

الفصل الثالث

الانحدار الخطي البسيط

Simple Linear Regression

أ. لولوه بن سعيد

مقدمة

- ▶ يعتبر الانحدار أحد الأساليب الإحصائية التي تستخدم في قياس العلاقات الاقتصادية، بحيث يختص بقياس العلاقة بين المتغير التابع و المتغير المستقبل / المفسّر.
- ▶ في أسلوب الانحدار لا يتم تحديد أي المتغيرات تابع و أيها مستقل الا من قبل الباحث نفسه بالاستعانة بالنظرية الاقتصادية أو الملاحظة.
- ▶ تنقسم نماذج الانحدار إلى عدة أنواع:



- ▶ **تحدد درجة الخطية على أساس درجة العلاقة المراد قياسها:**
 - في الانحدار الخطي تكون المعادلة الممثلة للعلاقة من الدرجة الأولى
 - في الانحدار الغير خطي تكون المعادلة الممثلة للعلاقة من الدرجة الثانية أو الثالثة و غيرها

- ▶ **أما عن صفتي بسيط و متعدد فإنهما يتحددان بعدد المتغيرات المستقلة في المعادلة:**
 - فالانحدار البسيط يقيس العلاقة بين متغيرين أحدهما تابع و الآخر مستقل
 - أما الانحدار المتعدد فهو يقيس العلاقة بين متغير تابع و أكثر من متغير مستقل

- ▶ و بالتالي يمكن تقسيم نماذج الانحدار إلى أربعة أنواع:



أولاً: نموذج الانحدار الخطي البسيط

- يعد الانحدار البسيط أداة تهتم بوصف و تحليل و تقييم العلاقة بين متغير تابع و متغير آخر مستقل. و يعرف بأنه دالة رياضية أو معادلة تبين العلاقة بين المتغيرين.
- يعتبر من أبسط نماذج الانحدار و أكثر الأدوات المستعملة في التحليل القياسي، و هناك نماذج عديدة للعلاقات الاقتصادية البسيطة التي يمكن قياسها بأسلوب الانحدار البسيط مثل:

- العلاقة بين الاستهلاك و الدخل المتاح (دالة الاستهلاك)
- العلاقة بين الادخار و الدخل المتاح (دالة الادخار)
- العلاقة بين الكمية المطلوبة و سعر السلعة (دالة الطلب)

► يعتبر نموذج الانحدار البسيط لدراسة العلاقة بين متغيرين من أكثر النماذج استخداماً إلا أن عليه بعض القيود كأداة عامة للتحليل التجريبي، و استخدامه مناسب كأداة تجريبية و سوف نتعلم كيفية تفسير هذا النموذج.

► يبدأ تطبيق التحليل الاقتصادي القياسي بالفرضية التالية:

أن X و Y متغيران يمثلان بعض المجتمع، و لدينا اهتمام بشرحهما أو دراسة كيفية تغير Y بتغير X مثلاً Y هي أجر العامل في الساعة و X هي سنوات التعليم، فيكون علينا لدراسة هذه العلاقة أن نواجه ثلاثة قضايا:

أولاً: لأنه لا يوجد علاقة دقيقة بين متغيرين أبداً، كيف نسمح لعوامل أخرى أن تؤثر في Y ؟

ثانياً: ماهي العلاقة الدالية بين X و Y ؟

ثالثاً: كيف يمكننا أن نتأكد من التقاط علاقة ثبات بقية العناصر *ceteris paribus* بين X و Y ؟

يمكننا حل القضايا السابقة بكتابة علاقة المتغيرين بمعادلة بسيطة تأخذ الصيغة

$$Y = \beta_0 + \beta_1 X + u \quad (\dots 1) \quad \text{التالية:}$$

تسمى المعادلة السابقة بنموذج الانحدار الخطي بمتغيرين لأنه يربط بين المتغيرين X و Y :

Y : هي المتغير التابع dependent variable أو التوضيحي explained variable أو متغير الاستجابة

response variable أو متغير التوقع predicted variable او المفسّر regressand.

X : تمثل المتغير المستقل independent variable أو التفسيري explanatory variable أو الرقابي

control variable أو متغير المقدر regressor .

β_0 : الحد الثابت أو المقطوع من محور المتغير التابع، و هي قيمة Y عندما تكون قيمة X تساوي

صفر و تسمى بالمعلمة التقاطعية.

β_1 : ميل الخط المستقيم و تسمى بالمعلمة الانحدارية و تشير إلى مقدار التغير في المتغير التابع

Y إذا تغير المتغير المستقل X بوحدة واحدة.

u : حد الخطأ العشوائي أو اضطراب العلاقة و يمثل العوامل الأخرى التي تؤثر في y بخلاف x .

▶ تتناول المعادلة (١) مسألة علاقة الدالة بين y و x مع بقاء العوامل الأخرى ثابتة في u ، و بذلك يكون التغير في حد الخطأ العشوائي صفر أي: $0 = \Delta u$

و بالتالي يصبح تأثير X على Y خطي: $\Delta y = \beta_1 \Delta x$ إذا كانت $\Delta u = 0$

▶ لذا فإن التغير في y هو ببساطة β_1 مضروباً في تغير x و هذا يعني أن β_1 هي معلمة الميل في العلاقة بين X و Y مع بقاء العوامل الأخرى ثابتة على حالها في u التي هي ذات أهمية أساسية في الاقتصاد التطبيقي.

▶ ما المقصود بالتأثير الخطي؟

تعني الخطية أن تغير وحدة واحدة في المتغير المستقل ← تغير المتغير التابع بنفس التأثير.
و لكن ذلك غير واقعي في تطبيقات اقتصادية كبيرة غالباً.

مثال: معادلة الأجر البسيطة

النموذج المتعلق بأجر شخص ما على مستوى التعليم المشاهد و العوامل الأخرى غير المشاهدة هو:

$$\text{Wage} = \beta_0 + \beta_1 \text{educ} + u$$

إذا تم قياس الأجر wage ريال لكل ساعة و educ عدد سنوات التعليم فإن:

β_1 : تقيس التغير في أجر الساعة لكل عام جديد من التعليم مع بقاء جميع العوامل الأخرى

ثابتة مثل خبرة قوة العمل و أخلاقيات العمل أو القدرة الجسمانية ... الخ

u : العلاقة السابقة تحتوي على حد الخطأ العشوائي مما يعني أن تلك العلاقة محتملة و

ليست مؤكدة، لأن تلك العلاقة قد تتأثر بحذف متغيرات ذات أهمية في العلاقة أو إضافة متغيرات

ليست ذات أهمية. كما تتأثر بأي خطأ من الباحث في جمع البيانات أو خطأ في المشاهدة أو

التحليل أو وضع صياغة رياضية غير صحيحة... الخ

ثانياً: تقدير نموذج الانحدار الخطي البسيط

▶ الطريقة الإحصائية التي تستخدم في قياس معاملات الانحدار للنموذج هي طريقة المربعات الصغرى العادية (Ordinary Least Squares (OLS)، وهناك عدة افتراضات احتمالية تتعلق بهذه الطريقة من شأنها تحويل الطريقة الإحصائية التي تستخدم في قياس العلاقة إلى طريقة قياسية تتلاءم مع الطبيعة الاحتمالية للعلاقات الاقتصادية:

الافتراض الأول:

أن المتغير (U) متغير حقيقي عشوائي ويفترض أن يكون حقيقياً أي يأخذ قيماً رقمية محددة و هذه القيم قد تكون موجبة أو سالبة أو صفرية، ويفترض أيضاً أن يكون عشوائياً بمعنى أن القيم التي يأخذها تعتمد على الصدفة فهي غير مؤكدة الحدوث و لها احتمال أقل من الواحد.

الافتراض الثاني:

أن القيمة المتوسطة للمتغير العشوائي (U) عند كل قيمة من قيم المتغير المستقل تساوي صفراً، فلن تكون هناك حاجة لتحديد القيمة المتوقعة للحد العشوائي في التنبؤ نظراً لأنها تساوي صفر.

الافتراض الثالث:

أن تباين الحد العشوائي (U) يكون ثابتاً عند جميع قيم المتغير المستقل، و يعني أن الفرق أو المدى بين الحد الأقصى و الحد الأدنى لقيم المتغير العشوائي عند كل قيمة يكون ثابت.

الافتراض الرابع:

أن المتغير العشوائي له توزيع معتدل عند كل قيمة من قيم المتغير المستقل أي أنه سيكون ناقوسي الشكل.

∴ حسب الافتراضات السابقة يمكن القول أن حد الخطأ العشوائي له قيم محددة و احتمالية و موزعة توزيعاً معتدلاً، متوسطها صفر و تباينها ثابت

الافتراض الخامس:

أن القيم التي يأخذها المتغير العشوائي عند كل قيمة من قيم المتغير المستقل تكون مستقلة عن بعضها البعض، أي أن تباير قيم العشوائي في الفترات المتتالية يكون صفر. و هذا يعفينا من مشكلة مؤداها ان الخطأ العشوائي في فترة واحدة قد يكون هو السبب في توليد الأخطاء العشوائية في كل الفترات التالية.

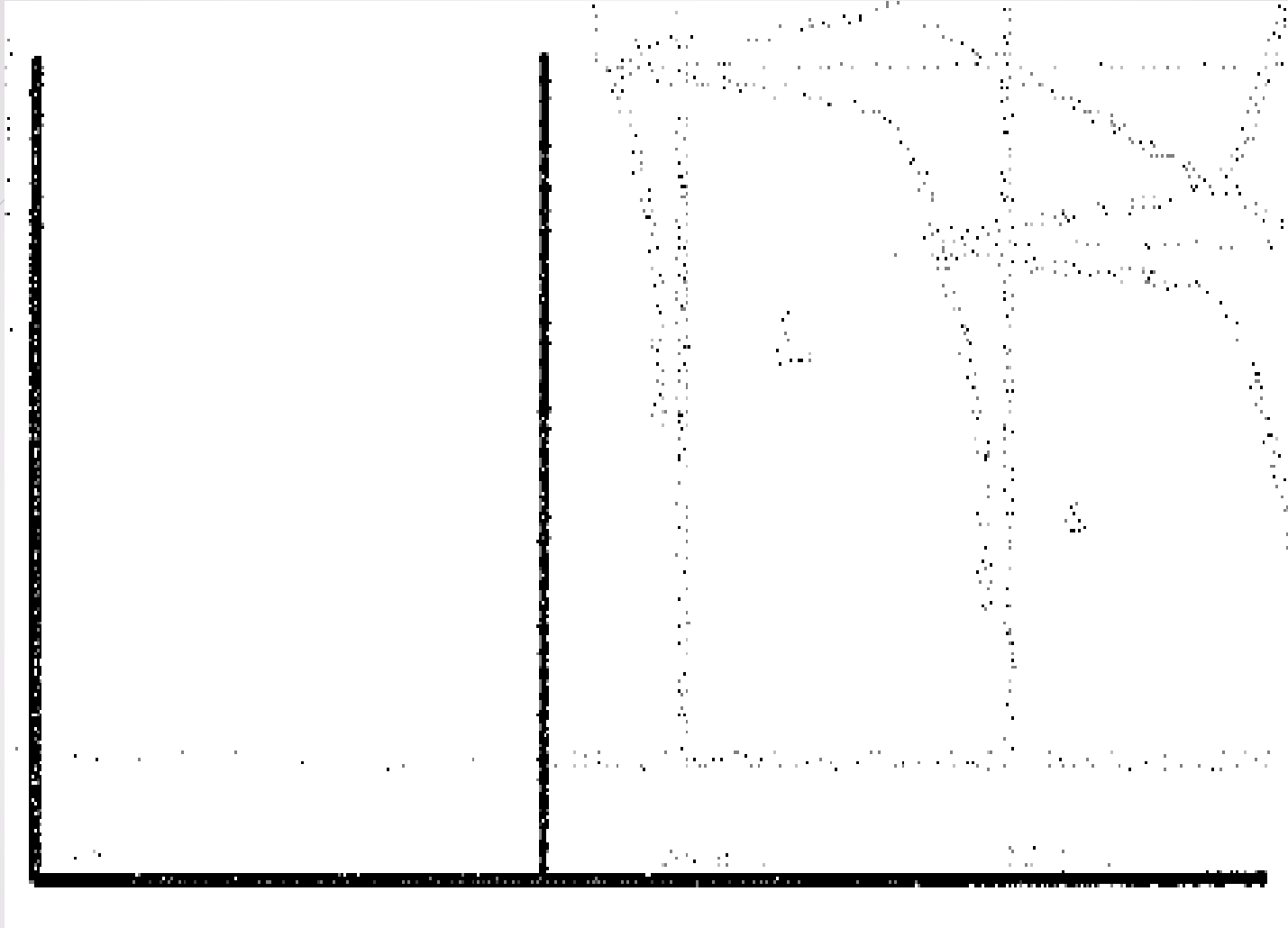
الافتراض السادس:

أن قيم المتغير (U) مستقلة عن قيم المتغير التفسيري (X)، بمعنى أن X لا يؤثر و لا يتأثر بـ U و هذا يعني أن تباير (U ، X) = صفر. و ذلك لأن التداخل بينهما يصعب تحديد النسبة التي يمكن تفسيرها من Y بدلالة X ، و من ثم فإن أخطاء القياس المتضمنة في U تتعلق بالمتغير التابع Y فقط.

الافتراض السابع:

ليست كل قيم المتغير المستقل متساوية (ثابتة)، حيث يجب أن تكون هناك على الأقل قيمة واحدة مختلفة عن باقي القيم، فإذا كان شكل الانتشار رأسي فإن طريقة المربعات الصغرى تصبح غير صالحة لتقدير علاقة الانحدار

Y



X

شكل انتشار المتغير المستقل الثابت

أ. لولوه بن سعيد

الافتراض الثامن:

أن المتغيرات المستقلة كالدخل مثلاً تقاس بلا أخطاء، و هذا يعني ان الأخطاء التي يمكن أن تقع تكون عند قياس المتغير التابع، و من ثم فإن أخطاء القياس المتضمنة في (U) تتعلق بالمتغير التابع Y فقط.

الافتراض التاسع:

أن المتغيرات المستقلة إحصائياً أي إذا وجد أكثر من متغير مستقل فإن الارتباط بينهم يكون منعماً أو ضعيفاً، فلو أن هناك كتغيرين مرتبطين ارتباطاً خطياً تاماً يمكن اعتبارهما متغيراً واحداً و من ثم فإن إدراجهما سوياً في معادلة الانحدار سيؤدي إلى عدم دقة قياس المعلمات.

الافتراض العاشر:

عدم وجود مشاكل للتجميع، أي ان المتغيرات التجميعية تكون بطريقة سليمة. كما يجب ان تكون الصيغة الرياضية متميزة و لا تتشابه مع صيغ أخرى في نفس النموذج بالإضافة إلى صحة تعيين النموذج و التأكد من درجة الخطية و عدد المتغيرات.

إذا تحققت كل هذه الافتراضات في الواقع فإن النتائج المقدره التي سنتحصل عليها بطريقة

أ. لولوه بن سعيد

المربعات الصغرى العادية يمكن الاطمئنان إلى صحتها

نشاط عملي:

■ يحتوي ملف corn.xlsx على بيانات ناتج الذرة و الأسمدة و الذي يعطي الدونم الواحد y شوال ذرة نتيجة استخدام كميات مختلفة من الأسمدة في الدونم x ، لإنتاج مزرعة خلال ١٠ سنوات من سنة ٢٠٠٠ حتى ٢٠٠٩. فضلاً قومي بتقدير العلاقة بين المتغيرين x و y (نموذج انحدار خطي بسيط) و أجيبي على جميع الفقرات و ذلك في ١٠ دقائق ☺



ثالثاً: الخصائص المرغوبة للمقدرات في حالة العينة الصغيرة

- يمكن حصر الخصائص المرغوبة للمقدرات في حالة العينة الصغيرة و التي سيتم اعتمادها في هذا المقرر فيما يلي:
 - 1- عدم التحيز Unbiasedness
 - 2- أقل تباين Least Variance
 - 3- الكفاءة Efficiency
 - 4- الخطية Linearity
 - 5- المثلى الخطية Best Linear Unbiased Estimator (BLUE)
 - 6- أدنى متوسط مربعات خطأ Minimum Mean-Square Error (MSE)
 - 7- الكفاية Sufficiency

١- عدم التحيز:

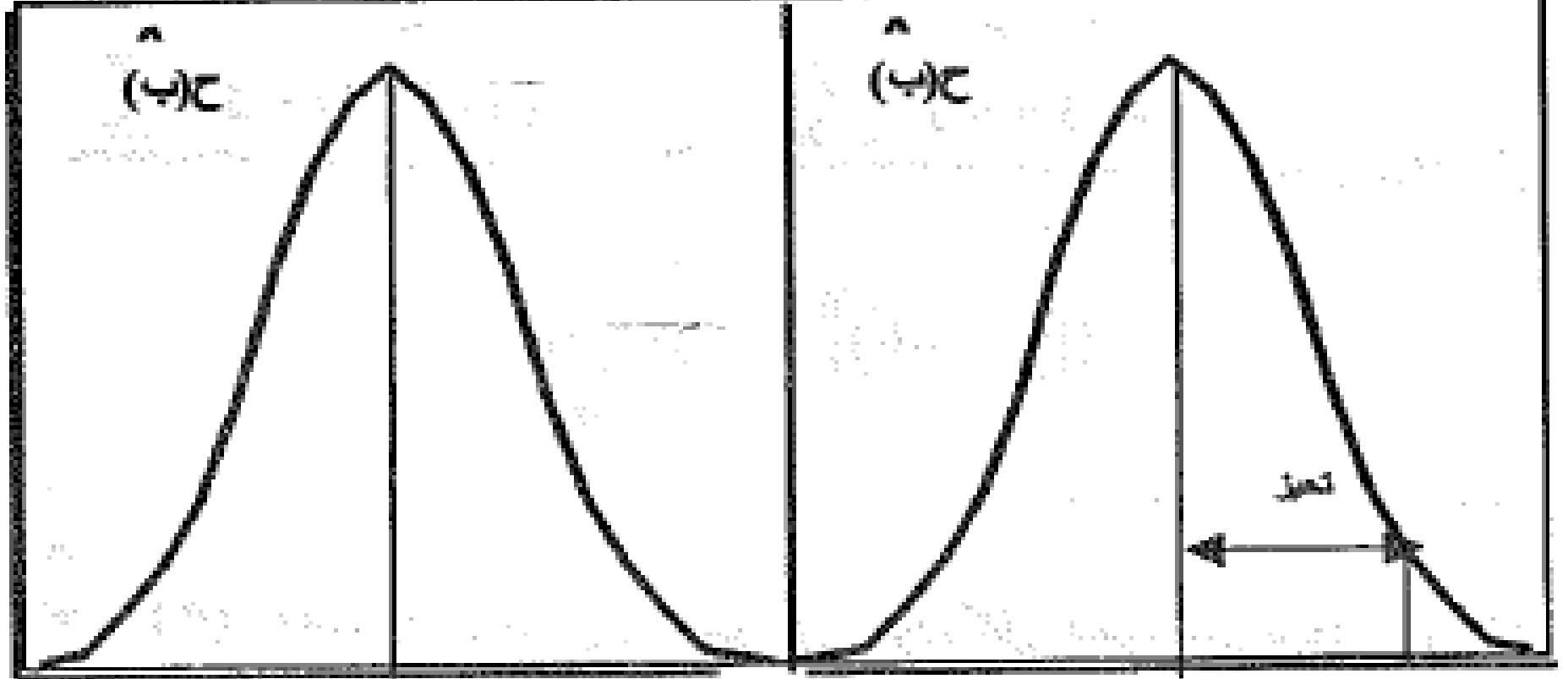
- ▶ يتمثل التحيز في وجود فرق أو انحراف بين القيمة المتوقعة للمقدر و معلمة المجتمع.
- ▶ إذا قمنا بسحب عينات صغيرة من المجتمع و قدرنا $\hat{\beta}$ لكل عينة منها، فإن هذا المقدر سوف يكون متحيزاً إذا كان الفرق بين القيمة المتوسطة له و القيمة المتوقعة لا يساوي الصفر

أي أن :

$$E(\hat{\beta}) - \beta \neq 0 \quad \text{متحيزاً}$$

$$E(\hat{\beta}) - \beta = 0 \quad \text{غير متحيز}$$

- ▶ صفة عدم التحيز غير مرغوب فيها إلا أنها لا تعتبر صفة مهمة في حد ذاتها، و لكن تعتبر مهمة إذا اقترنت بصفات أخرى نراها لاحقاً.



ق (ب) = ب

ق (ب) ب

شکل (۱-۵)

شکل (۲-۵)

ب غیر متحيز

ب متحيز

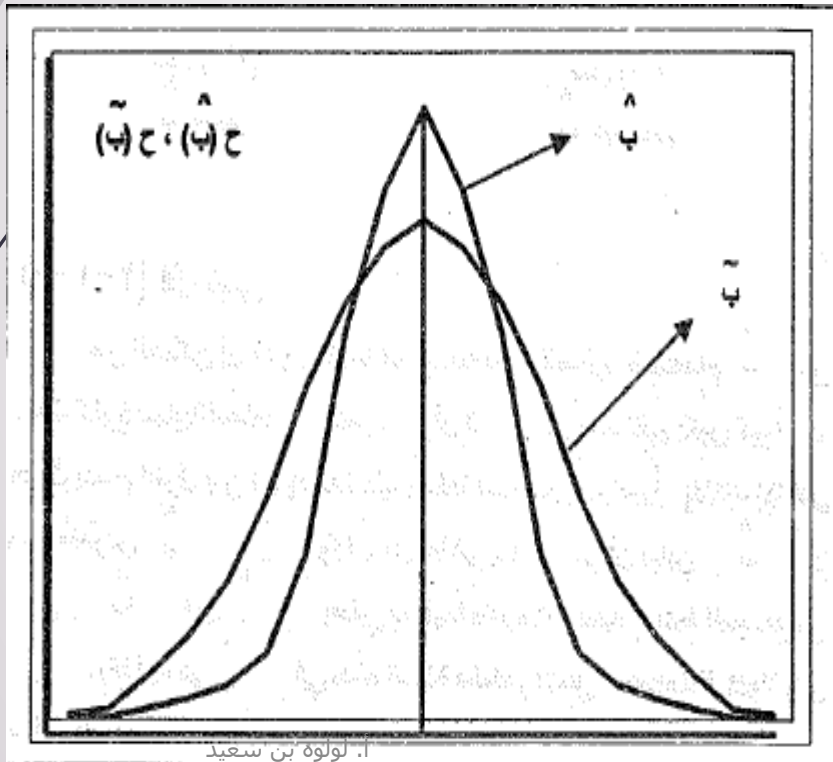
أ. لولوه بن سعيد

٢- أقل تباين:

كما قل تباين المقدّر $\hat{\beta}$ يصبح أكثر تمثيلاً لمعلمة المجتمع و يسمى بالمقدر الأمثل Best Estimator

من الشكل البياني نلاحظ أن المقدرين غير متحيزين و لكن $\hat{\beta}$ أقل تباين من $\tilde{\beta}$.

تعد صفة الأقل تباين مرغوبة إلا أنها في حد ذاتها غير هامة، فهي تستمد أهميتها من اقترانها بصفة عدم التحيز. كما تستمد صفة عدم التحيز أهميتها من اقترانها بصفة أقل تباين



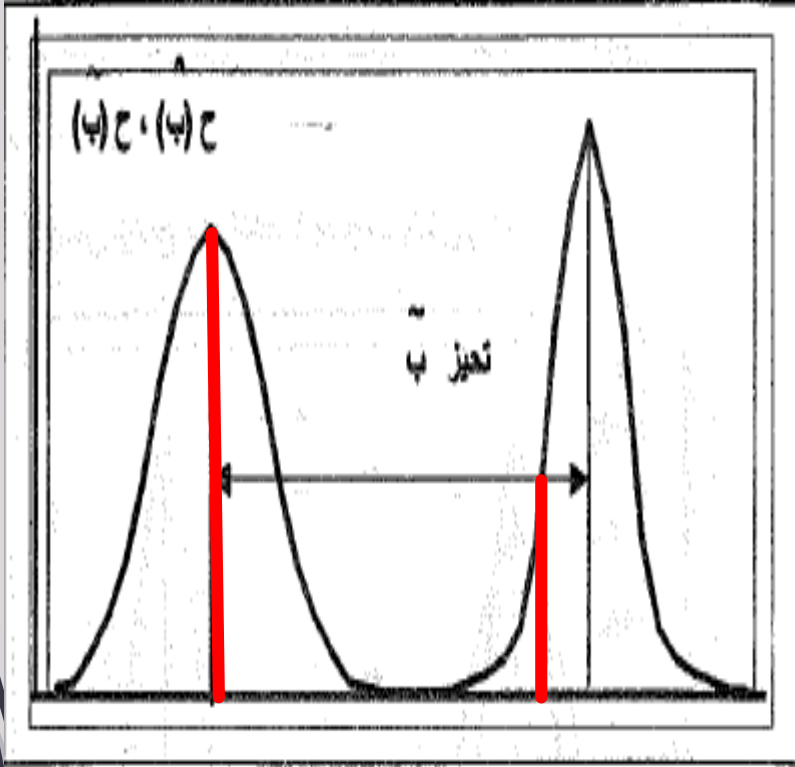
١. لؤلؤة بن سعيد

عدم اقتران صفة عدم التحيز و أقل تباين:

► يتضح من الشكل التالي أن :

$\hat{\beta}$ تتصف بصفة أقل تباين و بالرغم من ذلك فهي متحيزة بدرجة كبيرة جداً عن المعلمة الحقيقية للمجتمع β ، و لذلك فإن صفة أقل تباين ليست هامة في هذه الحالة طالما أن $\hat{\beta}$ لا يمكن أن تعطي قيمة تمثل معلمة المجتمع بسبب صفة التحيز.
كذلك الأمر في $\tilde{\beta}$ فهي غير متحيزة و لكنها لا يمكن أن تمثل الواقع بسبب صفة أقل تباين.

► لابد ان يكون المقدر غير متحيز و يقترن بصفة أقل تباين حتى يعطي تمثيلاً حقيقياً للمجتمع.



ق(ب) = (ب)ق

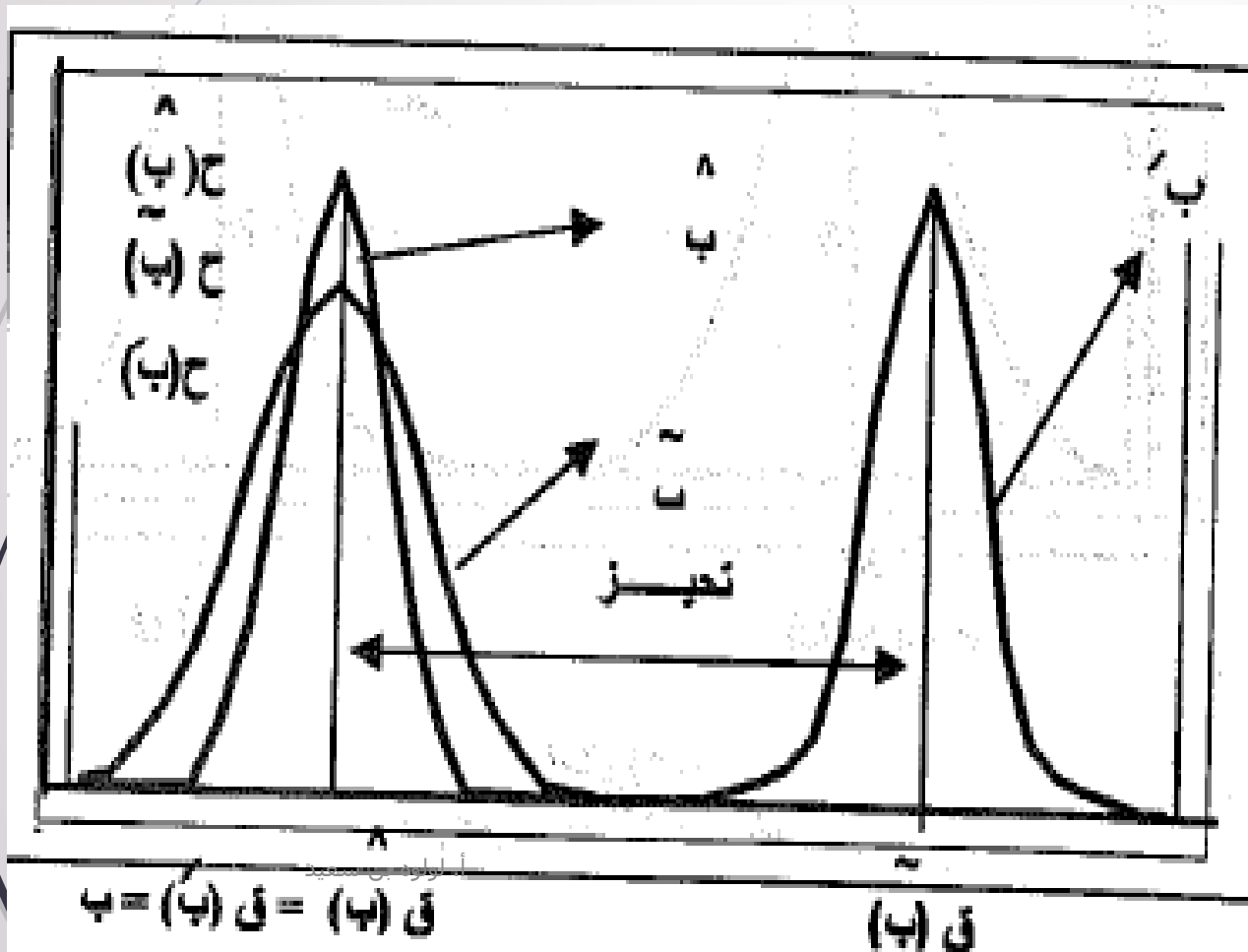
ق(ب)

شكل (٤.٥)

أ. لولوه بن سعيدي
عدم اقتران عدم التحيز بأقل تباين

٣- الكفاءة:

- يعتبر المقدر كفوفاً إذا توفرت فيه خاصية عدم التحيز و أقل تباين.
- من الشكل نستنتج أن $\hat{\beta}$ هو المقدر الكفو بسبب اقتران خاصية عدم التحيز و أقل تباين.



٤- الخطبة:

- يعتبر المقدر خطياً إذا كان على علاقة خطية من الدرجة الأولى مع القيم المشاهدة للمتغير التابع كم تناولنا سابقاً في هذا الفصل.

٥- المثلية الخطية:

- يعتبر المقدر خطياً و غير متحيزاً و أمثل إذا جمع بين صفات ثلاثة: عدم التحيز، أقل تباين، و الخطية.
- و هذا تتحقق خاصية BLUE

٦- أدنى متوسط لمربعات الخطأ MSE:

- ▶ يعتبر هذا المعيار توليفة من صفتي عدم التحيز و أقل تباين، و يتمتع المقدر بهذه الخاصية إذا كانت القيمة المتوقعة لمربع انحرافات القيم المقدرة بواسطته عن معلمة المجتمع (ادنى ما يمكن).
- ▶ فنجد ان القيمة المتوقعة لأي مقدر لابد أن تنحرف عن معلمة المجتمع الأساسية، كما أن تباين القيم لابد ان يكون أكبر من الصفر. و من ثم فإن المقدرات التي تعطي ادنى متوسط لتوليفة عدم التحيز و أدنى تباين يعتبر هو الأفضل و يحقق معيار أدنى متوسط لمربعات الخطأ MSE

٧- الكفاية:

- ▶ يعتبر المقدر كافي عندما يستخدم كل معلومات العينة، بحيث لو تمت إضافة مقدر جديد للنموذج فإنه لن يضيف شيئاً جديداً باستخدام نفس العينة.
- ▶ لا تعتبر الكفاية صفة هامة في ح ذاتها، و لكنها تعتبر شرطاً ضرورياً لخاصية الكفاءة التي تم شرحها سابقاً

الخلاصة

- ▶ يعتبر المقدر أفضل من غيره إذا كان يتصف بعدد من الخصائص المرغوبة أكثر من غيره.
- ▶ لا تعتبر صفة أدنى تباين في حد ذاتها هامة، حيث قد يوجد تباين صغير جداً و لكن هناك تحيز كبير، و من ثم فإن التباين الصغير يكون حول المتوسط الخطأ. كما أن صفة عدم التحيز لا تعتبر هامة إلا إذا اقترنت بخاصية أقل تباين.
- ▶ يلاحظ ان معياري المثلية الخطية و ادنى متوسط لمربعات الخطأ يحتويان على عنصري التحيز و التباين و هما يفضلان بوجود عام أي معيار فردي آخر.
- ▶ فيما يتعلق بطريقة المربعات الصغرى العادية OLS كإحدى المقدرات لمعلمات النماذج فإنها تتصف (عدم التحيز، الخطية، المثلية الخطية)