الاسبوع الثاني ٢٢/١٢/١٤٠٤م الاثنين

16) The Simple Linear Regression model and Correlation Coefficient — نموذج الانحدار الخطي البسيط ومعامل الارتباط

لنفترض وجود متغير مستقل (X_i) ومتغير تابع (Y_i):

$$Y_i = f(X_i)$$

حيث ان i تأخذ القيم 1، 2، 3، .... n

وفي المثال الحالي الاتي نريد تحديد العلاقة بين المتغيرين مع افتراض ان العلاقة خطية والمتمثلة بالمعادلة التالية

$$Y_i = a + bX_i$$

وهذا الشكل يدل على وجود علاقة سببية وحيدة الاتجاه بين المتغيرين X و Y، بمعنى ان X هو السبب فيما يطرأ على Y من تغيرات والعكس ليس صحيحاً.

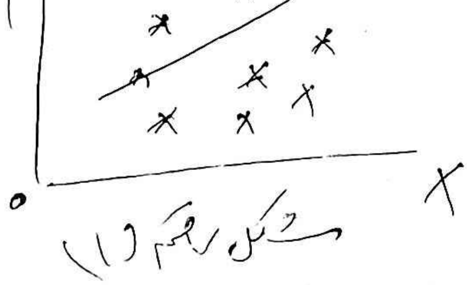
ومعلمات الدالة اعلاه هي a و b وتهدف الى ايجاد التقديرات الكمية لهذه المعلمات حيث ان

a: تمثل ثابت التقاطع او الجزء المحصور بين نقطة الاصل ونقطة تقاطع المنحنى مع محور Y

b: تمثل مقدار التغير في Y نتيجة تغير X بمقدار وحدة واحدة

وعلى اساس ما تقدم فالشكل اعلاه للدالة او لشكل يعني ان العلاقة بين (X و Y) هي علاقة مضبوطة (exact relationship) بمعنى ان Y سيكون التغير فيه (X) فقط وانه ليس هنالك اي عوامل اخرى تؤثر في المتغير التابع Y، فاذا كان هذا صحيحاً فان جميع النقاط (X و Y) سوف تقع على خط مستقيم

الا انه لو حاولنا جمع مشاهدات عن Y (والكمية المعروضة مثلا) عند اسعار (X) مختلفة، ووضعناها في شكل الانتشار نجد انها لن تقع كلها على خط مستقيم لوأي متحول



شكل رقم (1)

17

ومن شكل الانتشار نجد أن العلاقة بين المتغيرين X و Y تتخذ شكلاً قريباً من الخط المستقيم وانحرافات النقاط عن هذا الخط يمكن ارجاعها الى عدة عوامل ؟

١- حذف بعض المتغيرات من الدالة : كل متغير في الواقع الاقتصادي يتأثر بعدد كبير من المتغيرات او العوامل إلا أنه لا يمكن أن نأخذها جميعها بنظر الاعتبار عند دراسة دالة معينة، وذلك لعدة اسباب

أ- بعض هذه العوامل قد لا تكون معروفة حتى للشخص الأكثر إلماماً بالعلاقة المدروسة

ب- حتى اذا توفرت لدينا معرفة بمثل هذه العوامل، فإن بعضها غير قابل للقياس أحصائياً (مثل الذوق، التوقعات، الرغبات ... الخ)

ج- بعض العوامل تكون عشوائية وتظهر بشكل وأوقات لا يمكن التنبؤ بها مثل الكوارث والهزات الارضية والحروب ... الخ.

د- بعض العوامل ذات تأثير ضئيل على المتغير التابع عند النظر الى كل واحدة لذلك تكون معاملاتها ضئيلة بحيث لا يمكن قياسها قياساً معنوياً ولكن الا انه لو جمعت هذه العوامل مع بعضها البعض فإنها قد تفسر جزءاً لا بأس به من التغيرات في المتغير التابع.

هـ- عدم توفر البيانات حول المتغير أو عدم صحتها لاستخدامه في دالة بصورة كافية

٢- السلوك العشوائي للكيان البشري — إن انتشار النقاط حول الخط يمكن إرجاعه الى عنصر عشوائي في السلوك البشري، فردود فعل البشر لا يمكن التنبؤ بها الى حد ما وقد تسبب انحرافات عن السلوك الاعتيادي الذي نتوقعه، كمثال ذلك أن المستهلك قد يغير من نموذج نفقاته من لحظة لأخرى بدون تغير الدخل والأسعار.

٣- عدم كفاية صياغة الشكل الرياضي للنموذج : فقد نجعل العلاقة خطية في حين أنه كان يجب أن تكون لا خطية، أو أننا نكون قد حذفنا بعضاً من معادلات النموذج

٤- أخطاء التجميع Aggregation Errors : غالباً ما نستخدم بيانات مجمعة (إجمالي الاستهلاك، إجمالي الدخل ... الخ) تجمع فيها عناصر متعلقة بأفراد ذوي سلوك متماثل، وفي هذه الحالة تتحول المتغيرات التي تضم خصائص فردية تكون مفقودة.

5- اخطاء القياس Measurment Errors

قد تكون انحرافات النقاط عن الخط ؛ راجع ايضاً لوجود اخطاء في قياس المتغيرات نتيجة لطرق جمع ومعالجة البيانات الاحصائية.

مصادر الاخطاء الاربعة الاولى تؤدي الى الحصول على شكل خاطئ للمعادلة، ولذلك فاطلق عليها الاخطاء في المعادلات Errors in the equations او اخطاء الحذف Omission Errors، اما المصدر الخامس للخطأ فيمس الخطأ في القياس او المشاهدات، ومن ثم انه يظهر حتى لو المصادر الاربعة لم تكن موجودة.

ولاجل معالجة الخطأ هذه والاخطاء الاخرى يقوم بادخال متغير عشوائي في دالة الاقتصاد القياسي، ويرمز لهذا المتغير بالرمز (U_i) اما القيمة التقديرية له فيرمز لها بالرمز (e_i) ويطلق عليه عنصر الاضطراب العشوائي Random disturbance term او حد الخطأ (error term) او العنصر العشوائي stochastic term، وقد اطلق هذا الاسم لانه يؤدي الى اضطراب العلاقة الخطية المضبوطة المفترض وجودها بين X و Y وادخال العنصر العشوائي في الدالة يصبح النموذج القياسي Econometric model

$$Y_i = a + bX_i + U_i$$

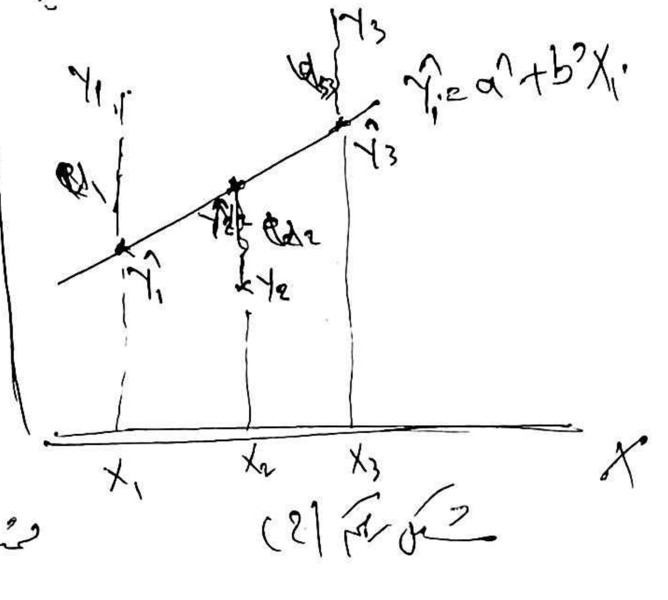
وعليه فان التغيرات الكلية في المتغير Y سوف تقسم الى جزئين:

أ- الجزء الاول يتم التعبير عنه بواسطة خط الانحدار $a + bX_i$

ب- الجزء الاخر يتم التعبير عنه بواسطة حد الخطأ وكما في الشكل

في الشكل (2): الخط يمثل الجزء المضبوط من العلاقة بين X و Y، اما انحرافات المشاهدات عن هذا الخط فتمثل الجزء العشوائي من العلاقة، ولولا وجود العنصر العشوائي random variable لكانت جميع النقاط Y_1, Y_2, Y_3 تقع على الخط $\hat{Y}_i = \hat{a} + \hat{b}X_i$

فالفرق بين القيمة الحقيقية Y والقيمة التقديرية $\hat{Y}$ تمثل مقدار الخطأ (e_i) او ما يسمى العنصر العشوائي random term



شكل رقم (2)

• وكلمة فإن الأولى من قسم المتغير التابع (Y) سيتم التعبير عنه عبر لبنتين الأولى والتي تعزى الى (X_i) والمركبة الثانية تعزى الى حد الخطأ e_i بمعنى ان التغيرات ما المتغير التابع والتي لا يتم تفسيرها بالمتغير المستقل (X_i) سيتم تفسيرها بحد الخطأ

$$Y_i = \underline{a + bX_i} + \underline{U_i}$$

التباين Variation in Y_i = Explained variation (تباين موضح) + random variations (تباين عشوائي) / unexplained variation (غير موضح)

Explained predetermined variation (pre-d)
هو كمية من قبل متغيرات محددة مسبقا

Expressed by u_i ويعبر عنه ب

- exogenous variable
- Lagged V.

(٤٠)

التصادفية
افتراضات نموذج الانحدار الخطي

Assumptions of the Linear stochastic Regression model

افتراضات حد الخطأ العشوائي
Assumptions of the Error trem (random)

ع - الافتراضات التصادفية (العشوائية)

١- ان حد الخطأ هو متغير عشوائي حقيقي يتوزع توزيعاً طبيعياً بوسط حسابي يساوي صفر $E(u_i) = 0$ وتباين ثابت مقداره (σ_u^2)

$u_i \sim (0, \sigma^2)$

والمقصود بأنه متغير عشوائي حقيقي يعني ان اي قيمة ما قد يأخذها المتغير العشوائي (u_i) في فترة زمنية ما تعتمد على الصدفة وقد تكون سالبة او موجبة او صفر. اما كونه يمتلك وسط حسابي مقدره صفر فهذا يعني ان لكل قيمة من قيم (X_i) سوف تأخذ (u_i) قيماً متعددة بعضها اكبر من الصفر وبعضها اقل من الصفر. اما اذا اخذنا جميع هذه القيم المحتملة لـ (u_i) فان متوسط هذه القيم سوف تكون صفر.

$E(u_i) = 0$

وكونه يمتلك تباين ثابت يعني ان

$$Var(u_i) = E(u_i - E(u_i))^2 = E(u_i)^2 = \sigma_u^2$$ متجانس

اي ان تباين قيم (u_i) حول متوسط هذه القيم تكون ثابتاً بالنسبة لجميع قيم X، بعبارة اخرى سوف تظهر قيم (u_i) نفس التشتت حول متوسطها.

كما ان المتغير العشوائي يمتلك توزيع معتدل Normal Distribution اي ان توزيع قيم u_i حول متوسط المساوي للصفر تكون على شكل جرس ومتماثل وذلك لكل قيمة من قيم X

(21)

2 - ان القيم المتتابعة او المتسلسلة لـ u_i تكون مستقلة اي ان:

$$cov(u_i, u_j) = E(u_i, u_j) = 0$$

$i \neq j \quad i, j = 1, 2, 3 \ldots n$

covariance
covariance

ويمكن اثبات ذلك كالاتي

$$cov(u_i u_j) = E(u_i - E(u_i))[u_j - E(u_j)]$$

$$cov(u_i u_j) = E(u_i u_j)$$

مستقل

وبما ان $0 = E(u_i) = E(u_j)$

$$= E(u_i) E(u_j) = 0$$

$$cov(u_i, u_j) = 0$$

بعبارة اخرى ، ان القيم المختلفة للمتغير العشوائي u_i تكون مستقلة عن بعضها البعض واختصار ان التباين المشترك لاي u_i مع اي من u_j مساوية للصفر وعليه فان قيمة العنصر العشوائي ان فترة لا تعتمد على قيمته في فترة اخرى.

3 - ان حد الخطأ هو متغير مستقل تماما عن المتغيرات التوضيحية الاخرى بمعنى انه لا يتغير بتغير اي من $X's$ وهذا يعني ان التباين المشترك بين حد الخطأ والمتغيرات المستقلة تكون صفر

التباين المشترك $cov(X_i, u_i) = 0$

$$cov(X_i \cdot u_i) = E\{[X_i - E(X_i)][u_i - E(u_i)]\} = 0$$

وبما ان
$E(u_i) = 0$

$$= E\{[X_i - E(X_i)] u_i\}$$

$$= E(X_i u_i) - E(X_i) E(u_i)$$

$$cov(X_i u_i) = E(X_i u_i)$$

$$= X E(u_i)$$

لان $E(u_i) = 0$

$$cov(X_i u_i) = 0$$

من افتراضنا ان قيم X_i ثابتة لعينة العينة

(22)

٤ - ان المتغيرات المستقلة مقاسة بدون اخطاء.
او ان u_i ستستوعب تأثير المتغيرات المحذوفة
واخطاء قياس المتغير التابع معنى اننا سنفترض ان المتغيرات المستقلة
ليست اخطاءها Y_i فقد تكون ذات اخطاء او بدونها.

ب - الشروط الاخرى

٥ - ان المتغيرات المستقلة لا ترتبط بشكل تام ارتباطاً قطعياً
بمعنى ان المتغيرات المحددة مسبقاً عموماً لا ترتبط بقوة اي انها
لا ترتبط ارتباطاً متعدداً قطعياً قوياً. وذلك ليسهل التعرف على التأثير
المنفرد لكل متغير على حدة.

٦ - ان المتغيرات الكلية التجميعية ينبغي ان تجمع بشكل صحيح لكي تكون خالية
من اخطاء التجميع

٧ - ان العلاقة المراد تقديرها ينبغي ان تكون مشخصة Identified
بمعنى ان تكون للمعلمات المراد تقديرها قيمة وحيدة Unique
وهو تعبير صحيح عن العلاقة الدالية.

٨ - ان العلاقة قد صيغت بشكل صحيح بمعنى ان تكون خالية من
اخطاء الصياغة من حيث عدد وتحديد المتغيرات التوضيحية (وهنا
صحة شمولها لجميع العلاقات الهامة والصورة الدالية الصحيحة
الخطية أو غير الخطية)

توزيع المتغير التابع Y_i
العشوائي

$$Y_i \sim (\alpha + BX_i, \sigma^2)$$

توزيع المتغير التابع هو توزيع معتدل، تكون قيمة المتوسط له

$$E(Y_i) = \alpha + BX_i$$

$$\text{mean of } Y_i = E(Y_i)$$
$$= E(\alpha + BX_i + u_i)$$
$$= \alpha + BX_i + E(u_i)$$
$$E(Y_i) = \alpha + BX_i$$

لان X_i الكميات ثابتة و $E(u_i) = 0$