−9.133	11.200	4.435
6	−26.067	−5.733	1.032
7	−32.000	−16.750	0.726
8	−28.933	−23.850	−4.119
9	4.133	−6.033	1.859
10	54.200	23.700	0.022
$\star\ \hat{Y}_i = 166.467 + 19.933\,X_i$ (19.021) (3.066) (8.752) + (6.502)			$R^2 = 0.8409$ $\bar{R}^2 = 0.8210$ $d = 0.716$
$\dagger\ \hat{Y}_i = 222.383 - 8.0250\,X_i + 2.542X_i^2$ (23.488) (9.809) (0.869) (9.468) (−0.818) (2.925)			$R^2 = 0.9284$ $\bar{R}^2 = 0.9079$ $d = 1.038$

$^{**}\widehat{Y}_i = 141.767 + 63.478\,X_i + 12.962_i^2 + 0.939X_i^3$ $(6.375) \quad (4.778) \quad (0.9856) \quad (0.0592)$ $(22.238) \quad (13.285) \quad (-13.151) \quad (15.861)$	$R^2 = 0.9983$ $\bar{R}^2 = 0.99975$ $d = 2.70$

وبالمثل فإن قيمة d المحسوبة بناء على دالة التكلفة التربيعية هي 1.038 وتكون الـ 5% قيم حرجة هي $d_L = 0.697$ و $d_U = 1.641$ مما يعني عدم القدرة على اتخاذ قرار في هذه الحالة. ولكن إذا استخدمنا اختبار d المعدل (انظر الفصل 12) سنجد أن هناك ارتباط طرديٌ في البواقي. حيث إن قيمة d، المحسوبة أقل من d_U. أما بالنسبة لدالة التكلفة التكعيبية، التوصيف الصحيح، فإن قيمة d_U المقدرة لا تشير إلى أي "ارتباط " طردي في البواقي[1].

"الارتباط" الموجب المشاهد في البواقي، عندما نستخدم نموذجًا خطيًا أو تربيعيًا ليس دليلاً على وجود ارتباط ذاتي (من الدرجة الأولى) في البواقي ولكن دليل على وجود خطأ في توصيف (النموذج). الارتباط المشاهد ببساطة يعبر عن المتغيرات المهمة التي لم يتم استخدامها في النموذج، وتوجد في الخطأ، ويجب فصلها من حد الخطأ وتقديمها كمتغيرات مفسرة في النموذج: فمثلاً إذا حذفنا X_i^3 من دالة التكلفة كما في (3.2.13)، فإن حد الخطأ في النموذج غير الصحيح (2.2.13) هو في الواقع يساوي $(u_{1i} + \beta_4 X_i^3)$ وسيظهر فيها نمط محدد (أي الارتباط الذاتي الموجب) وذلك دليل على أن X_i^3 تؤثر معنويًا على الـ Y.

لاستخدام اختبار Durbin Watson لاكتشاف خطأ التوصيف نقوم بالتالي:

1. من النموذج المفترض، نحصل على بواقي الـ OLS.
2. إذا اعتقدنا أن النموذج يعاني من خطأ التوصيف، حيث تم حذف متغير مفسر مثلاً Z من النموذج. لاحظ أن المتغير Z قد يكون أحد متغيرات X المتضمنة في النموذج المفترض، أو قد تكون دالة ما في هذا المتغير مثل X_2 أو X_3
3. احسب الإحصاء d من البواقي بناءً على المعادلة التقليدية التالية:

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

لاحظ أن: الترميز t يشير إلى المشاهدة وليس بالضرورة يعني أن البيانات تعتبر سلسلة زمنية.

[1] - في السياق الحالي، قيمة $d = 2$ تعني عدم وجود خطأ في التوصيف. (لماذا؟)

4. من جداول Durbin–Watson، إذا كانت قيمة d المقدرة معنوية، فإنه من الممكن قبول الفرض الخاص بوجود خطأ توصيف في النموذج، وإذا حدث ذلك فعلاً، بالتالي الطرق العلاجية تقترح نفسها بوضوح كالتالي:

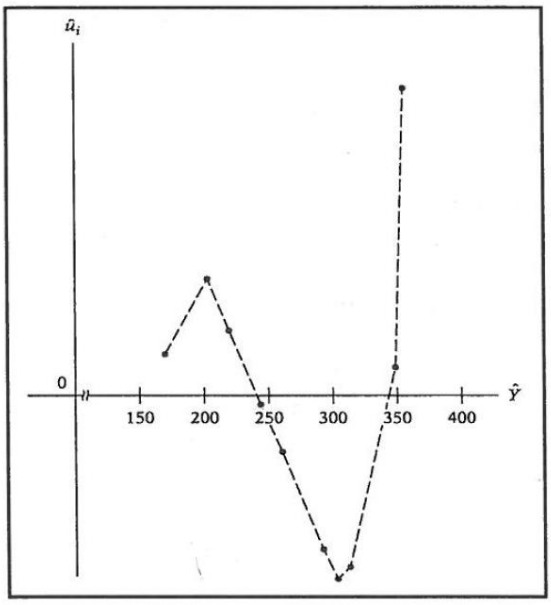
في مثالنا الحالي، المتغير $Z(= X)$(الناتج) كان مرتبًا بالفعل[2] وبالتالي، لا توجد ضرورة الحساب إحصاء d مجددًا. فقد رأينا أن إحصاء d لكل من دوال التكلفة الخطية والتربيعية يقترح وجود أخطاء في التوصيف. العلاج في هذه الحالة واضح ويكون كالتالي: قم بإدخال القيم التربيعية والتكعيبية في دالة التكلفة الخطية، وأدخل القيم التكعيبية على دالة التكلفة التربيعية. أو باختصار استخدم نموذج التكلفة التكعيبية.

اختبار RESET Test لـ Ramsey:

اقترح Ramsey اختبارًا عامًا لخطأ التوصيف وسماه RESET (اختبار خطأ التوصيف للانحدار (Regression Specification error test) سنستعرض هنا الشكل الأبسط لهذا الاختبار لتوضيح الفكرة دعنا نستخدم مرة أخرى مثال التكلفة – الناتج، ونفترض أن دالة التكلفة. هي دالة خطية في الناتج كالتالي:

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \quad \dots\dots\dots\dots\dots (6.4.13)$$

حيث $Y =$ الكلية التكلفة و x = الناتج . الآن إذا رسمنا البواقي $\hat{u}_i$ التي حصلنا عليها من هذا الانحدار ضد $\hat{Y}_i$ القيمة المقدرة لـ Y_i من النموذج، سنحصل على الشكل الموجود في (2.13).



شكل (2.13) البواقي $\hat{u}_i$ و Y المقدرة من دالة التكلفة الخطية: $Y_i = \lambda_1 + \lambda_2 X_i + u_i$

[2] - لا يهم ما إذا قمنا بترتيب $\hat{u}_i$، بناء على X_i^2 أو X_i^3 حيث إن هذه الدوال في x مرتبة بالفعل.

على الرغم أن $\sum \hat{u}_i \hat{Y}_i$ و $\sum \hat{u}_i$ بالضرورة يساويان الصفر (لماذا؟ انظر الفصل 3)، إلا أن البواقي تظهر في علاقتها بـ $\hat{Y}_i$ نمط محدّد في الشكل البياني الخاص بهما، ممَّا يعني أننا إذا استطعنا استخدام $\hat{Y}_i$ في شكل ما كمتغيرات محدّدة في (6.4.13) فإن ذلك سيؤدي إلى زيادة R^2، وإذا كانت هذه الزيادة معنوية إحصائيًا (على أساس اختبار F السابق مناقشته في الفصل 8) فإن ذلك يعني أنه كان من الخطأ استخدام دالة التكلفة الخطية (6.4.13) وهذه هي الفكرة وراء استخدام $RESET$ خطوات الـ $RESET$ تكون كالتالي:

1. من النموذج المختار مثل (6.4.13) نحصل على رقم Y_i المقدرة أي $\hat{Y}_i$.
2. نعيد عمل (6.4.13) مع تقديم $\hat{Y}_i$ للنموذج في بعض الصور الدالية المختلفة كمتغير أو متغيرات منحدرة من الشكل (2.13) نرى أن هناك علاقة غير خطية بين $\hat{u}_i$ و $\hat{Y}_i$ ممَّا يعني إمكانية استخدام $\hat{Y}_i^2$ و $\hat{Y}_i^3$ كمتغيرات منحدرة إضافية. وبالتالي نتقوم بعمل الانحدار التالي:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + u_i \dots\dots (7.4.13)$$

3. ارمز R^2 التي حصلت عليها من (7.4.13) بـ R^2_{new} والتي حصلت عليها من (6.4.13) ارمز لها بـ R^2_{old} . وبالتالي تستطيع استخدام اختبار F الذي سبق وقدمناه في (18.5.8) وهو كالتالي :

$$F = \frac{(R^2_{new} - R^2_{old})/\text{عدد المتغيرات المنحدرة}}{(1 - R^2_{new})/\left(n - \text{عدد المعالم في النموذج الجديد}\right)} \dots\dots (18.5.8)$$

ونرى ما إذا كانت الزيادة في R^2 باستخدام (7.4.13) لها معنوية إحصائية.

4. إذا كانت قيمة F المحسوبة معنوية، مثلاً عند مستوى المعنوية 5% يمكن قبول الفرض القائل بأن النموذج (7.4.13) نموذج غير صحيح.

بالعودة إلى مثالنا التوضيحي، لدينا النتائج التالية (الأخطاء المعيارية موجودة بين الأقواس):

$$\underset{(19.021)}{\hat{Y}_i = 166.467} + \underset{(3.066)}{19.933 X_i} \qquad R^2 = 0.8409 \quad \dots\dots (8.4.13)$$

$$\underset{(132.0044)}{\hat{Y}_i = 2140.7223} + \underset{(33.3951)}{476.6557 X_i} - \underset{(0.00620)}{0.09187 \hat{Y}_i^2} + \underset{(0.0000074)}{0.000119 \hat{Y}_i^3} \qquad (9.4.13)$$

$$R^2 = 0.9983$$

لاحظ أن: $\hat{Y}_i^2$ و $\hat{Y}_i^3$ في (9.4.13) تم الحصول من (8.4.13) والآن بتطبيق اختبار F نحصل على:

$$F = \frac{(0.9983 - 0.8409)/2}{(1 - 0.9983)/(10 - 4)} = 284.4035 \dots\dots (10.4.13)$$

يمكن للقارئ بسـهولة أن يثبت أن هذه القيمة لـــــ F لها معنوية إحصـائية عالية، ممَّا يعني أن النموذج (8.4.13) نموذج غير صـحيح. بالطبع سـنصـل إلى نفس الاسـتنتاج مسـتخدمين الاختبار البصـري للبواقي أو قيمة إحصاء d لـ Durbin–Watson.

إحدى مميزات اسـتخدام RESET هي سـهولة التطبيق، حيث إنها لا تتطلب تحديد النموذج البديل. ولكنها في نفس الوقت تعتبر نقطة ضـعف في هذه الطريقة، حيث إن معرفة النموذج غير صـحيح فقط لا يسـاعدنا في اختيار بديل أفضل.

اختبار مضـــروب lagrange (LM) لإضـــافة المتغيرات: Lagrange Multiplier (LM) Test for Adding Variables

هذا الاختبار يعتبر بديلاً لاختبار RESET لــ Ramsey. لتوضيح هذا الاختبار، سنستعرضه وفقا لمثالنا السابق التوضيحي.

إذا قارنا بين دالة التكلفة الخطية (6.4.13) ودالة التكلفة التكعيبية (4.13.4)، الأولى تعتبر صـــورة مقيدة من الأخيرة (تذكر ما سبق وناقشناه بخصوص المربعات الصغرى المقيدة في الفصل 8).

الانحدار المقيد (6.4.13) يفترض أن معامل الحدود المربعة والمكعبة من الناتج تسـاوي الصـفر. لاختبار ذلك باستخدام أسلوب LM نقوم بالخطوات التالية:

1. قدر الانحدار المقيد (6.4.13) باستخدام OLS واحصل على البواقي $\hat{u}_i$.
2. إذا كان في الحقيقة الانحدار المقيد (4.4.13) هو النموذج الصـحيح، فإن البواقي التي حصـلنا عليها من (6.4.13) ستكون مرتبطة مع الحدود المربعة والمكعبة للناتج أي X_i^2 و X_i^3
3. وذلك يعني أن نقوم بعمل انحدار لـــ $\hat{u}_i$ التي حصلنا عليها من الخطوة 1 على كل المتغيرات المنحدرة (مشتملة على المتغيرات الموجودة في النموذج المقيد) وذلك في مثالنا الحالي.

يعني التالي (جوجاراتي، 2015 صفحة 670):

$$\hat{u}_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + \alpha_4 X_i^3 + v_i \dots\dots\dots\dots (11.4.13)$$

حيث v هو حد الخطأ ومتوافر فيه الخصائص التقليدية المعروفة.

4. في حالة إحكام العينات الكبيرة، أثبت Engle أن n (حجم العينة) مضـــروب في قيمة R^2 المقدرة من الانحدار (المساعد) (11.4.13) تتبع توزيع كلي تربيعي بدرجة حرية تساوي عدد القيود المفروضة في

نموذج الانحدار المقيد، وتلك الدرجة تســـاوي أثنين في مثالنا الحالي، حيث إن القيمة X_i^2 و X_i^3 تم إسقاطهما من النموذج.

رمزيًا لدينا التالي:

$$nR^2 \widetilde{asy} X^2_{(\text{number of restrictions})} \dots\dots\dots\dots\dots (12.4.13)$$

حيث إن asy تعني تقاربيًا، أي في الحالات التي يكون فيها حجم العينة كبيرًا.

5. إذا زادت قيمة chi-square التي حصـــلت عليها من (12.4.13) من القيمة الحرجة عند مســـتوى المعنوية المختار نرفض الانحدار المقيد، وبخلاف ذلك نقبله.

نتائج الانحدار في مثالنا الحالي كالتالي:

$$\hat{Y}_i = 166.467 + 19.333X_i \dots\dots\dots (13.4.13)$$

حيث Y هي التكلفة الكلية و X هي الناتج. الأخطاء القياسية لهذا الانحدار معطاة في جدول (1.13).

إذا قمنا بعمل انحدار البواقي (13.4.13) كما هو مقترح في الخطوة 3. نحصل على النتائج التالية:

$$\begin{aligned} \hat{u}_i &= -24.7 + 43.5443X_i - 12.9615\,X_i^2 + 0.9396\,X_i^3 \\ se &= (6.375) \quad (4.779) \quad (0.986) \quad (0.059) \qquad (14.4.13) \\ R^2 &= 0.9896 \end{aligned}$$

على الرغم من أن حجم العينة يســاوي 10 وهو ليس بالحجم الكبير. إلا أننا نرغب في توضــيح آلية LM كالتالي: ســنحصــل على $nR^2 = (10)(0.9896) = 9.896$ عن جدول كاي التربيعي، وباســتخدام 2 درجات حرية، و1% مستوى معنوية، نجد أن قيمة كاي التربيعية تساوي تقريباً 9.21. وبالتالي القيمة المشاهدة والمساوية لـــ 9.896 لها معنوية إحصائية عند مستوى المعنوية 1% ونستنتج أنه يجب رفض النموذج المقيد (أي دالة التكلفة الخطية). ونلاحظ أننا قد توصلنا إلى نفس النتيجة باستخدام اختبار RESET لـ Ramsey.

5.13 أخطاء القياس: ERRORS OF MEASUREMENT

في كل ما ســبق، تم الافتراض صــراحةً بأن المتغير التابع Y، والمتغيرات المفســرة $X's$ تم قياســها بدون أخطاء. وبالتالي في انحدار نفقات الاستهلاك على الدخل وثروة رب الأسرة، افترضنا أن بيانات هذه المتغيرات "دقيقة"، وبالتالي لا مجال للتخمين أو التدثير أو التقريب بأي شكل منتظم مثل التقريب إلى مئة دولار وهكذا.

للأسف هذا الوضـع النموذجي لا يحدث كثيرًا في الواقع العملي، وذلك وفقًا للعديد من الأسباب مثل أخطاء عدم الاســتجابة، وأخطاء التســجيل، وأخطاء الحســاب. وبغض النظر عن الســبب فإن أخطاء القياس مشــكلة

محتملة عن التطبيق العملي، وهي تعتبر أحد مصادر تحيز التوصيف. وسنستعرض فيما يلي العواقب المحتملة لذلك:

أخطاء قياس في المتغير التابع Y: Errors of Measurement in the Dependent Variable Y

اعتبر النموذج التالي:

$$Y_i^* = \alpha + \beta X_i + u_i \quad \dots\dots\dots\dots (1.5.13)$$

حيث ل Y_i^* : نفقات الاستهلاك الدائمة.

X_i: الدخل الحالي خطأ عشوائي

u_i: خطأ عشوائي

بما أن Y_i^* لا يمكن قياسها مباشرة، نستخدم النفقات المشاهدة (المتغير Y_i) كالتالي:

$$Y_i = Y_i^* + \varepsilon_i \quad \dots\dots\dots (2.5.13)$$

حيث ε_i ترمز للخطأ في قياس Y_i^*. وبالتالي بدلاً من تقدير (1.5.13)، فإننا نقدر:

$$Y_i = (\alpha + \beta X_i + u_i) + \varepsilon_i = \alpha + \beta X_i + (u_i + \varepsilon_i) = \alpha + \beta X_i + v_i \dots\dots (3.5.13)$$

حيث $v_i = u_i + \varepsilon_i$ وهو يعتبر حد خطأ مركب، يشتمل على خطأ المجتمع (والذي يمكن تسميته معادلة حد الخطأ) وحد الخطأ في القياس.

للتبسيط افترض أن $E(u_i) - E(\varepsilon_i) = 0$, $cov(X_i, \varepsilon_i) = 0$ (والذي يعتبر فرضًا من فروض الانحدار الخطي التقليدي و $cov(X_i, \varepsilon_i) = 0$ أي أن أخطاء القياس في Y_i^* غير مرتبطة مع X_i و $cov(X_i, \varepsilon_i) = 0$ أي أن خطأ المعادلة وخطأ القياس غير مرتبطين. بالإضافة إلى بعض الفروض الخاصة بـ β المقدرة سواء من (1.5.13) أو (3.5.13) والتي تنص على أنها مقدر غير متحيز للمعلمة الحقيقيةβ (انظر تمرين 7.13). وبالتالي أخطاء القياس الموجودة في المتغير التابع Y لا تؤثر على خاصية عدم التحيز لمقدرات OLS. عمومًا التباين والأخطاء القياسية لـβ المقدرة من (1.5.13) و (3.5.13) ستكون مختلفة، باستخدام المعادلات المعتادة (انظر الفصل 3) نحصل على:

$$\text{نموذج (1.5.13)} : \quad var(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2} \quad \dots\dots\dots (4.5.13)$$

$$\text{نموذج (3.5.13)} : \quad var(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2} = \frac{\sigma_u^2 + \sigma_\varepsilon^2}{\sum x_i^2} \quad \dots\dots\dots (5.5.13)$$

نلاحظ أن التباين الأخير أكبر من التباين الأول[3]. وبالتالي على الرغم من وجود أخطاء القياس في المتغير التابع Y فإن مقدرات المعالم والتباينات مازالت غير متحيزة، إلا أن التباين المقدر أكبر من نظيره في الحالة التي لا يوجد فيها مثل هذا الخطأ من أخطاء القياس.

أخطاء القياس في المتغير المفسر X: Errors of Measurement in the Explanatory Variable X

الآن افترض أنه بدلاً من (1.5.13) لدينا النموذج التالي:

$$Y_i = \alpha + \beta X_i^* + u_i$$

حيث: Y_i نفقات الاستهلاك الحالية.

X_i^*: الدخل الدائم

u_i: حد الخطأ العشوائي (معادلة الخطأ)

وافترض أنه بدلاً من مشاهدة X_i^*، نشاهد التالي:

$$Y_i = X_i^* + w_i \dots\dots\dots (7.5.13)$$

حيث w_i تمثل خطأ في قياس X_i^*. وبالتالي بدلاً من تقدير (6.5.13) نقدر التالي:

$$Y_i = \alpha + \beta(X_i - w_i) + u_i = \alpha + \beta X_i + (u_i + \beta w_i) = \alpha + \beta X_i + z_i \dots (8.5.13)$$

حيث $z_i = u_i - \beta w_i$، وبالتالي لدينا مزيج من معادلة الأخطاء وأخطاء القياس.

والآن حتى لو افترضنا أن w_i لها وسط يساوي الصفر، وغير مرتبطة تسلسليًا، وغير مرتبطة بالـــu_i، لا نستطيع افتراض أن حد الخطأ المركب z_i مستقل عن المتغير المفسر x_i، وذلك بسبب التالي (افترض أن $E(z_i) = 0$):

$$cov(z_i, X_i) = E[z_i - E(z_i)][X_i - E(X_i)] = E(u_i - \beta w_i)(w_i) \quad using\ (13.5.7)$$
$$= E(-\beta w_i^2) = -\beta\sigma_w^2 \dots\dots\dots\dots\dots (9.5.13)$$

وبالتالي المتغير المفسر وجد الخطأ في (8.5.13) مرتبطان، وذلك يخالف فرضًا مهمًا في نموذج الانحدار الخطي التقليدي، وهو الفرض القائل بأن المتغير المفسر غير مرتبط مع – حد الخطأ العشوائي. إذا خالفنا هذا

[3] - لكن نلاحظ أن هذا التباين مازال غير متحيز، حيث إنه وفقا للشروط الموضوعة الخطأ المركب $v_i = u_i + \varepsilon_i$ مازال مستوفيا لكل الفروض الخاصة بطريقة المربعات الصغرى.

الفرض، فإنه يمكن إثبات أن مقدرات OLS لم تعد فقط متحيزة، ولكن غير منسقة أيضًا، أي ستكون مقدرات متحيزة حتى لو زاد حجم العينة n وأصبح غير محدود.

بالنسبة للنموذج (8.5.13) مثبت في الملحق A13، الفقرة 3A.13 أن (جوجاراتي، 2015 صفحة 671):

$$plim\hat{\beta} = \beta\left[\frac{1}{1+\sigma_w^2/\sigma_{X^*}^2}\right] \quad \ldots\ldots\ldots\ldots (10.5.13)$$

حيث σ_w^2 و $\sigma_{X^*}^2$ تمثل تباين w_i و X^* بالترتيب و $plim\hat{\beta}$ تعني النهاية الاحتمالية لـ β.

بما أن المقدار ما بين الأقواس متوقع أن يكون أقل من 1 (لماذا؟) فإن (10.5.13) توضح أنه حتى إذا زاد حجم العينة بشكل غير محدود فإن $\hat{\beta}$ لن يؤول إلى β. في واقع الأمر إذا افترضنا أن β موجبة، فإن $\hat{\beta}$ ستقدر قيمة β بأقل من قيمتها الحقيقية، وبالتالي يكون التحيز أقل من الصفر. بالطبع إذا لم يكن هناك أي أخطاء في القياس في X (أي أن$\sigma_w^2 = 0$) فإن $\hat{\beta}$ سيعتبر مقدرًا متسقا بالنسبة لـβ .

وبالتالي تمثل أخطاء القياس مشكلة حقيقية عندما تظهر في المتغير أو المتغيرات المفسرة، حيث إنهم يؤدون إلى عدم إمكانية الحصول على مقدر متسق. بالطبع، كما سبق ورأينا، إذا ظهرت أخطاء القياس في المتغير التابع فقط، فإن المقدرات لا تزال غير متحيزة ومتسقة أيضًا. أما إذا ظهرت أخطاء القياس في المتغيرات المفسرة. ما الحل؟ الإجابة ليست بالسهولة المتوقعة. إحدى الإجابات المتطرفة، أننا يمكن أن نفترض أن σ_w^2 صغير للغاية مقارنة مع $\sigma_{X^*}^2$، وبالتالي لكل التطبيقات العلمية يمكن أن نفترض عدم حدوث المشكلة، ونستمر في استخدام تقدير OLS التقليدي. بالطبع، هنا يمكننا مشاهدة أو قياس أي من σ_w^2 أو $\sigma_{X^*}^2$، وبالتالي لا يمكن الحكم على مدى تأثيرهما.

إحدى الإجابات الأخرى والتي يمكن اعتبارها أسلوب علاج للمشكلة الحالية هو التعامل مع المتغيرات المساعدة، وعلى الرغم من أنها ستكون مرتبطة بدرجة كبيرة مع المتغيرات X الأصلية، إلا أنها غير مرتبطة مع حد خطأ المعادلة أو حدود أخطاء القياس (أي u_i وw_i). إذا استطاع إيجاد مثل هذا المتغير المساعد، بالتالي يمكنك الحصول على مقدر متسق لـβ.

ولكن هذه العملية، التحدث عنها أسهل بكثير من تنفيذها. في الواقع العملي ليس من السهل إيجاد مثل هذه المتغيرات المساعدة، فنحن في الأغلب كثيرو الشكوى من المناخ، ولكننا لا تستطيع فعل أي شيء حياله، وإلى جانب صعوبة الحصول على المتغيرات المساعدة، فإنه أيضًا من الصعب معرفة ما إذا كانت المتغيرات المساعدة بالفعل مستقلة عن حدود الخطأ u_i و w_i أم لا.

في الأدبيات هناك اقتراحات أخرى لحل هذه المشكلة.

ولكن أغلب هذه الحلول خاصـــة بأوضـــاع محددة ومبنية على فروض مقيدة عديدة وبالتالي لا توجد حلول محددة لمشكلة أخطاء القياس. ولذلك لابد من القياس الدقيق على قدر المستطاع للبيانات محل الدراسة.

مثال: An Example

سنلخص الفقرة السابقة في المثال التالي:

جدول (213) يعطي بيانات افتراضية عن نفقات الاستهلاك الحقيقية Y^*، الدخل الحقيقي X^*، الاستهلاك المقاس Y والدخل المقاس X ، الجدول أيضًا يتضح فيه كيفية قياس تلك المتغيرات.

أخطاء القياس في المتغير التابع Y فقط: Measurement Errors in the Dependent Variable Y only

بناءً على البيانات المعطاة، دالة الاستهلاك الحقيقية هي:

$$\begin{matrix}\hat{Y}_i^* = 25.00 + 0.6000X_i^* \\ (10.477) \quad (0.0584) \\ t = (2.3861) \ (10.276) \\ R^2 = 0.9296\end{matrix} \qquad (11.5.13)$$

في حين أنه إذا استخدمنا Y_i بدلاً من Y_i^* ستحصل على:

$$\begin{matrix}\hat{Y}_i = 25.00 + 0.6000X_i^* \\ (12.218) \quad (0.0681) \\ t = (2.0461) \ (8.8118) \\ R^2 = 0.9066\end{matrix} \qquad (12.5.13)$$

كما يتضـــح من النتائج الســـابقة وبناءً على النظرية، القيمة المقدرة للمعاملات تظل كما هي. الأثر الوحيد لخطأ القياس في المتغير التابع هو أن الأخطاء القياســية المقدرة للمعاملات تزداد [انظر (5.5.13)] ويتضــح ذلك في (12.5.13) عمومًا لاحظ أن معاملات الانحدار في (11.5.13) و (12.5.13) | متساوية، حيث أن البيانات تم توليدها لتتماشى مع فروض نموذج أخطاء القياس.

أخطاء القياس في X: Errors of Measurement in X

نحن نعرف أن الانحدار الحقيقي هو (11,5,13)، افترض الآن أنه بدلاً من استخدام X_i^*، استخدمنا X_i (لاحظ أنه في الواقع نادرًا ما نستطيع قياس X_i^*) نتائج الانحدار ستكون كالتالي:

$$\hat{Y}_i^* = 25.992 + 0.5940X_i$$
$$(11.0810) \quad (0.068117) \qquad (13.5.13)$$
$$t = (2.3457) \ (9.6270)$$
$$R^2 = 0.90205$$

هذه النتائج عندما تكون هناك أخطاء قياس في المتغير أو المتغيرات المفسرة.

المعاملات المقدرة متحيزة. لحسن الحظ في هذا المثال التحيز صغير نسبيًا – من (10.5.13) نرى أن التحيز يعتمد على $\sigma_u^2/\sigma_{X^*}^2$، وعند توليد البيانات افترضنا أن $\sigma_u^2 = 36$ و $\sigma_{X^*}^2 = 3667$ ما يجعل معامل التحيز صغيرًا يساوي تقريبًا $0.98\%(36/3667)$

سنترك للقارئ تحديد ما يحدث إذا كان هناك خطأ في قياس كل من Y وX. أي عندما نقوم بعمل انحدار لـY_i علىX_i بدلاً من Y_i^* على X_i^* (انظر تمرين 23.13) .

جدول (213) بيانات افتراضية عن Y^* (نفقات الاستهلاك الحقيقية)، X^* (الدخل الحقيقي) وY (نفقات الاستهلاك المقاسة) و X (الدخل المقاس) البيانات كلها بالدولار.

Y^*	X^*	Y	X	e	w	u
75.4668	80.00	67.6011	80.0940	−7.8655	0.0940	2.4666
74.9801	100.00	75.4438	94.5721	0.4638	−8.4279	−10.0199
102.8242	120.00	109.6956	112.1406	6.8714	2.1406	5.8242
125.7651	140.00	129.4159	145.5969	3.6509	5.5969	16.7651
106.5035	160.00	104.2388	168.5579	−2.2647	8.5579	−14.4965
131.4318	180.00	125.8319	171.4793	−5.5999	−8.5207	−1.5662
149.3693	200.00	153.9926	203.5333	4.6233	3.5366	4.3693
143.8628	220.00	152.9208	222.8533	9.0579	2.8533	−13.1372
177.5218	240.00	176.3344	232.9879	−1.1874	−7.0120	8.5218
182.2748	260.00	174.5252	261.1813	−7.7496	1.1813	1.2748

Note: The data on X^* ate assumed to be given In deriving the other variable the assumption made were as follows : (1) $E(u_i) = E(\varepsilon_i) = E(w_i) = 0$; (2)$cov(X, u) = cov(X, \varepsilon) = cov(U, \varepsilon) = cov(w, u) = cov(\varepsilon, w) = 0$; (3)$\sigma_u^2 = 100, \sigma_x^2 = 36, and\ \sigma_w^2 = 36; and$ (4)$Y_i^* = 25 + 0.6X_i^* + u_i, Y_i = Y_i^* + \varepsilon_i$, $and\ X_i = X_i^* + w_i$

6.13 التوصيف الخاطئ لحد الخطأ العشوائي:

INCORECT SPECIFICATION OF THE STOCHASTIC ERROR TERM

المشكلة الأكثر ظهورًا أمام الباحث هي توصيف حد الخطأ u_i داخل نموذج الانحدار. فيما أن حد الخطأ لا يشـاهد مباشـرةً، فإنه لا توجد طريقة لتحديد الشـكل الذي يظهر به في النموذج. لتوضـيح ذلك دعنا نعود إلى النماذج المعطاة في (8.2.13) و (9.2.13). للتبسـيط، افترضـنا عدم وجود الحد الثابت المقطوع من المحور الصـادي في النموذج، وافترضـنا أيضًـا أن u_i الموجود في (8.2.13) يسـتوفي فيه $\ln u_i$ كل فروض OLS التقليدية.

إذا افترضـنا أن (8.2.13) هو النموذج "الصـحيح"، ولكن قدرنا (9.2.13) ما هي عواقب ذلك؟ يتضـح من ملحق A13، الفقرة 4A.13 أنه إذا كان $\ln u_i \sim N(O, \sigma^2)$ فإن:

$$u_i \sim \log normal \left[e^{\sigma^2/2}, e^{\sigma^2}\left(e^{\sigma^2} - 1\right)\right] \ldots\ldots\ldots\ldots\ldots (1.6.13)$$

وكنتيجة لذلك يكون لدينا التالي:

$$E(\hat{\alpha}) = \beta e^{\sigma^2/2} \qquad \ldots\ldots\ldots (2.6.13)$$

حيث e هي أساس اللوغاريتم الطبيعي.

كما ترى $\hat{\alpha}$ مقدر متحيز، حيث إن قيمة المتوسط لا تساوي القيمة الحقيقية لـــ β. سنستعرض المزيد من التفاصيل عن توصيف حد الخطأ العشوائي في الفصل الخاص بنماذج الانحدار غير الخطية في المعالم.