الجمهورية الجزائرية الدّيمقراطية الشّعبية

وزارة التّعليم العالي والبحث العلمي

جامعة ابن خلدون -تيارت -

كلّية العلوم الاقتصادية، التّجارية وعلوم التّسيير

قسم: العلوم التجارية

دروس محاضرات في مقياس – اقتصاد قياسي-

(Econometrics)

من إعداد:

- شـــــداد محمد

السنة الجامعية 2024-2025

فهرس المحتويات

مقدمـــــــــة

يُعد الاقتصاد القياسي فرعًا محوريًا من فروع العلوم الاقتصادية، إذ يجمع بين التحليل النظري والتقنيات الكمية لتقدير النماذج الاقتصادية والتحقق من صحة الفرضيات العلمية. وقد ساهم تطور أدوات الاقتصاد القياسي في تعزيز فهم العلاقات الاقتصادية المعقدة، ما جعله أداة أساسية لكل باحث وممارس اقتصادي يسعى إلى تفسير الظواهر الاقتصادية وتحليلها استنادًا إلى بيانات كمية دقيقة.

انطلاقًا من أهمية هذا التخصص أو المقياس، تأتي هذه المطبوعة بعنوان "محاضرات في مقياس: اقتصاد قياسي Econometrics"، الموجهة إلى طلبة العلوم الاقتصادية بصفة عامة، وطلبة قسم العلوم التجارية بصفة خاصة، حيث تهدف إلى تقديم أساسيات الاقتصاد القياسي بصورة منهجية ومبسطة. نشير أنه قد اقتصرت هذه المطبوعة على تغطية بعض الجانب النظري فقط، لتشكل قاعدة معرفية للطلبة، على أن تكون هناك مطبوعة مكملة تغطي الجانب التطبيقي عبر مجموعة من التمارين المحلولة التي تعزز الاستيعاب العملي للمفاهيم المدروسة.

تتضمن هذه المطبوعة عدة محاور مترابطة تغطي المبادئ الأساسية للاقتصاد القياسي. تبدأ أولاً بالتعريف بـ مفهوم الاقتصاد القياسي، حيث تتناول باختصار مفهومه وأهدافه الرئيسية، مع إبراز علاقاته بالتخصصات الأخرى كالاقتصاد النظري والإحصاء. ثم تتناول محور الانحدار الخطي البسيط وطريقة المربعات الصغرى العادية، حيث نشرح من خلاله كيفية نمذجة العلاقة بين متغيرين فقط، مع تقديم طريقة تقدير المعلمات وأهم الفرضيات التي يقوم عليها النموذج.

بعد ذلك، نتوسع في الدراسة إلى الانحدار الخطي المتعدد، حيث نتعامل مع نماذج تحتوي على أكثر من متغير تفسيري، مما يسمح بتحليل علاقات أكثر واقعية وتعقيدًا. يلي ذلك محور بعنوان التوسع في نموذج الانحدار الخطي المتعدد، حيث نتناول عناصر التحول الهيكلي، النماذج المقيدة والمتغيرات الصورية.

من بين الإشكاليات التي قد تعترض الباحث عند التعامل مع النماذج القياسية، إشكالية تعدد الارتباط الخطي وطرق اختيار المتغيرات التفسيرية، في هذا المحور سنحاول توضيح كيفية تشخيص هاته الإشكالية وتأثيرها على التقديرات والنموذج بصفة

عامة مع عرض بعض الحلول العملية لها. كذلك نعرض لاحقا إشكالية الارتباط الذاتي للأخطاء (Autocorrelation) ، التي تظهر عادة عند التعامل مع البيانات الزمنية، مبينين أسبابها وكيفية الكشف عنها وطرق معالجتها.

تتناول المطبوعة أيضًا إشكالية عدم ثبات تباين الأخطاء (Heteroscedasticity) ، التي تؤثر على كفاءة مقدرات المربعات الصغرى، موضحين أيضا كيفية الكشف عنها والتعامل معها. ثم نعالج إشكالية التوزيع غير الطبيعي للأخطاء (Normality)، نظرًا لأهميتها في إجراء اختبارات الفرضيات وتفسير نتائج النماذج.

بعد ذلك، نتناول موضوع النماذج غير الخطية (Nonlinear Models)، حيث نقدم بعض الأمثلة على النماذج التي لا تخضع لافتراض الخطية، مع عرض طرق تقديرها وأبرز خصائصها. كما نخصص محورًا لدراسة مشاكل البيانات وأخطاء القياس، مركزين على أثر هذه المشاكل في تحريف نتائج النماذج الاقتصادية وكيفية التعامل معها للحد من آثارها السلبية.

في الأخير، نختم بمحور المعادلات الآنية، حيث نتطرق إلى النماذج التي تحتوي على علاقات متبادلة بين المتغيرات، ونشرح كيفية التعامل مع هذا النوع من النماذج الذي يعكس واقع التفاعلات الاقتصادية المعقدة.

جدير بنا أن نذكر أننا في تحرير هاته المطبوعة قد اعتمدنا بشكل أساسي على كتاب "Econométrie" لمؤلفه "Régis Bourbonnais"، لما يتميز به من منهجية دقيقة وشمولية وبساطة في عرض مبادئ الاقتصاد القياسي.

.

1. مفهوم الاقتصاد القياسي:

قد تتعد تعاريف الاقتصاد القياسي لكن يمكن أن نلخصها في كونه العلم الذي يهتم بإيجاد العلاقات التي تربط بين مجموعة من المتغيرات الاقتصادية من خلال تحديد صيغها الرياضية وتقدير معالمها واختبار كفاءتها، أين تحديد صيغها الرياضية يعتمد على مفهوم النماذج وتقدير معالمها يعتمد على نظرية الارتباط واختبار كفاءتها يعتمد على نظرية المعاينة. قبل عرض أساسيات الاقتصاد القياسي نعرج أولا على توضيح مفهوم النموذج والتذكير بنظرية الارتباط.

1.أ. مفهوم النموذج:

إن مفهوم النموذج يشمل جميع العلوم القابلة متغيراتها للقياس. إن التطور الكبير في القدرة على قياس المتغيرات الاجتماعية عامة والاقتصادية خاصة جعل منها ميدانا مهما في بناء النماذج بل تعداه إلى أن أصبحت مصدرا ومرجع ابداع بالنسبة للعلوم الأخرى.

1.أ.1. تعريف نموذج الاقتصاد القياسي:

من الصعب كذلك إعطاء تعريف واحد لمفهوم النموذج. في الاقتصاد القياسي يمكننا اعتبار النموذج كتمثيل لظاهرة على شكل معادلة أو مجموعة معادلات رياضية المعبر عنها عبر مجموعة من العلاقات بين المتغيرات. تهدف هاته المعادلات إلى تمثيل الخصائص الأساسية للحقيقة المراد تمثيلها. إذا، النموذج هو أداة للباحث أو المنمذج الذي يبحث عن فهم وشرح ظاهرة معينة الذي يقوم بطرح ووضع فرضيات واضحة وصريحة للعلاقات.

يمكن القول أيضا أن النموذج، دائما في سياق الاقتصاد القياسي، هو عبارة عن تمثيل جزئي لحقيقة تكون في الواقع أكثر تعقيدا أين تكمن الصعوبة الكاملة لبناء نموذج في القدرة على تحديد واختيار عناصر الحقيقة أو الإشكالية الأكثر أهمية، هذا الاختيار يكون مرتبط أساسا بطبيعة الإشكالية وشكل القرار المراد تفعيله.

1.أ.2. بناء نموذج الاقتصاد القياسي:

في العلوم الاجتماعية عامة وفي العلوم الاقتصادية خاصة، تكون الظواهر المدروسة متعلقة بفهم طبيعة عمل الأنظمة الاقتصادية أين يهدف المنمذج، في إطار الاقتصاد القياسي، إلى مساعدة الأعوان الاقتصاديين (الأفراد، العائلات، المؤسسات،

الحكومات،...إلخ) على التدخل بطريقة أكثر فعالية في مواجهة المشكلات الاقتصادية. على هذا الأساس فإن بناء نموذج فعال يتم وفق عدد من المراحل المتسلسلة والمرتبطة والمهمة أين يؤدي ضعف مرحلة أو حلقة إلى ضعف فعالية النموذج أو حتى إلى عدم صحته. نلخص المراحل الأساسية التي يمر بها بناء نموذج قياسي:

- الاعتماد أو البناء على نظرية: يعتمد بناء نموذج على نظرية معبر عنها بمجموعة من الفرضيات،
- اختيار المتغيرات المحددة أو المفسرة للإشكالية،
- قياس وتكميم متغيرات الدراسة أين يتم ذلك عبر جمع بيانات متغيرات الدراسة بالطريقة المناسبة،
- تشخيص وصياغة العلاقات بين المتغيرات وتحديد شكلها الرياضي،
- بناء النموذج القياسي وتقدير معالمه،
- مرحلة التحقق من صحة النموذج تمثل آخر محطة في بناء نموذج أين تطرح تساؤلات من الشكل: ما مدى صحة العلاقات المشخصة؟ هل بالإمكان تحسين معالم النموذج؟ هل النموذج صحيح ومقبول على طول الفترة المدروسة؟....إلخ

نشير إلى أنه توجد أشكال عديدة للبيانات التي تتحدد وفق طبيعة الظاهرة أو الإشكالية المراد نمذجتها. نميز أهمها:

أ‌- **بيانات زمنية**: أو بيانات سلاسل زمنية التي تمثل الشكل الأكثر مصادفة في بناء النماذج الاقتصادية والتي تكون عبارة عن مشاهدات للمتغير خلال فترات زمنية ثابتة،

ب‌- **بيانات المقاطع الآنية**: تكون البيانات مشاهدة في نفس الزمن (آنيا) لكنها متعلقة بمجموعة من الوحدات الإحصائية (مشاهدة معدل النمو لسنة 2024 لمجموعة من الدول، بيانات الإنفاق لشهر جانفي 2024 لعائلات ولاية تيارت، ....إلخ)،

ت‌- **البيانات المقطعية**: يمزج هذا الشكل من البيانات بين البيانات الزمنية وبيانات المقاطع الآنية أين تكون المتغيرات مشاهدة خلال فترات زمنية ثابتة بالنسبة لمجموعة من الوحدات الإحصائية (مشاهدة معدلات النمو السنوي خلال الفترة الممتدة من 2010 إلى 2024 لمجموعة من الدول، بيانات الإنفاق الشهري لسنة 2024 لعائلات ولاية تيارت، .... إلخ).

ملاحظة: في حالة بناء نماذج ذات بيانات زمنية قد تكون العلاقة بين المتغيرات غير متزامنة أي قد يؤثر سلوك (قيمة) متغير أثناء فترة معينة على متغير آخر في فترة لاحقة ونقول إن هناك تأخير زمني في العلاقة أو النموذج (نماذج ذو تأخير زمني).

1.ب. دور الاقتصاد القياسي أو النماذج القياسية:

يتأرجح دور الاقتصاد القياسي بين كونه أداة إثبات لصحة النظريات والفرضيات أو أداة تحقيق وتحليل.

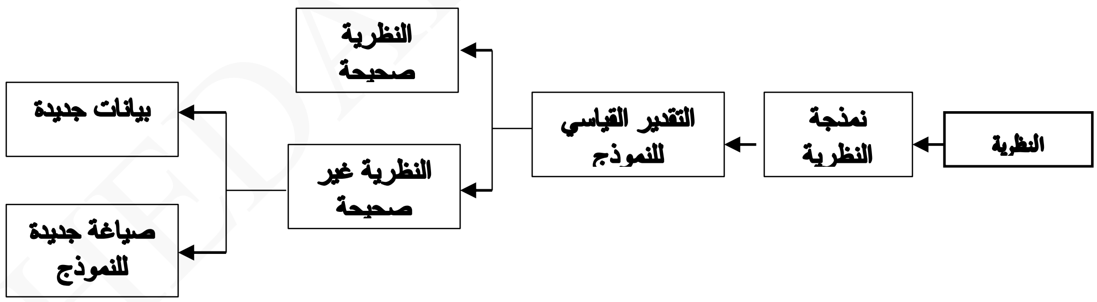
1.ب.1. الاقتصاد القياسي كأداة إثبات صحة النظريات:

الاقتصاد القياسي هو أداة متاحة للاقتصادي تسمح له بنفي أو إثبات النظريات والفرضيات التي يصيغها. الاقتصادي يقدم فرضيات وعلاقات والاقتصاد القياسي يقدم تقديرات لمعالم العلاقات مع القيام بالاختبارات الإحصائية والدقة الموافقة.

السؤال الذي قد يطرح: لماذا تقدر هاته العلاقات وتختبر إحصائيا؟

هناك العديد من المبررات وراء ذلك، كتوضيح العلاقات البينية حيث قد تؤدي الثقة المفرطة لحدس الاقتصادي إلى جهل بوجود علاقات مهمة أو إلى سوء استعمالها. أيضا، العلاقات الهامشية والمفسرة للظاهرة يجب تقديرها واختبارها. كذلك، من الضروري بقدر تقدير العلاقات تقدير وقياس درجة دقتها وموثوقيتها حيث استعمال الطرق الكيفية والوصفية لوحدها يكون غير كاف لقياس جودة وصحة العلاقة أو النموذج.

الشكل التالي يوضح مراحل اثبات صحة النظرية عن طريق الاقتصاد القياسي:



الشكل -1- النظرية الاقتصادية والاقتصاد القياسي

1.ب.2. الاقتصاد القياسي كأداة تحقيق أو تحليل:

الاقتصاد القياسي ليس فقط أداة إثبات بل أيضا أداة تحقيق وتحليل من خلال:

- إظهار العلاقات غير جلية مبدئيا بين المتغيرات،
- الاستقراء الاحصائي من خلال تعميم خصائص عينة إحصائية على المجتمع الاحصائي،

- محاكاة تأثير متغير على متغيرات أخرى،
- التنبؤ الذي يستعمل من طرف المؤسسات والسلطات العمومية بغرض استباق التغيرات المستقبلية.

سوف نحاول جاهدين من خلال ما يلي ابراز أهم استعمالات القياس الاقتصادي في ظروف وأهداف مختلفة.

1.ت. نظرية الارتباط:

عندما تكون لظاهرتين سلوك أو تطور مشترك نقول عنهما أنهما مرتبطتين. الارتباط البسيط يقيس درجة الارتباط الموجود بين الظاهرتين والمعبر عنهما بمتغيرين إحصائيين. عندما نتحدث عن الارتباط بين أكثر من ظاهرتين أو متغيرين فإننا نتحدث عن الارتباط المتعدد. كذلك نميز بين الارتباط الخطي والارتباط غير خط وبين الارتباط الموجب (علاقة طردية بين المتغيرات) والارتباط السالب (علاقة عكسية بين المتغيرات) أين تستعمل التمثيلات البيانية غالبا لتحديد نوع الارتباط بين المتغيرات.

1.ت.1. قياس ومحدودية معامل الارتباط الخطي:

إن التمثيل البياني يعطي فقط انطباع عن نوع الارتباط بين المتغيرات دون إعطاء صورة دقيقة كمية عن شدة وقوة الارتباط. لأجل ذلك وفي حالة دراسة الارتباط الخطي بين متغيرين X و Y نقوم بحساب إحصائية، تسمى بمعامل الارتباط الخطي البسيط $r_{X,Y}$، والذي يعطى بالعلاقة التالية:

$$r_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})\,(Y_i-\bar{Y})}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2}\,\sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i-\bar{Y})^2}} \dots\dots(1.1)$$

أين:

$Cov(X,Y)$: يمثل التباين المشترك بين المتغيرين X و Y،

σ_X و σ_Y: يمثلان على التوالي الانحرافين المعياريين للمتغيرين X و Y،

N: تمثل حجم المجتمع الكلي.

يبرهن على أن $\left(-1 \leq r_{X,Y} \leq 1\right)$، بحيث يقترب معامل الارتباط من "1" كلما زادت قوة الارتباط الموجب (الطردي) بين المتغيرين ويقترب من "-1" كلما زادت قوة الارتباط السالب (العكسي) الخطي بينهما. بينما يقترب من "0" في حالة غياب ارتباط خطي بينهما.

في العلوم الاقتصادية، من النادر أن يقترب معامل الارتباط من القيم الثلاث هاته مما يجعل من الصعوبة إعطاء تفسير دقيق له. يضاف إلى ذلك أن حسابه يتم في الغالب من خلال عينة إحصائية ذات حجم n $(n < N)$. نرمز ب $\hat{r}_{X,Y}$ لمعامل الارتباط الخطي البسيط المحسوب من العينة والذي يمثل تقدير لقيمته الحقيقية. إن نظرية المعاينة تسمح بالحصول على تقدير غير متحيز ومتقارب كما تسمح أيضا بوضع مجال ثقة له.

كون معامل الارتباط المحسوب من العينة هو متغير عشوائي، نقوم دائما باختبار معنوية معامل الارتباط الحقيقي من خلال اختبار الفرضية: $H_0: r_{X,Y} = 0$ مقابل الفرضية $H_1: r_{X,Y} \neq 0$.

للقيام بالاختبار يمكننا قبل ذلك إثبات أنه عند تحقق الفرضية H_0 فإن المتغير:

$$\frac{\hat{r}_{X,Y}}{\sqrt{\frac{\left(1 - \hat{r}_{X,Y}^2\right)}{n-2}}}$$

يتبع قانون ستودنت بدرجة حرية "$n-2$". إذا للقيام بالاختبار نحسب الإحصائية:

$$t^* = \frac{\left|\hat{r}_{X,Y}\right|}{\sqrt{\frac{\left(1 - \hat{r}_{X,Y}^2\right)}{n-2}}} \dots\dots (1.2)$$

ونقارنها بقيمة ستودنت المجدولة عند درجة حرية "$n-2$" ($t_{(n-2)}^{1-(\alpha/2)}$). عند العتبة α، يكون القرار على النحو التالي:

إذا كانت $\left(t^{*} > t_{(n-2)}^{1-(\alpha/2)}\right)$ فإننا نرفض الفرضية H_0، أي بقبول وجود ارتباط خطي معنوي بين المتغيرين X و Y. في الحالة العكسية، $\left(t^{*} \leq t_{(n-2)}^{1-(\alpha/2)}\right)$، فإننا نقبل الفرضية H_0، أي أنه لا يوجد ارتباط خطي معنوي بين المتغيرين X و Y.

1.ت.2. حدود معامل الارتباط الخطي:

نلخص حدود معامل الارتباط الخطي في النقطتين التاليتين.

- الارتباط المختبر هو الارتباط الخطي: إن استعمال العلاقتين (1.1) و (2.1) يسمح فقط بحساب واختبار وجود ارتباط خطي بين المتغيرين X و Y. قيمته المعدومة تشير فقط إلى معدومية التباين المشترك بينهما حيث قد يكونا المتغيرين مرتبطين لكن الاحصائية $r_{X,Y}$ هي معدومة. المثال الأكثر تداولا لتبيين ذلك هو الارتباط الدائري التام بين متغيرين X و Y و المعبر عنه بالمعادلة أو النموذج "$(X - x_1)^2 + (Y - y_1)^2 = h^2$" أين يكون معامل الارتباط الخطي $r_{X,Y}$ معدوم. لتجاوز هاته المحدودية، يكون من الأفضل أحيانا حساب الارتباط الخطي بين متغيرين بعد إدخال اللوغاريتم على قيمهما.
- الارتباط لا يعني السببية: وجود معامل ارتباط قوي بين متغيرين لا يعني بالضرورة وجود سببية أو علاقة حقيقية بينهما، قد لا يتعدى كونه ارتباط إحصائي فقط.

2. نموذج الانحدار الخطي البسيط:

نموذج الانحدار الخطي البسيط هو أبسط النماذج القياسية أين يحتوي على متغير تابع داخلي (مفسر) مفسر بمتغير خارجي (مستقل) واحد ومعبر عن العلاقة بينهما بعلاقة خطية بسيطة.

2.أ. تقديم النموذج:

لتبسيط المفاهيم المتعلقة بنموذج الانحدار الخطي البسيط نستعرض فيما يلي مثال توضيحي.

2.أ.1. مثال توضيحي:

لتكن دالة الاستهلاك الكينزية:

$$C = a_0 + a_1 Y$$

أين:

C: يمثل متغير الاستهلاك،

Y: يمثل متغير الدخل،

a_0: يمثل الاستهلاك التلقائي،

a_1: يمثل نسبة الاستهلاك الهامشية.

1.أ. مصطلحات:

- متغير الاستهلاك يسمى المتغير "المفسر" أو المتغير "التابع" أو المتغير "الداخلي".
- متغير الدخل يسمى المتغير "المفسر" أو المتغير "المستقل" أو المتغير "الخارجي".
- a_0 و a_1 يسميان بمعالم النموذج أو معالم الانحدار.

1.ب. التأشير:

التأشير متعلق بنوع البيانات أين يمكننا التمييز بين 3 أنواع من تأشير النماذج:

- التأشير الزمني، متعلق بالبيانات الزمنية، أين المتغيرات تمثل ظواهر تمت مشاهدتها على فترات زمنية منتظمة، مثل الاستهلاك والدخل السنوي خلال 20 سنة لبلد معين. يكتب النموذج حسب هذا التأشير:

$$C_t = a_0 + a_1 Y_t \quad t = \overline{1;20}$$

أين:

C_t: يمثل الاستهلاك في الزمن t،

Y_t: يمثل الدخل في الزمن t.

- التأشير المقطعي، متعلق ببيانات المقاطع الآنية، أين المتغيرات تمثل ظواهر تمت مشاهدتها في آن واحد من أجل العديد من الأفراد، مثلا الاستهلاك والدخل ل 20 بلد خلال سنة 2020. النموذج يكتب حسب هذا التأشير:

$$C_i = a_0 + a_1 Y_i \quad i = \overline{1;20}$$

أين:

C_i: يمثل استهلاك البلد i،

Y_i: يمثل دخل البلد i.

- التأشير المختلط، متعلق بالبيانات المقطعية، أين المتغيرات تمت مشاهدتها لفترات زمنية منتظمة ومن أجل وحدات إحصائية متعددة، مثلا الاستهلاك والدخل ل 20 بلدا خلال الفترة 2020 إلى 2024، النموذج يكتب بصفة عامة حسب هذا التأشير:

$$C_{it} = a_{0i} + a_{1i} Y_{it} \quad i = \overline{1;20} \quad t = \overline{1;5}$$

1.ت. دور المتغير أو العنصر العشوائي:

النموذج بالطريقة التي تم تشخيصه بها ما هو إلا تمثيل كاريكاتوري للحقيقة. في الواقع الأخذ فقط بالدخل لتفسير الاستهلاك هو غير كاف، توجد متغيرات أخرى قد تساهم في تفسير الاستهلاك. لأجل ذلك نقوم عند تشخيص النموذج بزيادة العنصر (المتغير) (ε_t) الذي يلخص مجموعة المعلومات غير مدرجة في النموذج ليصبح النموذج "$C_t = a_0 + a_1 Y_t + \varepsilon_t$" في حالة تأشير السلاسل الزمنية أو من الشكل "$C_i = a_0 + a_1 Y_i + \varepsilon_i$" في حالة تأشير المقاطع الآنية، أين الخطأ ε_t (ε_i) يمثل **خطأ تشخيص النموذج**، أي مجموعة الظواهر والمتغيرات المفسرة للاستهلاك غير مرتبطة بالدخل. العنصر ε_t يقيس الفرق بين القيم الحقيقية المشاهدة ل C_t والقيم التي كانت ستشاهد في حالة تشخيص تام للنموذج. بشكل عام، العنصر ε_t يقيس ثلاثة أشكال من الأخطاء:

- خطأ التشخيص، الناتج عن كون المتغير المفسر غير كاف لتفسير كلي للظاهرة،
- خطأ القياس، الناتج عن كون البيانات لا تمثل بدقة الظاهرة،
- خطأ تغير البيانات من عينة لأخرى.

1.ث. تأثير العنصر العشوائي ε_t:

عند تقدير نموذج الاستهلاك السابق، القيم الحقيقية للمعلمتين a_0 و a_1 لا نعرفهما بدقة لوجود أشكال الأخطاء الثلاثة، بل نعرف تقديرات لها، يرمز لها على التوالي ب $\hat{a}_0$ و $\hat{a}_1$ والتي تكون عبارة عن متغيرات عشوائية لها قوانين احتمالات مشتقة من القانون الاحتمالي للعنصر العشوائي ε_t، لأنها، كما سنرى لاحقا، دوال تابعة له. هاته الخاصية تسمح بوضع مجالات ثقة للمعالم الحقيقية المجهولة والقيم التنبئية.

2.ب. تقدير معالم النموذج:

من مراحل بناء نموذج قياسي المهمة تقدير معالمه الذي يعتمد على طرق رياضية بحتة ومجموعة من الفرضيات.

2.ب.1. النموذج والفرضيات:

ليكن النموذج التالي:

$$y_t = a_0 + a_1 x_t + \varepsilon_t \quad t = \overline{1, n}$$

أين:

y_t: يمثل المتغير التابع في الزمن t ،

x_t: يمثل المتغير المستقل في الزمن t،

a_0 و a_1 : تمثلان معلمتي النموذج،

ε_t: يمثل خطا التشخيص (الفرق بين النموذج الحقيقي والنموذج المشخص)، هذا الخطأ هو مجهول ويبقى مجهول[1]،

n: تمثل عدد المشاهدات أو حجم العينة.

الفرضيات:

عند تقدير نماذج القياس الاقتصادي نحتاج إلى مجموعة من الفرضيات التي تسمح بالحصول على نماذج فعالة والقيام بالاختبارات الإحصائية اللازمة عنها. إن عدم تحقق هاته الفرضيات إما قد يرهن جودة النموذج وفعاليته أو يجعل منه نموذج غير صحيح. يعبر عنها بالعناصر التالية:

- H_1: النموذج في الأساس خطي بالنسبة ل x_t أو من أجل أي تحويل رياضي ل x_t،
- H_2: قيم المتغير المستقل x_t تمت مشاهدتها بدون خطأ (x_t غير عشوائي)،
- H_3: $E(\varepsilon_t) = 0$، الأمل الرياضي لخطأ التشخيص هو معدوم، أي أن النموذج هو مشخص جيدا في المتوسط ،
- H_4: $E(\varepsilon_t^2) = \sigma^2$، تباين الأخطاء هو ثابت، أي مجال تضاعفه هو نفسه مهما كانت الفترة أو الوحدة الإحصائية، تسمى هاته الفرضية بفرضية ثبات أو تجانس تباين الأخطاء،

[1] يبقى مجهول لأننا سنرى فيما بعد أن القيم الحقيقة للمعلمتين a_0 و a_1 هي غير معلومة (مقدرة) .

- H_5: $E(\varepsilon_t, \varepsilon_{t'}) = 0$، لما $t \neq t'$، الأخطاء هي غير مرتبطة (أو مستقلة)، أي خطأ في الزمن t ليس له تأثير على آخر في الزمن t'،
- H_6: $COV(x_t, \varepsilon_t) = 0$، الخطأ هو مستقل عن المتغير المستقل.

سنرى في المحاور اللاحقة نتائج عدم تحقق هاته الفرضيات وكيفية التعامل مع ذلك.

2.ب.2. صياغة المقدرات:

تقدير المعلمتين a_0 و a_1 يتحصل عليه بتصغير مجموع الأبعاد مربعة بين كل مشاهدة y_t والقيمة التقديرية على المستقيم، تسمى طريقة التقدير هاته بطريقة المربعات الصغرى (GLS).

رياضيا يتم ذلك بحل:

$$Min \sum_{t=1}^{n} \varepsilon_t^2 = Min \sum_{t=1}^{n} (x_t - a_0 - a_1 y_t)^2 = Min\ S$$

بعدم مشتقتي S بالنسبة ل a_1 و a_0 نتحصل على:

$$\begin{cases} \dfrac{\delta S}{\delta a_1} = 0 \\ \dfrac{\delta S}{\delta a_0} = 0 \end{cases} \Rightarrow \begin{cases} \dfrac{\delta \sum_{t=1}^{n} (x_t - a_0 - a_1 y_t)^2}{\delta a_1} = 0 \\ \dfrac{\delta \sum_{t=1}^{n} (x_t - a_0 - a_1 y_t)^2}{\delta a_0} = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \sum_t x_t y_t - \hat{a}_0 \sum_t x_t - \hat{a}_1 \sum_t x_t^2 = 0 \\ \sum_t y_t - n\hat{a}_0 - \hat{a}_1 \sum_t x_t = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \hat{a}_1 = \dfrac{\sum_{t=1}^{n} (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} & \ldots\ldots (2.1) \\ \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} & \ldots\ldots (2.2) \end{cases}$$

يمكننا وضع ملاحظتين:

- كتابة النموذج ليست عشوائية: التشخيص $y_t = a_0 + a_1 x_t + \varepsilon_t$ لا يكافئ التشخيص $x_t = a'_0 + a'_1 y_t + \varepsilon_t$. في التشخيص الأول x_t يسبب y_t والعكس في التشخيص الثاني. يمكننا من خلال العلاقة (1.2) استنتاج العلاقة: $\hat{a}_1 \times \widehat{a'}_1 = P^2_{X,Y}$.
- المعلمة a_1 تمثل ميل المستقيم أو النسبة الهامشية.

لدينا: $\Delta y_t = \hat{a}_1 \Delta x_t$. تأثير تغير في x_t يقاس مباشرة على y_t من خلال المعلمة $\hat{a}_1$.

2.ب.3. حالة خاصة: نموذج بدون حد ثابت.

تقر النظرية الاقتصادية أحيانا بعلاقات يكون فيها $(a_0 = 0)$، كما هو الحال بالنسبة لدالة الإنتاج لمنتوج صناعي أين عامل إنتاج معدوم يؤدي إلى كمية إنتاج معدومة. في هاته الحالة تقدير المعلمة a_1 يعطى بالعلاقة:

$$\hat{a}_1 = \frac{\sum_{t=1}^{n} x_t\, y_t}{\sum_{t=1}^{n} {x_t}^2} \ldots\ldots (2.3)$$

نلاحظ أنها تطبيق للعلاقة (1.2) مع $(\bar{x} = \bar{y} = 0)$. كذلك في حالة متغيرات مركزة حول متوسطها نطبق العلاقة (3.2) لتقدير a_1 ويكون $(a_0 = 0)$ كون $(\bar{x} = \bar{y} = 0)$.

2.ب.4. الكتابات المختلفة للنموذج (الخطأ والباقي):

نموذج الانحدار الخطي البسيط يمكن أن يكتب وفق صيغتين تبعا إن كان نموذج نظري مشخص من طرف الاقتصادي أو نموذج مقدر من خلال عينة مشاهدات إحصائية.

- نموذج نظري مشخص من طرف الاقتصادي مع ε_t تمثل الخطأ المجهول:

$$y_t = a_0 + a_1 x_t + \varepsilon_t$$

- نموذج مقدر من خلال عينة مشاهدات إحصائية مع e_t تمثل الباقي:

$$y_t = \hat{a}_0 + \hat{a}_1 x_t + e_t = \hat{y}_t + e_t$$

الباقي المشاهد e_t هو إذا الفرق بين القيم المشاهدة للمتغير التابع والقيم المقدرة من خلال النموذج المقدر، يمكننا أيضا الكتابة:

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$$

2.ت. خصائص المعالم المقدرة:

لدينا:

$$y_t = a_0 + a_1 x_t + \varepsilon_t \dots\dots (2.4)$$

ومنه

$$\bar{y} = a_0 + a_1 \bar{x} + \bar{\varepsilon} \dots\dots (2.5)$$

أيضا

$$(2.4) - (2.4) = (y_t - \bar{y}) = a_1 (x_t - \bar{x}) + (\varepsilon_t - \bar{\varepsilon})$$

بتعويض قيمة $(y_t - \bar{y})$ في العلاقة (1.2) نتحصل على:

$$\hat{a}_1 = a_1 + \frac{\sum_{t=1}^{n} (x_t - \bar{x})\,(\varepsilon_t - \bar{\varepsilon})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \dots\dots (2.6)$$

- هل تقديرات المعالم هي غير متحيزة؟

العلاقة (6.2) تصبح:

$$\hat{a}_1 = a_1 + \frac{\sum_{t=1}^{n}(x_t - \bar{x})\,\varepsilon_t}{\sum_{t=1}^{n}(x_t - \bar{x})^2} - \frac{\sum_{t=1}^{n}(x_t - \bar{x})\bar{\varepsilon}}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

$$\hat{a}_1 = a_1 + \frac{\sum_{t=1}^{n}(x_t - \bar{x})\,\varepsilon_t}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \qquad \ldots\ldots(2.7)$$

ومنه

$$E(\hat{a}_1) = a_1 + \frac{\sum_{t=1}^{n}(x_t - \bar{x})\,E(\varepsilon_t)}{\sum_{t=1}^{n}(x_t - \bar{x})^2} - \frac{\sum_{t=1}^{n}(x_t - \bar{x})E(\bar{\varepsilon})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

إذا

$$E(\hat{a}_1) = a_1$$

كون فرضيا $\left(E(\bar{\varepsilon}) = E(\sum_{1}^{n}\varepsilon_t)\right)$ و $E(\varepsilon_t)$ هما معدومين.

بنفس الطريقة نبرهن أن:

$$E(\hat{a}_0) = a_0$$

لدينا (كون $\bar{y} = \bar{\hat{y}}$ ، انظر البرهان أسفله):

$$\begin{cases} \bar{y} = \hat{a}_0 + \hat{a}_1\bar{x} \\ \bar{y} = a_0 + a_1\bar{x} + \bar{\varepsilon} \end{cases} \Rightarrow \hat{a}_0 = a_0 + \bar{\varepsilon} - (\hat{a}_1 - a_1)\bar{x} \qquad \ldots\ldots(2.7')$$

ومنه

$$E(\hat{a}_0) = a_0 + E(\bar{\varepsilon}) - (E(\hat{a}_1) - a_1)\bar{x}$$

ومنه

$$E(\hat{a}_0) = a_0$$

ومنه تقديري المعلمتين هما تقديرين غير متحيزين.

(نذكر أن الأمل الرياضي يدخل فقط على المتغيرات العشوائية أم القيم الثابتة فأملها هي قيمتها نفسها)

- هل تقديرات المعالم هي مقاربة؟

لأن التقديرين غير متحيزين، يكفي ليكونا متقاربين:

$$\lim_{n\to\infty} V(\hat{a}_0) = \lim_{n\to\infty} V(\hat{a}_1) = 0$$

أين

$V(\hat{a}_0)$ و $V(\hat{a}_1)$ هما تبايني المقدرات.

لدينا

$$V(\hat{a}_1) = E\left(\hat{a}_1 - E(\hat{a}_1)\right)^2 = E(\hat{a}_1 - a_1)^2$$

وحسب العلاقة (7.2):

$$V(\hat{a}_1) = E\left(\frac{\sum_{t=1}^{n}(x_t - \bar{x})\,\varepsilon_t}{\sum_{t=1}^{n}(x_t - \bar{x})^2}\right)^2 = E\left(\sum_{t=1}^{n} w_t\,\varepsilon_t\right)^2$$

$$= E\left(\sum_t w_t^2 \varepsilon_t^2 + 2\sum_{t<t'} w_t w_{t'} \varepsilon_t \varepsilon_{t'}\right)$$

مع

$$w_t = \frac{(x_t - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

ويصبح:

$$V(\hat{a}_1) = \sum_t w_t^2 E(\varepsilon_t^2) + 2 \qquad \sum_t \underset{(t'>t)}{} \sum_{t'} w_t w_{t'} E(\varepsilon_t \varepsilon_{t'})$$

وحسب الفرضيتين H_4 و H_5 أعلاه يصبح:

$$V(\hat{a}_1) = \sum_t w_t^2 \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \sum_t w_t^2$$

$$V(\hat{a}_1) = \frac{\sigma_\varepsilon^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \dots \dots (2.8)$$

أين

$$\lim_{n\to\infty} V(\hat{a}_1) = \lim_{n\to\infty} \frac{\sigma_\varepsilon^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2} = 0$$

كون σ_ε^2 هو قيمة ثابتة و:

$$\lim_{n\to\infty} \sum_{t=1}^{n}(x_t - \bar{x})^2 = \infty$$

يمكننا الملاحظة من خلال تباين التقدير $\hat{a}_1$، العلاقة (8.2)، أن هذا التقدير يكون أكثر دقة (تباين صغير) كلما كانت قيمة $\sum_{t=1}^{n}(x_t - \bar{x})^2$ كبيرة، الأمر الذي يتحقق في الحالتين:

- عدد مشاهدات (حجم العينة) كبير كفاية،
- قيم المتغير المستقل (الخارجي) x_t تكون جد متشتتة حول متوسطها.

برهان مماثل بالنسبة لتباين $\hat{a}_0$ يعطينا:

$$V(\hat{a}_0) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2}\right) \dots \dots (2.9)$$

لدينا

$$V(\hat{a}_0) = E\big(\hat{a}_0 - E(\hat{a}_0)\big)^2 = E(\hat{a}_0 - a_0)^2$$

وحسب العلاقة ('7.2):

$$V(\hat{a}_0) = E(\bar{\varepsilon} - (\hat{a}_1 - a_1)\bar{x})^2 = E(\bar{\varepsilon}^2 - 2\bar{\varepsilon}(\hat{a}_1 - a_1)\bar{x} + (\hat{a}_1 - a_1)^2\bar{x}^2)$$

$$= E(\bar{\varepsilon}^2) - 2\bar{x}E(\bar{\varepsilon})E(\hat{a}_1 - a_1) + \bar{x}^2E(\hat{a}_1 - a_1)^2$$

$$= E\left(\left(\frac{\sum_t \varepsilon_t}{n}\right)^2\right) + \bar{x}^2V(\hat{a}_1) = E\left(\frac{(\sum_t \varepsilon_t)^2}{n^2}\right) + \bar{x}^2V(\hat{a}_1)$$

$$= \frac{\sum_t E({\varepsilon_t}^2) + \sum_t \sum_{t'\ (t<t')} E(\varepsilon_t\varepsilon_{t'})}{n^2} + \bar{x}^2V(\hat{a}_1)$$

$$= \frac{\sigma_\varepsilon^2}{n} + \bar{x}^2\frac{\sigma_\varepsilon^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2} = \sigma_\varepsilon^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2}\right)$$

أين يكون:

$$\lim_{n\to\infty} V(\hat{a}_0) = 0$$

يمكننا ملاحظة أن:

$$V(\hat{a}_0) = \frac{\sigma_\varepsilon^2}{n} + \bar{x}^2\ V(\hat{a}_1) \ldots\ldots (2.9')$$

و

$$COV(\hat{a}_0, \hat{a}_1) = -\bar{x}V(\hat{a}_1) \ldots\ldots (2.9'')$$

من أجل هاته الأخيرة، لدينا:

$$COV(\hat{a}_0, \hat{a}_1) = E\big((\hat{a}_0 - a_0)(\hat{a}_1 - a_1)\big)$$

$$COV(\hat{a}_0, \hat{a}_1) = E\big((\bar{\varepsilon} - (\hat{a}_1 - a_1)\bar{x})(\hat{a}_1 - a_1)\big)$$

$$= E(\bar{\varepsilon})\bar{x}E(\hat{a}_1 - a_1) - \bar{x}E(\hat{a}_1 - a_1)^2$$

ومنه

$$COV(\hat{a}_0, \hat{a}_1) = -\bar{x}V(\hat{a}_1)$$

نقول في النهاية أن تقديرات المعالم هي مقاربة.

2.ث. أثر فرضية طبيعية الأخطاء:

سوف نقوم الآن بإدخال فرضية جديدة تتعلق بطبيعية الأخطاء (تتبع التوزيع الطبيعي). هاته الفرضية ليست ضرورية للحصول على مقدرات غير متحيزة ومقاربة لكنها تسمح ببناء اختبارات إحصائية عليها واختبار صحة النموذج.

نعبر عن هاته الفرضية بكتابة:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

بداية، نبحث عن تقدير لتباين الأخطاء σ_ε^2.

الباقي يعطى ب:

$$e_t = y_t - \hat{y}_t = y_t - \hat{a}_1 x_t - \hat{a}_0$$

بدون تغيير، يمكننا كتابة هاته العبارة:

$$e_t = y_t - \hat{a}_1 x_t - \hat{a}_0 - \hat{a}_1 \bar{x} + \hat{a}_1 \bar{x}$$

أو أيضا

$$e_t = y_t - \hat{a}_1 \bar{x} - \hat{a}_0 - \hat{a}_1 x_t + \hat{a}_1 \bar{x}$$

$$e_t = y_t - \bar{y} - \hat{a}_1 (x_t - \bar{x})$$

أين

$$\bar{y} = \hat{a}_1 \bar{x} + \hat{a}_0$$

بتعويض y_t و $\bar{y}$ بعبارتيهما (4.2) و (5.2)، يصبح:

$$e_t = (a_1 - \hat{a}_1)(x_t - \bar{x}) + (\varepsilon_t - \bar{\varepsilon})$$

بتربيع طرفي هاته العلاقة وإدخال المجموع، نتحصل على:

$$\sum_t e_t^2 = (a_1 - \hat{a}_1)^2 \sum_t (x_t - \bar{x})^2 + \sum_t (\varepsilon_t - \bar{\varepsilon})^2 + 2(a_1 - \hat{a}_1) \sum_t (x_t - \bar{x})(\varepsilon_t - \bar{\varepsilon})$$

قبل أن نكمل، لدينا من العلاقة (6.2):

$$\sum_t (x_t-\bar{x})(\varepsilon_t-\bar{\varepsilon}) = -(a_1-\hat{a}_1)\sum_t (x_t-\bar{x})^2$$

التي نعوضها في العلاقة الأخيرة وبعد التبسيط نتحصل على:

$$\sum_t e_t^2 = \sum_t (\varepsilon_t-\bar{\varepsilon})^2 - (a_1-\hat{a}_1)^2 \sum_t (x_t-\bar{x})^2$$

بأخذ أملها الرياضي:

$$E\left(\sum_t e_t^2\right) = E\left(\sum_t (\varepsilon_t-\bar{\varepsilon})^2\right) - E\left((a_1-\hat{a}_1)^2\right)\sum_t (x_t-\bar{x})^2$$

نتفحص جزئي الطرف الأيمن من هاته العلاقة:

- $E(\sum_t(\varepsilon_t-\bar{\varepsilon})^2) = E(\sum_t(\varepsilon_t^2-2\varepsilon_t\bar{\varepsilon}+\bar{\varepsilon}^2)) = E(\sum_t\varepsilon_t^2 - 2\bar{\varepsilon}\sum_t\varepsilon_t + \sum_t\bar{\varepsilon}^2)$

$$= E\left(\sum_t \varepsilon_t^2 - 2\bar{\varepsilon}n\bar{\varepsilon} + n\bar{\varepsilon}^2\right) = E\left(\sum_t \varepsilon_t^2 - 2n\bar{\varepsilon}^2 + n\bar{\varepsilon}^2\right)$$

$$= E\left(\sum_t \varepsilon_t^2 - n\bar{\varepsilon}^2\right) = E\left(\sum_t \varepsilon_t^2 - \bar{\varepsilon}\sum_t \varepsilon_t\right)$$

$$= E\left(\sum_t \varepsilon_t^2 - \frac{(\sum_t \varepsilon_t)^2}{n}\right) = \left(\sum_t E(\varepsilon_t^2) - \frac{1}{n}E\left(\sum_t \varepsilon_t\right)^2\right)$$

نعلم أن

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2$$

ومنه

$$E\left(\sum_t(\varepsilon_t-\bar{\varepsilon})^2\right)=\left(n\sigma_\varepsilon^2-\frac{1}{n}E(\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_n)^2\right)$$

وحسب فرضية استقلالية الأخطاء، نتحصل على:

$$E\left(\sum_t(\varepsilon_t-\bar{\varepsilon})^2\right)=\ n\sigma_\varepsilon^2-\frac{1}{n}n\sigma_\varepsilon^2=(n-1)\sigma_\varepsilon^2$$

بالنسبة للجزء الثاني:

- $E((a_1-\hat{a}_1)^2)\sum_t(x_t-\bar{x})^2=\sigma_\varepsilon^2$

وذلك حسب العلاقة (8.2).

نجد في النهاية:

$$E\left(\sum_t e_t^2\right)=(n-1)\sigma_\varepsilon^2-\sigma_\varepsilon^2=(n-2)\sigma_\varepsilon^2$$

أي

$$\sigma_\varepsilon^2=\frac{1}{(n-2)}E\left(\sum_t e_t^2\right)$$

بالرمز لتقدير σ_ε^2 ب $\hat{\sigma}_\varepsilon^2$، للحصول على تقدير غير متحيز، العلاقة الأخيرة تستلزم:

$$\hat{\sigma}_\varepsilon^2=\frac{1}{(n-2)}\sum_t e_t^2\ \ldots\ldots(2.10)$$

بتعويض تباين الأخطاء في المعادلتين (8.2) و (9.2) بتقديره المعطى بالعلاقة (10.2) نتحصل على:

$$\begin{cases} \hat{V}(\hat{a}_1) = \hat{\sigma}^2_{\hat{a}_1} = \dfrac{\hat{\sigma}^2_\varepsilon}{\sum_{t=1}^n (x_t - \bar{x})^2} \\ \hat{V}(\hat{a}_0) = \hat{\sigma}^2_{\hat{a}_0} = \hat{\sigma}^2_\varepsilon \left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum_{t=1}^n (x_t - \bar{x})^2}\right) \end{cases} \quad \dots\dots (2.11)$$

حيث فرضية طبيعية الأخطاء تستلزم أن المتغيرين:

$$\frac{\hat{a}_0 - a_0}{\sigma_{\hat{a}_0}} \text{ و } \frac{\hat{a}_1 - a_1}{\sigma_{\hat{a}_1}}$$

يتبعان قانون التوزيع الطبيعي المعياري $N(0,1)$.

لدينا أيضا من (10.2):

$$\frac{\sum_t^n e_t^2}{\sigma_\varepsilon^2} = (n-2)\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2_{(n-2)}$$

بحيث $\chi^2_{(n-2)}$ يمثل التوزيع الاحتمالي ل كي-دو (*Chi-deux*) ذو درجة حرية $(n-2)$.

يمكننا ملاحظة أن $\sum_{t=1}^n (x_t - \bar{x})^2 \frac{\sum_t^n e_t^2}{\sigma_{\hat{a}}^2} = (n-2)\frac{\hat{\sigma}_{\hat{a}}^2}{\sigma_{\hat{a}}^2} = (n-2)\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}$ التي تتبع بدورها توزيع $\chi^2_{(n-2)}$

الأمر الذي يستلزم أن

$$\left(\frac{\hat{a}_0 - a_0}{\hat{\sigma}_{\hat{a}_0}} \text{ و } \frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}}\right) \sim t_{(n-2)}$$

أي يتبعان توزيع ستودنت ذو درجة حرية $(n-2)$.

في الواقع لدينا:

$$\frac{\hat{a}_i - a_i}{\hat{\sigma}_{\hat{a}_i}} = \frac{\dfrac{\hat{a}_i - a_i}{\sigma_{\hat{a}_i}}}{\sqrt{(n-2)\dfrac{\hat{\sigma}^2_{\hat{a}_i}}{\sigma^2_{\hat{a}_i}}\dfrac{1}{(n-2)}}} \quad i = 1,2$$

الذي يمثل متغير إحصائي حاصل قسمة متغير طبيعي معياري على الجذر التربيع لمتغير كي–دو مقسوم على درجة حريته الذي يمثل تعريف متغير يتبع توزيع ستودنت بدرجة حرية $(n-2)$.

هاته التعاريف تسمح لنا الآن بوضع اختبار إحصائيات تسمح لنا بالإجابة عن إشكاليات من الشكل:

- مقارنة معلمة نموذج انحدار بقيمة معينة،
- مقارنة معلمتي انحدار تم حسابهما من عينتين مختلفتين،
- وضع مجال ثقة لمعالم نموذج انحدار...

2.ج. معادلة وجدول تحليل التباين:

معادلة وجدول تحليل التباين هي أدوات إحصائية تسمح لنا باختبار جودة النموذج والقيام أيضا ببعض الاختبارات الإحصائية على معالمه.

2.ج.1. معادلة تحليل التباين:

نبرهن أولا العلاقتين التاليتين:

- $\sum_t^n e_t = 0$: مجموع البواقي معدوم (معادلة الانحدار تمر بالنقطة المتوسطة حيث يكون هذا صحيحا فقط من أجل النماذج التي تحوي حد ثابت أو تلك التي بياناتها ممركزة حول متوسطها، البرهان بين ذلك).

لدينا:

$$y_t = \hat{y}_t + e_t \Rightarrow y_t = \hat{a}_0 + \hat{a}_1 x_t + e_t \Rightarrow \sum_t y_t = \sum_t \hat{a}_0 + \hat{a}_1 \sum_t x_t + \sum_t e_t$$

$$\Rightarrow \sum_t e_t = \sum_t y_t - n\hat{a}_0 - \hat{a}_1 \sum_t x_t \Rightarrow \frac{\sum_t e_t}{n} = \bar{y} - (\hat{a}_0 + \hat{a}_1 \bar{x})$$

لكن من العلاقة (2.2) لدينا $\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$

أي

$$\frac{\sum_t e_t}{n} = 0 \Rightarrow \sum_t e_t = 0$$

وهو المطلوب.

- ن $\sum_t y_t = \sum_t \hat{y}_t$: المساواة بين متوسطي المتغير المفسَر و المتغير المقدر.

لدينا:

$$y_t - \hat{y}_t = e_t \Rightarrow \sum_t y_t = \sum_t \hat{y}_t + \sum_t e_t = 0 \Rightarrow \sum_t y_t = \sum_t \hat{y}_t \Rightarrow \bar{y} = \bar{\hat{y}}$$

من هاتين العلاقتين المبرهنتين يمكننا استنتاج العلاقة الأساسية لتحليل التباين:

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t e_t^2 \; \ldots \ldots (2.12)$$

$$SCT \quad = \quad SCE \quad + \quad SCR$$

مجموع مربعات البواقي + مجموع مربعات التغيرات المفسرة = مجموع مربعات التغيرات الكلية

لدينا

$$y_t = \hat{y}_t + e_t \Rightarrow y_t - \bar{y} = \hat{y}_t - \bar{y} + e_t \Rightarrow (y_t - \bar{y})^2 = (\hat{y}_t - \bar{y} + e_t)^2$$

$$\Rightarrow (y_t - \bar{y})^2 = (\hat{y}_t - \bar{y})^2 + e_t^2 + 2e_t(\hat{y}_t - \bar{y})$$

$$\Rightarrow \sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t e_t^2 + 2\sum_t e_t(\hat{y}_t - \bar{y})$$

لدينا الطرف:

$$\sum_t e_t(\hat{y}_t - \bar{y}) = \sum_t e_t\hat{y}_t - \bar{y}\sum_t e_t$$

هاته العلاقة تسمح بقياس جودة انحدار نموذج. في الواقع، كلما كان مجموع مربعات التغيرات المفسرة قريبا من مجموع مربعات التغيرات الكلية كلما زادت جودة النموذج وزاد تمثيله للسحابة النقطية. في هذا السياق من المتداول لقياس جودة النموذج حساب النسبة:

$$R^2 = \frac{\sum_t(\hat{y}_t - \bar{y})^2}{\sum_t(y_t - \bar{y})^2} \ldots\ldots\ldots (2.13)$$

الذي يسمى "معامل التحديد" و R يسمى "معامل الارتباط المتعدد" (يمثل معامل الارتباط الخطي البسيط في حالة نموذج الانحدار الخطي البسيط بين x و y).

جدول تحليل التباين:

الجدول التالي يمثل تحليل التباين لنموذج انحدار خطي بسيط:

مصدر التغير	مجموع المربعات	درجة الحرية ddl	متوسط مجموع المربعات
x	$SCE = \sum_t(\hat{y}_t - \bar{y})^2$	1	$SCE/1$
البواقي	$SCR = \sum_t e_t^2$	$(n-2)$	$SCR/(n-2)$
الكلي	$SCT = \sum_t(y_t - \bar{y})^2$	$(n-1)$	

نعني بدرجة الحرية عدد المتغيرات الحرة التي يمكن أن تأخذ أي قيمة دون قيد، مثلا بالنسبة لمجموع مربعات التغيرات الكلية SCT

هو مجموع $(n-1)$ متغير عشوائي حر أين القيمة الأخيرة تقيد بالمتوسط $\bar{y}$ ، أي يمكن معرفة قيمته عند معرفة $\bar{y}$. لذلك نقول أن درجة حريته هي $(n-1)$.

اختبار الفرضية $H_0: a_1 = 0$ يكافئ اختبار الفرضية $H_0: SCE = 0$ (المتغير x_t لا يساهم في تفسير النموذج أو المتغير التابع).

ليكن اختبار الفرضية $H_0: SCE = 0$ مقابل الفرضية $H_0: SCE \neq 0$. إحصائية هذا الاختبار تعطى ب:

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{\frac{\sum_t (\hat{y}_t - \bar{y})^2}{1}}{\frac{\sum_t e_t^2}{(n-2)}} \dots\dots\dots (2.14)$$

أو

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{\frac{R^2}{1}}{\frac{(1-R^2)}{(n-2)}} \dots\dots\dots (2.15)$$

الإحصائية F^* هي النسبة بين مجموع المربعات المفسر ب x_t و مجموع مربعات البواقي، كل منهما مقسوم على درجة حريته الموافقة (بالقسمة نتحصل على تباينين). بحيث، إذا كان مجموع المربعات المفسر ب x_t أكبر من مجموع مربعات البواقي، المتغير x_t يعتبر متغير حقيقة مفسر.

الإحصائية F^* هي متغير إحصائي يتبع توزيع فيشر بدرجة حرية 1 و $(n-2)$. بحيث إذا كان $F^* > F^{\alpha}_{1,\ (n-2)}$ فإننا نرفض عند العتبة α الفرضية H_0، نقبل فرضية عدم المساواة بين التباينين H_1، أي أن x_t هو متغير حقيقة مفسر. في الحالة العكسية، نقبل فرضية مساواة التباينين، أي الفرضية H_0 ونقول أن المتغير x_t لا يفسر حقيقة المتغير التابع y_t أو لا يفسر الظاهرة.

لبرهنة العلاقة (14.2) أو التكافؤ بين الاختبارين أعلاه، لدينا استنادا للعلاقة (8.2) العلاقة:

$$\left(\frac{\hat{a}_1 - a_1}{\sigma_{\hat{a}_1}}\right)^2 = \frac{(\hat{a}_1 - a_1)^2}{\sigma_\varepsilon^2 / \sum_{t=1}^{n}(x_t - \bar{x})^2} \sim \chi_1^2$$

هي مربع متغير طبيعي معياري.

ولدينا العلاقة:

$$\frac{\sum_t e_t^2}{\sigma_\varepsilon^2} = (n-2)\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{(n-2)}^2$$

هي مجموع مربعات $(n-2)$ متغير طبيعي معياري.

بقسمة المتغيرين نتحصل على:

$$F^* = \frac{\dfrac{(\hat{a}_1 - a_1)^2}{\sigma_\varepsilon^2 / \sum_{t=1}^{n}(x_t - \bar{x})^2}}{(n-2)\dfrac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}} = \frac{(\hat{a}_1 - a_1)^2 \sum_{t=1}^{n}(x_t - \bar{x})^2}{\dfrac{\sum_t e_t^2}{(n-2)}}$$

حيث تحت فرضية تحقق $H_0: a_1 = 0$ يصبح:

$$F^* = \frac{\hat{a}_1^2 \sum_{t=1}^{n}(x_t - \bar{x})^2}{\dfrac{\sum_t e_t^2}{(n-2)}} = \frac{\dfrac{SCE}{1}}{\dfrac{SCR}{(n-2)}} \sim F_{1,\ (n-2)}^{\alpha}$$

كتكملة لدينا:

$$SCE = \sum_t (\hat{y}_t - \bar{y})^2 = \sum_t \left((\hat{a}_0 + \hat{a}_1 x_t) - (\hat{a}_0 + \hat{a}_1 \bar{x})\right)^2 = \hat{a}_1^2 \sum_{t=1}^{n}(x_t - \bar{x})^2$$

نلاحظ في النهاية أن:

$$F^* = \left(t^*_{\hat{a}_1}\right)^2 = \left(\frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}}\right)^2 = \frac{\hat{a}_1^2}{\hat{\sigma}_\varepsilon^2/\sum_{t=1}^n (x_t-\bar{x})^2} = \frac{\hat{a}_1^2\sum_{t=1}^n (x_t-\bar{x})^2}{\sum_t e_t^2/(n-2)}$$

2.ح. التنبؤ باستعمال نموذج الانحدار الخطي البسيط :

عندما تقدر معالم النموذج يكون بالإمكان حساب التنبؤ للأفق "h". ليكن النموذج المقدر:

$y_t = \hat{a}_0 + \hat{a}_1 x_t + e_t$، $t = \overline{1,n}$، إذا كانت قيمة المتغير المستقل (المفسر) معلومة عند $n+1$ (x_{n+1})، التنبؤ ل y_{n+1} يعطى ب

$$\hat{y}_{n+1} = \hat{a}_0 + \hat{a}_1 x_{n+1}$$

نبرهن أن هذا التنبؤ هو غير متحيز.

خطأ التنبؤ يعطى ب $e_{n+1} = y_{n+1} - \hat{y}_{n+1}$ الذي يمننا كتابته

$$e_{n+1} = (a_0 + a_1 x_{n+1} + \varepsilon_{n+1}) - (\hat{a}_0 + \hat{a}_1 x_{n+1}) = (a_0 - \hat{a}_0) - (a_1 - \hat{a}_1)x_{n+1} + \varepsilon_{n+1}$$

بإدخال الأمل الرياضي نجد

$$E(e_{n+1}) = 0$$

أي

$$E(\hat{y}_{n+1}) = y_{n+1}$$

ومنه $\hat{y}_{n+1}$ هو تقدير غير متحيز ل y_{n+1}.

بطريقة مماثلة يمكن البرهان أيضا أن:

$$E(\hat{y}_{n+h}) = y_{n+h}$$

إذا التنبؤ غير متحيز يتحصل عليه مباشرة باستعمال النموذج المقدر.

في الواقع، لا يكون من الأهمية معرفة القيمة التنبئية المتوقعة دون معرفة درجة الدقة أو الثقة التي نلحقها بها. حساب تباين خطأ التنبؤ يسمح لنا بتحديد مجال ثقة للقيمة التنبئية.

تباين خطأ التنبؤ يعطى ب:

$$V(e_{n+1}) = V\big((\hat{a}_0 - a_0) - (\hat{a}_1 - a_1)x_{n+1} + \varepsilon_{n+1}\big)$$

باستعمال خاصية تباين مجموع متغيرات عشوائية إضافة إلى كون a_0، a_0 و x_{n+1} هي ثوابت نجد:

$$V(e_{n+1}) = V(\hat{a}_0) + x_{n+1}^2 V(\hat{a}_1) + 2x_{n+1} COV(\hat{a}_0, \hat{a}_1) - 2COV(\hat{a}_0, \varepsilon_{n+1}) + 2x_{n+1} COV(\hat{a}_1, \varepsilon_{n+1})$$

وحسب الفرضية الأساسية H_5، ε_{n+1} هو غير مرتبط مع ε_n، أين نتحصل على:

$$V(e_{n+1}) = V(\hat{a}_0) + x_{n+1}^2 V(\hat{a}_1) + 2x_{n+1} COV(\hat{a}_0, \hat{a}_1)$$

وحسب العلاقتين ('9.2) و ("9.2):

$$V(e_{n+1}) = \frac{\sigma_\varepsilon^2}{n} + \bar{x}^2\, V(\hat{a}_1) + x_{n+1}^2 V(\hat{a}_1) - 2x_{n+1}\bar{x} V(\hat{a}_1) + \sigma_\varepsilon^2$$

في النهاية، باستعمال (8.2) وبعد التبسيط و تعويض σ_ε^2 بمقدره نجد:

$$\hat{V}(e_{n+1}) = \hat{V}(y_{n+1} - \hat{y}_{n+1}) = \hat{\sigma}_\varepsilon^2 \left(\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2} + 1\right) \dots\dots\dots (2.16)$$

يمكننا الملاحظة من هاته العلاقة الأخيرة أن تقدير تباين خطأ التنبؤ هو دالة (تابع) للبعد التربيعي بين القيمة التنبئية للمتغير المفسر (المستقل) ومتوسط هذا المتغير نفسه، حيث كلما زاد هذا البعد كلما زاد خطر خطأ التنبؤ. نلاحظ أيضا أن تقدير تباين خطأ التنبؤ هو دالة عكسية لتباين أو تشتت المتغير المفسر (المستقل أو الخارجي) ولحجم العينة.

إن فرضية طبيعية خطأ التشخيص ε_t تسمح بتحديد ووضع مجال ثقة عند $(1-\alpha)\%$ لقيمة التنبؤ على النحو التالي:

$$e_{n+1} = y_{n+1} - \hat{y}_{n+1} \sim N\left(0, \sigma_\varepsilon^2\left(\frac{1}{n} + \frac{(x_{n+1}-\bar{x})^2}{\sum_{t=1}^{n}(x_t-\bar{x})^2} + 1\right)\right)$$

ومنه

$$\frac{y_{n+1} - \hat{y}_{n+1}}{\hat{\sigma}_\varepsilon\sqrt{\left(\frac{1}{n} + \frac{(x_{n+1}-\bar{x})^2}{\sum_{t=1}^{n}(x_t-\bar{x})^2} + 1\right)}} \sim t_{n-2}^{\alpha/2}$$

أي

$$P\left(y_{n+1} \in \left(\hat{y}_{n+1} \pm \hat{\sigma}_\varepsilon\sqrt{\left(\frac{1}{n} + \frac{(x_{n+1}-\bar{x})^2}{\sum_{t=1}^{n}(x_t-\bar{x})^2} + 1\right)}\right)\right) = 1-\alpha \ldots\ldots\ldots (2.17)$$

حالة خاصة:

عندما نستعمل نموذج الانحدار الخطي البسيط يكون فيه المتغير الخارجي الزمن (t)، يكتب على الشكل:

$$y_t = \hat{a}_0 + \hat{a}_1 t + e_t, \qquad t = \overline{1, n}$$

من أجل حساب التنبؤ في الزمن n من أجل الأفق h، نستعمل العلاقة:

$$y_{n+h} = \hat{a}_0 + \hat{a}_1(n+h)$$

وطول مجال التنبؤ يزداد طرديا كلما زاد الأفق h كون تباين خطأ التنبؤ يكتب بدلالة $\left((n+h)-\bar{t}\right)^2$ حسب العلاقة (17.2) أين يكون n و $\bar{t}$ ثابتين.

3. نموذج الانحدار الخطي المتعدد:

نموذج الانحدار الخطي المتعدد هو امتداد لنموذج الانحدار الخطي البسيط المشار إليه في المحور السابق. بعد تقديم النموذج الخطي العام نقوم بتقدير معالمه ودراسة خصائصها. بعد ذلك نعرج على مختلف الاختبارات الإحصائية لمعالم النموذج المقدرة ونشير إلى تحليل التباين والاختبارات المتعلقة به. لاحقا نذكر فئة خاصة من المتغيرات المفسرة (الخارجية)، فئة المتغيرات التأشيرية، قبل التطرق إلى التنبؤ باستعمال النموذج الخطي العام.

3.أ. النموذج الخطي العام:

3.أ.1. تقديم:

في نموذج الانحدار الخطي البسيط فرضنا أن هناك متغير تابع واحد مفسر بمتغير خارجي واحد. لكن من النادر أن تفسر ظاهرة اقتصادية أو اجتماعية بمتغير واحد. النموذج الخطي العام أو نموذج الانحدار الخطي المتعدد هو تعميم وتوسعة لنموذج الانحدار الخطي البسيط أين يكون به عديد المتغيرات المفسرة (الخارجية)، تكتب صيغته المبسطة على الشكل:

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \cdots + a_k x_{kt} + \varepsilon_t \quad t = \overline{1, n}$$

أين:

y_t: المتغير التابع في الزمن t،

x_{1t}: المتغير المستقل الأول في الزمن t،

x_{2t}: المتغير المستقل الثاني في الزمن t،

$\vdots$

x_{kt}: المتغير المستقل k في الزمن t،

a_0، a_1... a_k معالم النموذج،

ε_t: خطأ التشخيص (الفرق بين النموذج الحقيقي والنموذج المشخص)، هذا الخطأ هو مجهول ويبقى مجهول.

3.أ.2. الكتابة بصيغة المصفوفات:

الكتابة السابقة للنموذج الخطي العام في الواقع هي أقل ملاءمة ولجعلها أكثر سهولة للاستعمال والتعامل معها ومع نتائجها نلجأ إلى كتابته على شكل مصفوفات.

بعد كتابة النموذج من أجل مشاهدات العينة على النحو التالي:

$$y_1 = a_0 + a_1x_{11} + a_2x_{21} + \cdots + a_kx_{k1} + \varepsilon_1$$
$$y_2 = a_0 + a_1x_{12} + a_2x_{22} + \cdots + a_kx_{k2} + \varepsilon_2$$
$$\vdots$$
$$y_n = a_0 + a_1x_{1n} + a_2x_{2n} + \cdots + a_kx_{kn} + \varepsilon_n$$

يمكننا كتابة النموذج الخطي العام على شكل مصفوفات:

$$\underset{(n,1)}{Y} = \underset{(n,k+1)}{X}\ \underset{(k+1,1)}{a} + \underset{(n,1)}{\varepsilon}$$

مع:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1t} & x_{2t} & \cdots & x_{kt} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}, \quad a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_k \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

أين نلاحظ أن العمود الأول من المصفوفة X يتشكل من العدد 1 الذي يمثل معاملات a_0 في النموذج الخطي العام أين يكون بعدها $(n, k+1)$ أي n سطر و $(k+1)$ عمود. كتابة النموذج الخطي العام على شكل مصفوفات يسهل التعامل معه.

3.ب. تقدير وخصائص مقدرات النموذج الخطي العام:

3.ب.1. تقدير معالم الانحدار:

ليكن النموذج الخطي العام ذو k متغير مستقل والمشاهدة على عينة حجمها n مشاهدة و المكتوب على شكل مصفوفات:

$$Y = Xa + \varepsilon \ldots \ldots (3.1)$$

لتقدير شعاع المعالم a المشكل من $a_0, a_1, \dots, a_k$، نطبق طريقة المربعات الصغرى المتمثلة في تصغير مجموع مربعات الأخطاء، ليكن:

$$Min \sum_{t=1}^{n} \varepsilon_t^2 = Min(\varepsilon'\varepsilon) = Min\big((Y - Xa)'(Y - Xa)\big) = Min\, S \dots\dots (3.2)$$

بحيث ε' هو مقلوب الشعاع ε.

نعلم أنه لا يمكننا تقدير المعالم من العلاقتين (1.3) و(2.3) كون خطأ التشخيص ε كما أشرنا هو مجهول ويبقى مجهول. نعوضه بشعاع البواقي e لتصبح العلاقة (1) $Y = X\hat{a} + e$ التي نطبق عليها طريقة المربعات الصغرى لنتحصل بدل العلاقة (2.3) على:

$$Min \sum_{t=1}^{n} e_t^2 = Min(e'e) = Min\big((Y - X\hat{a})'(Y - X\hat{a})\big) = Min\, \hat{S}$$

لتصغير $\hat{S}$، نشتقها بالنسبة ل $\hat{a}$ لنتحصل على:

$$\frac{\partial \hat{S}}{\partial \hat{a}} = -2X'Y + 2X'X\hat{a} = 0$$

$$\Rightarrow \hat{a} = (X'X)^{-1}(X'Y) \dots\dots (3.3)$$

مع تحقق الشرط الثاني (المشتقة الثانية موجبة):

$$\frac{\partial^2 S}{\partial \hat{a}^2} = 2X'X > 0$$

كون المصفوفة $(X'X)$ هي مصفوفة معرفة موجبة.

الحل الممثل بالعلاقة (3.3) يكون مكنا لما تكون المصفوفة $(X'X)$ ذات البعد $(k+1, k+1)$ قابلة للعكس.

المصفوفة $(X'X)$ تمثل الجداء المتقاطع بين المتغيرات المفسرة، في حالة ارتباط تام أو أي علاقة خطية بينها، تكون مصفوفة غير قابلة للعكس (المحدد معدوم) وتصبح طريقة المربعات الصغرى غير صالحة لتقدير معالم الشعاع a (محور تعدد الارتباط الخطي يعالج جزئيا هاته الإشكالية).

نسمي **المعادلات الطبيعية** المعادلات المستخرجة من العلاقة:

$$(X'X)\hat{a} = (X'Y)$$

النموذج الخطي العام المقدر يكتب:

$$y_t = \hat{y}_t + e_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \cdots + \hat{a}_k x_{kt} + e_t$$

أين $(e_t = y_t - \hat{y}_t)$ يمثل الباقي، أي الفرق بين القيمة المشاهدة y_t والقيمة المقدرة $\hat{y}_t$.

ملاحظة: ننبه مرة أخرى على الفرق بين خطأ تشخيص النموذج ε_t، الذي يبقى مجهولا، والباقي e_t الذي يكون معلوما عند تقدير النموذج.

حالة خاصة:

عند التعامل مع بيانات ممركزة حول متوسطها، $X_{it} = x_{it} - \bar{x}_i$ التقدير الممثل بالشعاع $\hat{a}$ يمكن أن يكتب بدلالة مصفوفة التباين والتباين المشترك على النحو التالي:

$$\hat{a} = \Omega^{-1} COV(X, Y)$$

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} v(x_1) & cov(x_1, x_2) & \cdots & \cdots & cov(x_1, x_k) \\ cov(x_2, x_1) & V(x_2) & & & cov(x_2, x_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ cov(x_k, x_1) & cov(x_k, x_2) & \cdots & \cdots & V(x_k) \end{pmatrix}^{-1} \begin{pmatrix} cov(x_1, y) \\ cov(x_2, y) \\ \vdots \\ cov(x_k, y) \end{pmatrix}$$

مع

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}_1 - \hat{a}_2 \bar{x}_2 - \cdots - \hat{a}_k \bar{x}_k$$

تأثير تغير متغير مستقل واحد:

ليكن النموذج العام المقدر:

$$y_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \cdots + \hat{a}_k x_{kt} + e_t$$

إذا تغيرت قيمة المتغير x_k من x_{kt} إلى $(x_{kt} + \Delta x_{kt})$، مع بقاء كل شيء آخر على حاله ($k-1$ المتغيرات الأخرى تبقى ثابتة)، فإن قيمة المتغير التابع المقدرة $\hat{y}$ تتغير ب $(\hat{a}_k \Delta x_{kt})$، أي: $\Delta \hat{y}_t = \hat{a}_k \Delta x_{kt}$.

3.ب.2. فرضيات وخصائص التقديرات:

إضافة إلى فرضية خطية النموذج بالنسبة ل X (في محور النماذج غير خطية سنفصل الفرق بين الخطية بالنسبة للمتغيرات X والخطية بالنسبة للمعالم a)، نميز بين الفرضيات الاحتمالية المتعلقة بخطأ التشخيص ε والفرضيات الهيكلية.

3.ب.2.1. الفرضيات الاحتمالية:

- H_1: النموذج خطي بالنسبة ل x_t،
- H_2: قيم المتغير المستقل x_t تمت مشاهدتها بجون خطأ (x_t غير عشوائي)،
- H_3: $E(\varepsilon_t) = 0$، الأمل الرياضي لخطأ التشخيص هو معدوم، أي أن النموذج هو مشخص جيدا في المتوسط ،
- H_4: $E(\varepsilon_t^2) = \sigma^2$، تباين الأخطاء هو ثابت $\forall t$، أي مجال تضاعفه هو نفسه مهما كانت الفترة أو الوحدة الإحصائية، تسمى هاته الفرضية بفرضية ثبات تباين الأخطاء،
- H_5: $E(\varepsilon_t, \varepsilon_{t'}) = 0$، لما $t \neq t'$، الأخطاء هي غير مرتبطة (أو حتى مستقلة)، أي خطأ في الزمن t ليس له تأثير على خطأ آخر في الزمن t'،
- H_6: $COV(x_t, \varepsilon_t) = 0$، الخطأ هو مستقل عن المتغير المستقل.

3.ب.2.2. الفرضيات الهيكلية:

- H_7: المتغيرات المستقلة هي مستقلة فيما بينها عن أي شكل من الارتباطات، هذا يستلزم أن المصفوفة $(X'X)$ قابلة للقلب.
- H_8: المصفوفة $(X'X)/n$ هي مصفوفة منتهية غير شاذة.
- H_9: حجم العينة أكبر من عدد معالم النموذج، $n > k + 1$.

3.ب.3. خصائص المقدرات:

ليكن النموذج المقدر بالعلاقة (3.3). النموذج الخطي العام المكتوب على شكل مصفوفات، مثله مثل النموذج الخطي البسيط، يمكن أيضا أن يكتب بطرق مختلفة:

$$\left.\begin{array}{l} Y = Xa + \varepsilon \\ Y = X\hat{a} + e \\ \hat{Y} = X\hat{a} \end{array}\right\} \rightarrow e = Y - \hat{Y} \text{ الباقي}$$

من العلاقة (3.3) نتحصل على:

$$\hat{a} = (X'X)^{-1}(X'Y) = (X'X)^{-1}X'(Xa + \varepsilon)$$

$$\Rightarrow \hat{a} = (X'X)^{-1}X'(Xa) + (X'X)^{-1}(X'\varepsilon)$$

$$\Rightarrow \hat{a} = a + (X'X)^{-1}X'\varepsilon \ldots \ldots (3.4)$$

بإدخال الأمل الرياضي على الطرفين نجد:

$$E(\hat{a}) = a + (X'X)^{-1}X'E(\varepsilon)$$

كون $E(\varepsilon) = 0$ نجد:

$$E(\hat{a}) = a$$

ونقول أن التقدير، المعبر عنه بالعلاقة (3.3)، هو تقدير غير متحيز.

نحسب الآن مصفوفة التباين والتباين المشترك للمعالم المقدرة التي نرمز لها $\Omega_{\hat{a}}$:

$$\Omega_{\hat{a}} = E[(\hat{a} - a)(\hat{a} - a)']$$

من العلاقة (4) لدينا:

$$(\hat{a} - a) = (X'X)^{-1}X'\varepsilon$$

و

$$(\hat{a} - a)' = \varepsilon'X(X'X)^{-1}$$

كون $(X'X)^{-1}$ هي مصفوفة متناظرة أي $[(X'X)^{-1}]' = (X'X)^{-1}$

ومنه

$$\Omega_{\hat{a}} = E[(\hat{a} - a)(\hat{a} - a)'] = (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}$$

لدينا

$$E(\varepsilon\varepsilon') = \Omega_\varepsilon = \begin{pmatrix} E(\varepsilon_1\varepsilon_1) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2\varepsilon_2) & \dots & E(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E(\varepsilon_n\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 I_n$$

لنجد في النهاية

$$\Omega_{\hat{a}} = \sigma_\varepsilon^2 (X'X)^{-1} \dots \dots (3.5)$$

يمكننا كتابة

$$\Omega_{\hat{a}} = \frac{\sigma_\varepsilon^2}{n} \left(\frac{X'X}{n}\right)^{-1}$$

حسب الفرضيتين H_3 و H_7:

$$\lim_{n\to\infty} \Omega_{\hat{a}} = \lim_{n\to\infty} \frac{\sigma_\varepsilon^2}{n} \left(\frac{X'X}{n}\right)^{-1} = 0$$

ونقول أن التقدير (3.3) هو تقدير مقارب.

نظرية غوس–ماركوف:

التقدير (3.3) بطريقة المربعات الصغرى هو أحسن تقدير خطي غير متحيز (المقدرات له أقل تباين ممكن).

نشير أنه في حالة تحقق فرضية الأخطاء الطبيعية فإن تقدير المعالم بطريقة المعقولية العظمى يعطي نتائج مماثلة لتلك المتحصل عليها بطريقة المربعات الصغرى.

يمكننا عبر حسابات مصفوفاتية تبيين أن تقدير تباين خطأ التشخيص يعطى ب:

$$\hat{\sigma}_\varepsilon^2 = \frac{e'e}{n-k-1} \dots \dots (3.6)$$

باستبدال تباين خطأ التشخيص في العلاقة (5.3) بتقديره نتحصل على:

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_\varepsilon^2 (X'X)^{-1} \dots \dots (3.7)$$

3.ب.4. نظرية Frisch, Waugh et Lovell (FWL):

نفرض أن المتغيرات المستقلة هي مقسمة إلى مجموعتين ممثلتين بالمصفوفتين X_1 و X_2. يكتب النموذج:

$$Y = X_1 a_1 + X_2 a_2 + \varepsilon$$

نظرية Frisch, Waugh & Lovell (FWL) تقول أن تقدير المربعات الصغرى GLS للمعالم a_2 والبواقي هو نفسه لما نقدر النموذج:

$$M_1 Y = M_1 X_2 a_2 + v$$

مع

$$M_1 = I - X_1 (X_1{}' X)^{-1} X_1{}'$$

المصفوفة M_1 هي مصفوفة متناظرة و ثابتة ($M_1 M_1 = M_1$)، وأيضا $M_1 X_1 = 0$.

تسمح هاته النظرية بفهم إشكالية تشخيص نموذج: إذا كان المتغير Y مفسر فعلا بمجموعة من المتغيرات X_1 و X_2، فمن اللازم والضروري أن تظهر هاته المتغيرات X_1 و X_2 حتى وإن كنا فقط نهتم بتأثير بمجموعة واحدة من المتغيرات.

3.ب.5. معادلة وجدول تحليل التباين وجودة النموذج :

مثل نموذج الانحدار الخطي البسيط، لدينا:

- $\sum_t y_t = \sum_t \hat{y}_t \Rightarrow \overline{\hat{Y}} = \bar{Y}$،
- $\sum_t^n e_t = 0$.

من هاتين العلاقتين نستخلص العلاقة الأساسية لتحليل التباين:

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t e_t^2 \ldots\ldots (3.8)$$

$$SCT \quad = \quad SCE \quad + \quad SCR$$

التغير الكلي (SCT) يساوي التغير المفسر (SCE) + تغير البواقي (SCR).

ستسمح لنا هاته العلاقة بقياس جودة النموذج بحيث كلما اقترب التغير المفسر من التغير الكلي كلما زادت الجودة العامة للنموذج.

كون قيمتهما مرتبطتين بوحدة القياس، نفضل لقياس جودة النموذج استعمال مؤشر نسبي بدون وحدة، معامل التحديد، الذي يعطى بالعلاقة:

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2} \dots \dots (3.9)$$

يسمى R بمعامل الارتباط المتعدد.

في حالة بيانات ممركزة حول متوسطها وفقط في هاته الحالة يمكننا التعبير عن معامل التحديد على شكل مصفوفات بالعلاقة:

$$R^2 = \frac{\hat{Y}'\hat{Y}}{Y'Y} = 1 - \frac{e'e}{Y'Y} \dots \dots (3.10)$$

جودة النموذج هاته يجب أن تعدل بدرجة الحرية. في الواقع، عندما تكون درجة الحرية للنموذج ضعيفة، من الأفضل تصحيح R^2 من أجل الأخذ بعين الاعتبار عدد المشاهدات (حجم العينة) مقارنة بعدد المتغيرات المفسرة (المعالم المقدرة) بحساب معامل التحديد المصحح، $\bar{R}^2$، الذي يعطى بالعلاقة:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) \dots \dots (3.11)$$

يكون لدينا $\bar{R}^2 < R^2$ و $\bar{R}^2 \cong R^2$ لما حجم العينة n يكون كبير كفاية. سنرى لاحقا معياري أكاييك وشوارز (Akaike و Schwarz) للفصل بين فقدان درجة حرية وإضافة معلومة عند إدخال متغير أو مجموعة متغيرات إلى النموذج.

الاختبارات الاحصائية :

3.ب.6. دور الفرضيات:

تحقق فرضية طبيعية الأخطاء يستلزم:

$$\frac{\sum_{t=1}^{n} e_t^2}{\sigma_\varepsilon^2} = (n-k-1)\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} = (n-k-1)\frac{\hat{\sigma}_{\hat{a}_i}^2}{\sigma_{\hat{a}_i}^2} \dots\dots (12)$$

تتبع توزيع كي-دو (χ^2) بدرجة حرية $(n-k-1)$ (مجموع مربعات $(n-k-1)$ متغيرات طبيعية معيارية مستقلة). ينتج من ذلك:

$$\frac{\hat{a}_i - a_i}{\hat{\sigma}_{\hat{a}_i}} \sim t_{n-k-1}$$

لدينا في الواقع القيمة:

$$\frac{\hat{a}_i - a_i}{\hat{\sigma}_{\hat{a}_i}} = \frac{\dfrac{\hat{a}_i - a_i}{\sigma_{\hat{a}_i}}}{\sqrt{(n-k-1)\dfrac{\hat{\sigma}_{\hat{a}_i}^2}{\sigma_{\hat{a}_i}^2}\dfrac{1}{(n-k-1)}}}$$

تمثل النسبة بين متغير طبيعي معياري ومتغير كي دو مقسوم على درجة حريته، وهو تعريف متغير ستودنت.

لدينا $(\hat{a}-a)'\Omega_{\hat{a}}^{-1}(\hat{a}-a)$ تتبع توزيع كي دو بدرجة حرية $(k+1)$ (مجموع مربعات $(k+1)$ متغيرات معيارية، $(k+1)$ معلمة). بتعويض مصفوفة التباين والتباين المشترك النظرية للمعالم بتقديرها $\hat{\Omega}_{\hat{a}} = \hat{\sigma}_\varepsilon^2 (X'X)^{-1}$، قانون الاحتمالات ل $\frac{1}{k+1}(\hat{a}-a)'\hat{\Omega}_{\hat{a}}^{-1}(\hat{a}-a)$ هو قانون فيشر بدرجة حرية $(k+1)$ و $(n-k-1)$.

في الواقع، القيمة:

$$F^* = \frac{\dfrac{1}{k+1}(\hat{a}-a)'(\hat{\sigma}_\varepsilon^2 (X'X)^{-1})^{-1}(\hat{a}-a)}{(n-k-1)\dfrac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}\dfrac{1}{(n-k-1)}}$$

تمثل النسبة بين متغيري كي دو مقسومين على درجتي حريتهما والتي تمثل تعريف متغير فيشر.

3.ب.7. بناء الاختبارات:

من خلال مختلف العلاقات المعرفة أعلاه، يمكننا وضع مجموعة من الاختبارات الإحصائية.

3.ب.1.7. مقارنة معلمة a_i بقيمة ثابتة $\bar{a}$:

كتابة الفرضية الموافقة يكون على النحو التالي:

$$\begin{cases} H_0: a_i = \bar{a} \\ H_1: a_i \neq \bar{a} \end{cases} \quad i = \overline{0, k}$$

نعلم أن:

$$\frac{\hat{a}_i - a_i}{\hat{\sigma}_{a_i}} \sim t_{n-k-1}$$

تحت الفرضية الصفرية تصبح:

$$\frac{|\hat{a}_i - \bar{a}|}{\hat{\sigma}_{\hat{a}_i}} = t^*_{\hat{a}_i} \sim t_{n-k-1} \ldots\ldots (3.13)$$

$t^*_{\hat{a}_i}$ تسمى بقيمة ستودنت المحسوبة.

ويكون قرار اختبار الفرضية على النحو التالي:

- إذا كان $t^*_{\hat{a}_i} > t^{\alpha/2}_{n-k-1}$ نرفض الفرضية H_0، أي أن a_i هو معنويا يختلف عن $\bar{a}$ عند عتبة α.
- إذا كان $t^*_{\hat{a}_i} < t^{\alpha/2}_{n-k-1}$ نقبل الفرضية H_0، أي أن a_i هو معنويا يساوي $\bar{a}$ عند عتبة α.

حالة خاصة: في حالة معرفة معنوية معلمة في النموذج (معنوية الحد الثابت أو معنوية المتغير المستقل الموافق في تفسير المتغير التابع) فإننا نستبدل فقط في الاختبار السابق $\bar{a}$ ب 0 ليصبح الاختبار:

$$\begin{cases} H_0: a_i = 0 \\ H_1: a_i \neq 0 \end{cases} \quad i = \overline{0, k}$$

وتصبح العلاقة (13.3):

$$\frac{|\hat{a}_i|}{\hat{\sigma}_{\hat{a}_i}} = t^*_{\hat{a}_i} \sim t_{n-k-1} \ldots\ldots (3.14)$$

ويتم اتخاذ القرار بنفس الطريقة السابقة.

هذا الاختبار هو مهم، بحيث إذا كان في نموذج مقدر معلمة غير معنوية (معدومة احصائيا) فيجب إقصاء المتغير المستقل الموافق لها من النموذج وإعادة تقدير النموذج الجديد. سبب عدم المعنوية قد يكون إما ل:

- غياب ارتباط مع المتغير التابع،
- ارتباط خطي قوي مع متغير مستقل آخر.

3.ب.7.2.مقارنة مجموعة من المعالم مع مجموعة من الثوابت:

نبحث الآن عن اختبار في آن واحد مساواة مجموعة جزئية من معالم النموذج مع قيم ثابتة. نعبر عن ذلك ب:

$$\begin{cases} H_0: a_q = \bar{a}_q \\ H_1: a_q \neq \bar{a}_q \end{cases}$$

q يمثل عدد معالم النموذج المختبرة وتمثل بعد الشعاع a_q.

بينا سابقا أن $\frac{1}{k+1}(\hat{a}-a)'\hat{\Omega}^{-1}_{\hat{a}}(\hat{a}-a)$ تخضع لقانون فيشر بدرجة حرية $(k+1)$ و $(n-k-1)$، كذلك من أجل مجموعة جزئية من q معلمة، تكون العبارة $\frac{1}{q}\left(\hat{a}_q-a_q\right)'\hat{\Omega}^{-1}_{\hat{a}_q}\left(\hat{a}_q-a_q\right)$ تخضع لقانون فيشر بدرجة حرية (q) و $(n-k-1)$. من أجل قبول الفرضية H_0 يكفي أن يكون:

$$\frac{1}{k+1}(\hat{a}-a)'\hat{\Omega}^{-1}_{\hat{a}}(\hat{a}-a) \leq F^{\alpha}(q, n-k-1) \ldots\ldots (3.15)$$

$F^{\alpha}(q, n-k-1)$ هي قيمة فيشر المجدولة عند عتبة α و درجة حرية (q) و $(n-k-1)$.

3.ب.7.3. مجال الثقة لتباين الأخطاء σ_ε^2:

مجال الثقة لتباين الأخطاء يسمح بوضع حدود لمدى مجال تغير الأخطاء. من أجل مجال ثقة عند 95%، يعطى المجال ب:

$$P\left(\sigma_\varepsilon^2 \in \left[\frac{(n-k-1)\hat{\sigma}_\varepsilon^2}{\chi_{1(n-k-1)}^{2\,(\alpha/2)}};\frac{(n-k-1)\hat{\sigma}_\varepsilon^2}{\chi_{2(n-k-1)}^{2(1-\alpha/2)}}\right]\right)\ldots\ldots(3.16)$$

3.ت. تحليل التباين:

3.ت.1. تشكيل جدول تحليل التباين واختبار المعنوية الكلية للنموذج:

في هاته الفقرة سنتساءل عن المعنوية الكلية لنموذج الانحدار، أي إن كانت المتغيرات المستقلة (المفسرة) لها تأثير على المتغير التابع. يمكن صياغة هذا الاختبار على الشكل التالي: هل يوجد على الأقل متغير مستقل معنوي؟

ليكن الاختبار:

$$\begin{cases} H_0: a_1 = a_2 = \cdots = a_k = 0 \\ H_1: \exists i\ t.q\ \ a_i \neq 0,\ \ i = \overline{1,k} \end{cases}$$

لا نختبر هنا معنوية الحد الثابت a_0 لأننا نهتم فقط بمعنوية المتغيرات المستقلة.

في حالة قبول الفرضية الصفرية هذا يعني أنه لا توجد أي علاقة خطية معنوية بين المتغير التابع والمتغيرات المستقلة (مجموع المربعات المفسرة ليست معنويا مختلة عن الصفر).

نأخذ العلاقة الأساسية (8.3) لتحليل التباين:

$$\sum_t (y_t-\bar{y})^2 = \sum_t (\hat{y}_t-\bar{y})^2 + \sum_t e_t^2$$

يعتبر الانحدار معنوي إذا كان التغير المفسر مختلف عن الصفر معنويا. الجدول أسفله يمثل تحليل التباين الذي يسمح بالقيام بالاختبار (اختبار فيشر). لدينا من العلاقة (9.3):

$$F^* = \frac{\sum_t(\hat{y}_t - \bar{y})^2/k}{\sum_t e_t^2/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F^\alpha(k, n-k-1) \dots (3.17)$$

جدول تحليل التباين للنموذج الخطي العام

مصدر التغير	مجموع المربعات	درجة الحرية ddl	متوسط مجموع المربعات
$x_1, x_2, \dots, x_k$	$SCE = \sum_t (\hat{y}_t - \bar{y})^2$	k	SCE/k
البواقي	$SCR = \sum_t e_t^2$	$(n-k-1)$	$SCR/(n-k-1)$
الكلي	$SCT = \sum_t (y_t - \bar{y})^2$	$(n-1)$	

فرضية التوزيع الطبيعي للأخطاء تستلزم أنه تحت الفرضية الصفرية (عند تحقق الفرضية الصفرية) فإن F^* تتبع قانون فيشر (النسبة بين متغيري كي دو مقسومين على درجة حريتهما) وفق العلاقة (17.3) أعلاه. نرفض H_0 لما يكون $F^* > F^\alpha(k, n-k-1)$ ونقر بالمعنوية العامة للنموذج. هذا الاختبار، في الحالة التطبيقية، يتم مباشرة عند معرف معامل التحديد R^2 (فقط في نموذج به حد ثابت) وفق دائما العلاقة (17.3).

3.ت.2. اختبارات أخرى من خلال جدول تحليل التباين:

من خلال جدول تحليل التباين يمكننا القيام بمجموعة أخرى من الاختبارات، أهمها:

- **اختبار إضافة متغير أو مجموعة من المتغيرات المفسرة إلى النموذج:**

هل إضافة مجموعة من المتغيرات المستقلة إلى النموذج يحسن جودة الانحدار؟

نفرض أن النموذج يحوي k_1 متغير مستقل ونرغب في اختبار معنوية إضافة $(k-k_1)$ متغير مستقل.

يعبر عن هذا الاختبار ب:

$$\begin{cases} H_0: a_{k_1+1} = a_{k_1+2} = \cdots = a_k = 0 \\ H_1: \exists i\ t.q\ \ a_i \neq 0,\ \ i = \overline{k_1+1, k} \end{cases}$$

ويتم وفق أربعة (4) مراحل:

- المرحلة الأولى: يتم حساب التغير الكلي، التغير المفسر وتغير البواقي للنموذج الكلي (النموذج يحوي المتغيرات المستقلة المضافة المراد اختباها)، SCT، SCE و SCR.
- المرحلة الثانية: يتم حساب التغير الكلي ، التغير المفسر وتغير البواقي للنموذج المقدر بدون المتغيرات المضافة المراد اختبارها، SCT_1، SCE_1 و SCR_1.
- المرحلة الثالثة: يتم تشكيل جدول تحليل التباين على النحو التالي:

جدول تحليل التباين لاختبار معنوية إضافة متغيرات مستقلة إلى النموذج

متوسط مجموع المربعات	درجة الحرية ddl	مجموع المربعات	مصدر التغير
SCE_1/k_1	k_1	SCE_1	$x_1, x_2, \dots, x_{k_1}$
SCE/k	k	SCE	$x_1, x_2, \dots, x_k$
$SCR/(n-k-1)$	$(n-k-1)$	SCR	البواقي
	$(n-1)$	SCT	الكلي

- المرحلة الرابعة:

يبرهن أن:

$$F^* = \frac{(SCE - \mathrm{SCE}_1)/(k-k_1)}{SCR/(n-k-1)} \sim F^{\alpha}_{[(k-k_1),(n-k-1)]}$$

حيث نقبل الفرضية الصفرية، لا يوجد فرق معنوي بين التغيرين المفسرين، أي أن إضافة المتغيرات المستقلة $(k-k_1)$ لم يحسن النموذج عند العتبة α لما يكون $F^* \leq F^{\alpha}_{[(k-k_1),(n-k-1)]}$ ونقبل الفرضية البديلة، إضافة المتغيرات حسن من جودة النموذج، لم $F^* > F^{\alpha}_{[(k-k_1),(n-k-1)]}$.

3.ث. التنبؤ باستعمال النموذج الخطي العام والانحدار التراجعي:

3.ث.1. التنبؤ الشرطي:

تتمثل الإشكالية في تحديد القيمة المعطاة للمتغير الداخلي (التابع) عندما نعلم قيم المتغير أو المتغيرات الخارجية (المستقلة أو المفسرة). يكتب النموذج العام المقدر على الشكل:

$$y_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \cdots + \hat{a}_k x_{kt} + e_t$$

إذا التنبؤ من أجل قيمة الفترة $t+h$ (أو $i+h$ بالنسبة للبيانات المقطعية الآنية) هو:

$$\hat{y}_{t+h} = \hat{a}_0 + \hat{a}_1 x_{1t+h} + \hat{a}_2 x_{2t+h} + \cdots + \hat{a}_k x_{kt+h}$$

ويعطى خطأ التنبؤ ب:

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h}$$

باعتبار أن فرضيات النموذج الخطي العام هي محققة، التنبؤ $\hat{y}_{t+h}$ هي غير متحيزة.

افترضنا أنه للقيام بالتنبؤ فإننا نعرف بدون خطأ قيم المتغيرات الخارجية في الزمن $t+h$. هذه الفرضية قد تكون محققة بالنسبة للبيانات المقطعية الآنية لكن قد لا تكون كذلك بالنسبة للبيانات الزمنية التي تكون غالبا مقدرة أو متنبأ بها بهامش خطأ وبالتالي تدخل عناصر عدم دقة جديدة على عملية حساب التنبؤ. لا نتناول هاته الإشكالية هنا لأننا نتحدث عن التنبؤ الشرطي.

3.ث.2. دقة ومجال التنبؤ:

خطأ التنبؤ المحسوب في الزمن t للأفق h، e_{t+h}، يمكن أن يكتب على الشكل:

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h} = X_{t+h}a + \varepsilon_{t+h} - X_{t+h}\hat{a} = X_{t+h}(a-\hat{a}) + \varepsilon_{t+h}$$

نحسب تباينه:

$$V(e_{t+h}) = V(X_{t+h}(a-\hat{a}) + \varepsilon_{t+h})$$

$$= V\big(X_{t+h}(a-\hat{a})\big) + V(\varepsilon_{t+h}) + 2COV(X_{t+h}(a-\hat{a}), \varepsilon_{t+h})$$

من جهة لدينا $X_{t+h}(a-\hat{a})$ هي عبارة عن توليفة خطية X_t و ε_t و ε_{t+h} هي غير مرتبطة معهما (الفرضيتين H_4 و H_5) فيكون $COV(X_{t+h}(a-\hat{a}), \varepsilon_{t+h}) = 0$، ومنه:

$$V(e_{t+h}) = V\big(X_{t+h}(a-\hat{a})\big) + V(\varepsilon_{t+h})$$

ومن جهة لدينا:

$$V\big(X_{t+h}(a-\hat{a})\big) = X_{t+h}V(a-\hat{a})X'_{t+h} = \sigma_\varepsilon^2[X_{t+h}(X'X)^{-1}X'_{t+h}]$$

و

$$V(\varepsilon_{t+h}) = \sigma_\varepsilon^2$$

ومنه يعطى تباين خطأ التنبؤ بالعلاقة:

$$V(e_{t+h}) = \sigma_\varepsilon^2[X_{t+h}(X'X)^{-1}X'_{t+h} + 1] \dots\dots (3.18)$$

حيث X_{t+h} تمثل المصفوفة السطرية لقيم المتغيرات الخارجية (المستقلة) في الزمن ا $t+h$:

$$X_{t+h} = (1 \quad x_{1\,t+h} \quad x_{2\,t+h} \quad \cdots \quad x_{k\,t+h})$$

إن خطأ التنبؤ ($e_{t+h} = y_{t+h} - \hat{y}_{t+h}$) هو عبارة عن توليفة خطية بين المتغيرات المستقلة وخطأ التشخيص ε_{t+h}، إذا دائما حسب فرضية التوزيع الطبيعي، فإن خطأ التنبؤ يتبع التوزيع الطبيعي:

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h} \sim N\big(0, \sigma_{e_{t+h}}^2\big)$$

وباستبدال تباين خطأ التشخيص النظري في العلاقة (18.3) بتقديره، يمكننا أن نستنتج أن:

$$\frac{e_{t+h}}{\hat{\sigma}_{\varepsilon}\sqrt{[X_{t+h}(X'X)^{-1}X'_{t+h}+1]}}=\frac{y_{t+h}-\hat{y}_{t+h}}{\hat{\sigma}_{\varepsilon}\sqrt{[X_{t+h}(X'X)^{-1}X'_{t+h}+1]}}\sim t_{n-k-1}^{\alpha/2}$$

نلاحظ أيضا من العلاقة (18.3)، كما هو الحال بالنسبة للنموذج الخطي البسيط، أن تباين خطأ التنبؤ يكون صغيرا كلما كان:

- تباين البواقي صغير،
- قيم المتغيرات المستقلة التنبئية تقترب من متوسطاتها الموافقة.

مجال التنبؤ y_{t+h} ل عند مستوى $(1-\alpha)$ يعطى ب:

$$y_{t+h}=\hat{y}_{t+h}\pm t_{n-k-1}^{\alpha/2}\hat{\sigma}_{\varepsilon}\sqrt{[X_{t+h}(X'X)^{-1}X'_{t+h}+1]} \ldots\ldots (3.19)$$

4. التوسع في نموذج الانحدار الخطي المتعدد:

سنحاول من خلال هذا المحور التطرق إلى بعض العناصر البعدية المهمة لتقدير نموذج، بداية من اختبار وجود تغير هيكلي في النموذج (استقرارية النموذج) إلى اختبار قيود شرطية على معالمه ومفهوم المتغيرات الصورية (التأشيرية) ثم إلى طريقة التحقق من الصيغة الهيكلية للنموذج.

4.أ. التحول الهيكلي واختبار استقرارية النموذج على طول الفترة الكلية:

هل يمكننا اعتبار النموذج مستقر على طول فترة الدراسة، أو يمكننا وضع نموذجين أو أكثر لفترات متباينة (تغير هيكلي للنموذج)؟ تشخيص النموذج هو نفسه ولكن بمعالم مقدرة مختلفة بين الفترتين.

4.أ.1. اختبار الاستقرارية ل Chow:

ليكن النموذج المقدر على طول فترة الدراسة n:

$$y_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \cdots + \hat{a}_k x_{kt} + e_t \,, \quad t = \overline{1, n}$$

والنموذج المقدر على فترتين:

$$y_t = \hat{a}_0^1 + \hat{a}_1^1 x_{1t} + \hat{a}_2^1 x_{2t} + \cdots + \hat{a}_k^1 x_{kt} + e_t \,, \quad t = \overline{1, n_1}$$
$$y_t = \hat{a}_0^2 + \hat{a}_1^2 x_{1t} + \hat{a}_2^2 x_{2t} + \cdots + \hat{a}_k^2 x_{kt} + e_t \,,, \quad t = \overline{n_1, n}$$

يعبر عن الاختبار ب:

$$H_0 : \begin{pmatrix} a_0 = a_0^1 = a_0^2 \\ a_1 = a_1^1 = a_1^2 \\ \vdots \\ a_k = a_k^1 = a_k^2 \end{pmatrix}$$

يعرف هذا الاختبار باختبار Chow او اختبار استقرارية المعالم حيث يمكن صياغته احصائيا على النحو التالي: هل يوجد فرق معنوي بين مجموع مربعات البواقي لنموذج الفترة الكلية (SCR) ومجموع مربعات البواقي لنموذجي الفترتين $(SCR_1 + SCR_2)$؟

يبرهن أن:

$$F^* = \frac{[SCR - (SCR_1 + SCR_2\,)]/ddl_n}{(SCR_1 + SCR_2\,)/ddl_d} \sim F^{\alpha}_{(ddl_n, ddl_d)}$$

بحيث:

$$ddl_n = [(n - k - 1) - [(n_1 - k - 1\,) + (n_1 - k - 1\,)]]$$

و

$$ddl_d = [(n_1 - k - 1\,) + (n_1 - k - 1\,)]$$

أين $n = n_1 + n_2$.

حيث نقبل الفرضية الصفرية، بالقول أن النموذج مستقر على طول الفترة، أي لا يوجد تغير هيكلي للنموذج بين الفترتين وهذا عند العتبة α لما يكون $\left(F^* \leq F^{\alpha}_{(ddl_n, ddl_d)}\right)$ ونقبل الفرضية البديلة، النموذج غير مستقر على طول الفترة وبوجود تغير هيكلي بين الفترتين لما يكون $\left(F^* > F^{\alpha}_{(ddl_n, ddl_d)}\right)$.

ننبه أنه في حالة عدم تحقق فرضية تجانس تباين الأخطاء (الفرضية الاحتمالية H_3) فإن اختبار Chow هذا يكون متحيزا اتجاه توسيع مجال الرفض، أي يزيد احتمال رفض الفرضية H_0.

4.أ.2. اختبارات الاستقرارية بالانحدار التراجعي:

نسمي الانحدار التراجعي الطريقة التي تتمثل في تقدير متتابع لنماذج مع الزيادة المتتالية في كل مرة لعدد المشاهدات. نقدر أولا النموذج مع $k + 2$ المشاهدة الأولى (نموذج ذو درجة حرية واحدة) ثم نضيف المشاهدة التالية ونقدر مجددا النموذج وهكذا دواليك حتى إضافة جيع باقي المشاهدات. تفحص بسيط لتمثيل تغير معالم النماذج مرفوقة بمجالات ثقتها ($\pm$ ضعف الانحراف المعياري) يسمح بملاحظة التغيرات الهيكلية المحتملة للنموذج.

- اختبارات المجاميع التراكمية CUSUM تعتمد على ديناميكية أخطاء التنبؤ. تسمح هاته الاختبارات بالكشف عن عدم الاستقرارات الهيكلية للنموذج عبر الزمن. على عكس اختبار Chow، لا تتطلب هاته الاختبارات المعرفة المسبقة لتواريخ التغير الهيكلي. الفكرة العامة لها تتمثل في دراسة التطور الزمني لخطأ التنبؤ المعياري، نسمي خطأ تراجعي w_t تتابع خطأ التنبؤ المعياري المحسوب في الزمن $t-1$ للزمن t.

الخطأ التراجعي w_t في الزمن t يعرف على أنه النسبة بين:

- الفرق بين التنبؤ المحسوب في الزمن $t-1$ للزمن t والقيمة المحققة في الفترة t،
- وقيمة متغير سلمي محسوب من أجل نفس الفترة t.

أي:

$$w_t = \frac{e_t}{S_e} = \frac{y_t - X_t'\hat{a}_{t-1}}{(X_{t+h}(X'X)^{-1}X_{t+h}' + 1)}$$

مع $t = K+2, K+3, \dots, n$ و $K = k+1$ هو عدد معالم النموذج المقدر).من صيغته، يتبع الخطأ التراجعي w_t التوزيع الطبيعي $N\left(0, \sigma_{w_t}^2\right)$. ليكن الاختبارين:

- المجاميع التراكمية CUSUM التي تعتمد على المجموع التراكمي المحسوب من البواقي التراجعية،
- المجاميع التراكمية CUSUM التي تعتمد على المجموع التراكمي المحسوب من مربعات البواقي التراجعية.

من خلال الباقي التراجعي w_t، نحسب:

- إحصائية المجاميع التراكمية CUSUM:

$$W_t = \frac{n-K}{SCR} \sum_{j=K+2}^{t} w_j$$

دائما مع $t = K + 2, K + 3, \ldots, n$ ($K = k + 1$ هو عدد معالم النموذج المقدر) و SCR يمثل مجموع مربعات البواقي المحسوب من النموذج ذو n مشاهدة.

إذا كانت المعالم مستقرة زمنيا، إذا الأخطاء التراجعية يجب أن تبقى داخل المجال المحدد بالمستقيمين: $\left[K, \pm\alpha\sqrt{n-K}\right]$ و $\left[n, \pm 3\alpha\sqrt{n-K}\right]$ مع $\alpha = 1{,}143$ ، $0{,}984$ و $0{,}850$ على التوالي من أجل عتبات الثقة 1%، 5% و 10%.

في الحالة العكسية يكون النموذج غير مستقر.

- إحصائية المجاميع التراكمية CUSUM SQ تعطى بمربع الخطأ التراجعي:

$$S_t = \frac{\sum_{j=K+2}^{t} w_j^2}{\sum_{j=K+2}^{n} w_j^2} \text{ مع } t = K + 2, K + 3, \ldots, n \text{ و } 0 \leq S_t \leq 1$$

يسمح هذا الاختبار بالكشف عن التغيرات العشوائية لسلوك النموذج. إذا كانت المعالم مستقرة خلال الزمن، إذا البواقي التراجعية التربيعية يجب أن تبقى داخل المجال المعرف ب:

$$\left[\pm C \frac{(t-k)}{n-k}\right]$$

أين C هي إحصائية كولموغوروف–سميرنوف (*Kolmogorov–Smirnov*).

4.ب. النماذج المقيدة واختبار قيود وضوابط على المعالم:

هل القيود المرجوة والمفروضة على المعالم هي مبررة وصحيحة قياسيا؟ يعبر عن اختبار هاته القيود اختصارا ب:

$$H_0: a_i = c$$

تكمن الفكرة في حساب مجموع مربعات البواقي للنموذج بدون قيود SCR ومجموع مربعات البواقي للنموذج بع إدراج القيود عليه SCR_1 وحساب قيمة فيشر:

$$F^* = \frac{(SCR_1 - SCR)/ddl_n}{SCR/(n-k-1)}$$

مع $ddl_n = (n-k'-1) - (n-k-1)$ (k' تعبر عن تغيير القيود لدرجة الحرية).

ومقارنتها بقيمة فيشر المجدولة $F^{\alpha}_{(ddl_n,(n-k-1))}$ ويكون القرار على النحو التالي:

نقبل الفرضية الصفرية، القيود المفروضة هي صحيحة وهذا عند العتبة α لما يكون $F^* \leq F^{\alpha}_{(ddl_n,ddl_d)}$ ق ، هذا يعني إحصائيا أنه لا يوجد فرق بين SCR و SCR_1، ونقبل الفرضية البديلة، القيود المفروضة هي غير مقبولة، لما $F^* > F^{\alpha}_{(ddl_n,ddl_d)}$.

4.ت. بناء وأهداف المتغيرات الصورية (التأشيرية):

المتغير الصوري أو التأشيري هو متغير مستقل (مفسر) خاص الذي يأخذ فقط قيمتين 0 أو 1. يستعمل هذا المتغير، في نموذج قياسي، عندما نرغب في إدراج متغير وصفي ثنائي (يأخذ صفتين) إما لتصحيح متغير غير طبيعي (الصفة موجودة أو غير موجودة) أو لما يمون المتغير المفسر المستقل وصفي ثنائي في حد ذاته. يتمثل إذا في إدراج متغير أو مجموعة من المتغيرات الإضافية إلى النموذج المشخص وتطبيق الطرق الكلاسيكية للتقدير. تسمى هاته المتغيرات أحيانا بالمتغيرات الصامتة.

نموذج الانحدار يختلف وفقا حدوث وظهور الظاهرة من خلال قيم معلمة أو مجموعة من المعالم بينما تبقى العوامل الأخرى ثابتة. في حالة تغير هيكلي لمعلمة من النموذج، المتغير الصامت يؤثر على معلمة المتغير أو المتغيرات المفسرة.

مثلا، ليكن النموذج الخطي بمتغيرين مستقلين x_{1t} و x_{2t}:

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + b_0 D_t + b_1 D_t x_{1t} + b_2 D_t x_{2t} + \varepsilon_t$$

إذا تحققت الظاهرة، $D_t = 1$ و $D_t = 0$.

إذا كان $D_t = 0$، النموذج يكتب :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$$

إذا كان $D_t = 1$، النموذج يكتب :

$$y_t = (a_0 + b_0) + (a_1 + b_1)x_{1t} + (a_2 + b_2)x_{2t} + \varepsilon_t$$

إذا كان $b_1 = b_2 = 0$، يختلف النموذج فقط بالحد الثابت.

مجال استعمال المتغيرات التأشيرية (الصامتة) هو واسع جدا ومتنوع، يمكننا ذكر تصحيح القيم غير طبيعية (الشاذة)، التغيرات الهيكلية، تحليل وإدراج الفصلية، تشخيص الأفراد أو الوحدات الإحصائية، إدراج عوامل وصفية... الخ.

4.ث. اختبار التشخيص لرامسي (Ramsey):

اختبار رامسي يتحقق من فعالية الصيغة الهيكلية للنموذج من الشكل:

- علاقة هيكلية غير ملائمة بين المتغير التابع والمتغيرات المستقلة،
- غياب متغير مستقل في النموذج،
- الارتباط بين المتغيرات المستقلة وعنصر الخطأ،.....

بدل تقدير صيغ أخرى (مثلا غير خطية)، يهتم الاختبار بمعنوية معلمة أو مجموعة من المعالم لنموذج وسيطي أين تظهر السلسلة أو المتغير المفسر المقدر بأسس 2، 3، 4... يتم اختبار رامسي ضمن 3 مراحل:

- **المرحلة 1**: تقدير النموذج الأولي بطريقة المربعات الصغرى (GLS) وحساب القيم التقديرية للمتغير التابع:

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \cdots + \hat{a}_k x_{kt}$$

– **المرحلة 2**: تقدير النموذج الوسيطي:

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \cdots + b_k x_{kt} + \emptyset_2 \hat{y}_t^2 + \emptyset_3 \hat{y}_t^3 + \cdots + \emptyset_h \hat{y}_t^h + v_t$$

– **المرحلة 3**: اختبار الفرضية $H_0: \emptyset_2 = \emptyset_3 = \cdots = \emptyset_h = 0$ باستعمال اختبار فيشر لمجموعة جزئية من المعالم (أو اختبار ستودنت لما $h = 2$). إذا تم قبول الفرضية H_0، نقول أن النموذج هو خطي ولا توجد مشكلة تشخيص للنموذج.

5. إشكالية تعدد الارتباط الخطي وطرق اختيار المتغيرات التفسيرية:

عند بنائه للنماذج، قد يواجه القياسي غزارة المعلومات المتوفرة وتنوعها حيث يتردد عند تشخيصه للنموذج في عدد متغيراته وفي الاختيار بينها. يعبر عن هذه الإشكالية بكيفية معرفة عدد المتغيرات المفسرة (المستقلة) الأمثل في بناء النموذج القياسي ويعبر عن هذا إحصائيا بإيجاد المتغيرات المستقلة التي تعظم ارتباطها مع المتغير التابع وتكون أقل ارتباطا فيما بينها. لكن قبل ذلك نتطرق إلى مفهوم الارتباط الجزئي.

5.أ. الارتباط الجزئي:

معامل الارتباط الجزئي يقيس الارتباط بين متغيرين مع عزل تأثير متغير أو مجموعة من المتغيرات الأخرى على هذا الارتباط.

ليكن y متغير مفسر، x_1، x_2 و x_3 متغيرات مفسرة، المعاملات r_{y,x_1}^2، r_{y,x_2}^2 و r_{y,x_3}^2 تقيس على التوالي تغير y المفسر ب x_1 فقط، ب x_2 فقط و ب x_3 فقط.

يمكننا تشكيل 6 معاملات ارتباط جزئية من الدرجة الأولى:

r_{yx_1,x_2}^2، r_{yx_1,x_3}^2 r_{yx_2,x_1}^2، r_{yx_2,x_3}^2 r_{yx_3,x_1}^2، r_{yx_3,x_2}^2

و تشكيل 3 معاملات ارتباط جزئية من الدرجة ثانية:

$$r^2_{yx_3,x_1\,x_2} \text{ ، } r^2_{yx_2,x_1\,x_3} \text{ ، } r^2_{yx_1,x_2\,x_3}$$

مفهوم الارتباط الجزئي هذا هو مهم جدا لأنه يسمح بفعالية إضافة أو إدخال متغير مستقل جديد إلى النموذج.

كلما كانت قيمة معامل الارتباط الجزئي مرتفعة كلما زادت أهمية المتغير الموافق في تفسير النموذج. يمكن حساب معامل الارتباط الجزئي بطريقتين:

1- من خلال معامل الارتباط البسيط:

ليكن لدينا k متغير مفسر، يتم حساب الارتباط الجزئي بين المتغير y والمتغير المفسر x_i، $r^2_{yx_i,x_1\,\ldots\,x_{i-1}x_{i+1}\ldots\,x_k}$ بين باقيي انحداري المتغير التابع y والمتغير المفسر x_i ، e_1 و e_2، على المجموعة الجزئية المتشكلة من $k-1$ المتغير المفسر الآخر. ونكتب:

$$r^2_{yx_i,x_1\,\ldots\,x_{i-1}x_{i+1}\ldots\,x_k} = r^2_{e_1,e_2}$$

بحيث:

$$e_1 = y - y(x_1 \;\ldots\; x_{i-1}x_{i+1} \;\ldots\; x_k)$$

و

$$e_2 = x_i - x_i(x_1 \;\ldots\; x_{i-1}x_{i+1} \;\ldots\; x_k)$$

2- من خلال قيمة ستودنت t:

في نموذج ذو k متغير مفسر، توجد علاقة بين معامل الارتباط الجزئي و قيمة ستودنت t على النحو التالي:

$$r^2_{yx_i,x_1\,\ldots\,x_{i-1}x_{i+1}\ldots\,x_k} = \frac{t_i^2}{t_i^2 + (n-k-1)} \;\ldots\ldots\; (5.1)$$

ملاحظة: العلاقة (1.5) محققة فقط من أجل معامل الارتباط الجزئي من الدرجة $k-1$.

حالة خاصة: في حالة $k = 3$، x_3 x_2 ،x_1

معامل الارتباط الجزئي بين x_1 و x_2 مع عزل تأثير x_3، $r^2_{x_1x_2,x_3}$، يعطى ب:

$$r^2_{x_1x_2,x_3} = \frac{r^2_{x_1x_2} - r^2_{x_1\,x_3} r^2_{x_2\,x_3}}{\sqrt{\left(1 - r^2_{x_1\,x_3}\right)\left(1 - r^2_{x_2\,x_3}\right)}} \dots\dots (5.2)$$

ومعامل الارتباط الجزئي بين x_1 و x_3 مع عزل تأثير x_2، $r^2_{x_1x_3,x_2}$، يعطى ب:

$$r^2_{x_1x_3,x_2} = \frac{r^2_{x_1x_3} - r^2_{x_1\,x_2} r^2_{x_2\,x_3}}{\sqrt{\left(1 - r^2_{x_1\,x_2}\right)\left(1 - r^2_{x_2\,x_3}\right)}} \dots\dots (5.3)$$

5.ب. العلاقة بين معاملات الارتباط، البسيط، الجزئي والمتعدد:

في حالة نموذج ذو متغير مستقل واحد، x_1، مجموع مربعات البواقي، SCR، يساوي:

$$\mathrm{SCR} = \sum_t e_t^2 = \sum_t (y_t - \bar{y})^2 \left(1 - R^2_{y,x_1}\right) = \sum_t (y_t - \bar{y})^2 \left(1 - r^2_{y,x_1}\right)$$

بحيث R^2_{y,x_1} هو معامل التحديد لانحدار y على x_1. نفرض الآن نموذج به متغيرين مستقلين x_1 و x_2:

$$y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + e$$

مجموع مربعات البواقي، SCR، بعد عزل تأثير x_1 و x_2 تساوي:

$$\mathrm{SCR} = \sum_t e_t^2 = \sum_t (y_t - \bar{y})^2 \left(1 - R^2_{y,x_1x_2}\right)$$

بحيث R^2_{y,x_1x_2} هو معامل التحديد لانحدار y على x_1 و x_2.

لكن $r^2_{yx_2,x_1}$ هي نسبة الباقي المفسرة فقط ب x_2، هاته العبارة الأخيرة يمكن أن تكتب:

$$\sum_t (y_t - \bar{y})^2 \left(1 - R^2_{y,x_1x_2}\right) = \left(1 - r^2_{yx_2,x_1}\right) \sum_t (y_t - \bar{y})^2 \left(1 - r^2_{yx_1}\right) \dots (5.4)$$

ومنه

$$1 - R^2_{y,x_1x_2} = \left(1 - r^2_{yx_1}\right)\left(1 - r^2_{yx_2,x_1}\right) \dots \dots (5.5)$$

تحصلنا على تجزئة التأثير النسبي لكل من: مساهمة x_2 في تفسير y عندما يكون تأثير x_1 معزول ومساهمة x_1 في تفسير y (بعض الكتاب يستعملون مصطلح العائد لوصف هاته المساهمات النسبية).

العلاقة (5.5) يمكن أن تعمم على حالة نموذج ذو ثلاث متغيرات مستقلة:

$$1 - R^2_{y,x_1x_2x_3} = \left(1 - r^2_{yx_1}\right)\left(1 - r^2_{yx_2,x_1}\right)\left(1 - r^2_{yx_3,x_1x_2}\right) \dots \dots (5.6)$$

من أجل نموذج ذو أربع متغيرات مفسرة نتحصل على العلاقة:

$$1 - R^2_{y,x_1x_2x_3x_4} = \left(1 - r^2_{yx_1}\right)\left(1 - r^2_{yx_2,x_1}\right)\left(1 - r^2_{yx_3,x_1x_2}\right)\left(1 - r^2_{yx_4,x_1x_2x_3}\right) \dots (5.7)$$

نشير إلى إمكانية تبديل ترتيب مؤشرات المتغيرات المستقلة، لتصبح العلاقة الأخيرة:

$$1 - R^2_{y,x_1x_2x_3x_4} = \left(1 - r^2_{yx_3}\right)\left(1 - r^2_{yx_4,x_3}\right)\left(1 - r^2_{yx_1,x_3x_4}\right)\left(1 - r^2_{yx_2,x_1x_3x_4}\right)$$

5.ت. تعدد الارتباط الخطي:

يستعمل مصطلح "تعدد الارتباط الخطي" للنموذج الذي يحتوي على متغيرات مفسرة تكون مرتبطة فيما بينها. بالمقابل، من أجل متغيرات مفسرة ذات تباين مشترك معدوم $\left(Cov(x_i, x_j) = 0, i \neq j\right)$، نقول عنها أنها متعامدة (مستقلة). إذا كان، من أجل الدراسات النظرية، بالإمكان افتراض تعامد سلسلتين إحصائيتين، فإنه بالنسبة للقياسي الاقتصادي أو المنمذج تكونين غالبا مرتبطتين ولو بنسبة ضئيلة.

5.ت.1. نتيجة أو آثار تعدد الارتباط الخطي:

يمكننا ذكر 3 أثار أساسية:

1- زيادة قيمة تباين بعض المعالم المقدرة عند زيادة الارتباط الخطي بين المتغيرات المفسرة،

2- عدم استقرارية المعالم المقدرة، تغيرات طفيفة في البيانات ينجر عنها تغيرات كبير في قيم المعالم المقدرة،

3- في حالة **الارتباط الخطي التام**، المصفوفة $(X'X)$ تصبح غير قابلة للعكس (محددها معدوم)، يصبح تقدير المعالم غير ممكن وتبايناتها غير منتهية.

5.ت.2. اختبارات الكشف عن تعدد الارتباط الخطي:

يوجد العديد من اختبارات الكشف عن تعدد الارتباط الخطي التي تتقاطع أحيانا في مبادئها.

أ- اختبار كلين Klein:

يعتمد اختبار كلين على المقارنة بين معامل التحديد R_y^2 المحسوب للنموذج ذو k متغير مفسر:

$$y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \cdots + \hat{a}_k x_k + e$$

ومعاملات الارتباط الخطية البسيطة $r_{x_i x_j}$ بين المتغيرات المفسرة من أجل $i \neq j$.

إذا كان $r_{x_i x_j}^2 > R_y^2$ فإنه يكون احتمال وجود ارتباط بين المتغيرين x_i و x_j .

اختبار كلين ليس اختبار إحصائي بمعنى اختبار الفرضيات لكن ببساطة هو مؤشر لإمكانية وجود تعدد ارتباط خطي.

ب- اختبار فيرار وغلوبر (Ferrar & Glauber):

تتمثل المرحلة الأولى من الاختبار في حساب محدد مصفوفة معاملات الارتباط بين k المتغيرات المفسرة.

$$D = \begin{vmatrix} 1 & r_{x_1 x_2} & r_{x_1 x_3} & \cdots & r_{x_1 x_k} \\ r_{x_2 x_1} & 1 & r_{x_2 x_3} & \cdots & r_{x_2 x_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_k x_1} & r_{x_k x_2} & r_{x_k x_3} & \cdots & 1 \end{vmatrix}$$

عندما تؤول قيمة المحدد D إلى الصفر، يكون هناك احتمال كبير لوجود تعدد الارتباط الخطي.

تتمثل المرحلة الثانية من الاختبار في إجراء اختبار χ^2، وذلك بوضع الفرضية التالية:

H_0: $D = 1$ (السلاسل متعامدة أو مستقلة)،

H_1: $D \neq 1$ (السلاسل مرتبطة)،

القيمة المحسوبة لاختبار χ^2 المحسوبة من العينة تعطى ب:

$$\chi_c^2 = -\left[n - 1 - \frac{1}{6}(2K + 5)\right] . Ln\ D$$

حيصث n تمثل حجم العينة، K عدد المتغيرات المفسرة (الحد الثابت معدود $K = k + 1$) و Ln يمثل اللوغاريتم النبيري.

يكون القرار على النحو التالي:

- إذا كان $\chi_c^2 \geq \chi_{\frac{1}{2}K(K-1)}^{2\ (\alpha)}$ فإننا نرفض الفرضية H_0 ، أي يوجد تعدد ارتباط خطي عند المستوى α ،

➢ إذا كان $\chi_c^2 < \chi_{\frac{1}{2}K(K-1)}^{2\,(\alpha)}$ فإننا نقبل الفرضية H_0 ، أي لا يوجد تعدد ارتباط خطي عند المستوى α .

5.ت.3. معالجة تعدد الارتباط الخطي:

نشير إلى بعض الطرق الممكنة لإعطاء حلا لمشكلة تعدد الارتباط الخطي:

- زيادة حجم العينة: تكون هاته الطريقة أكثر فعالية لما تكون المشاهدات الإضافية تختلف قيمها عن تلك السابقة المستعملة في بناء النموذج،
- قمة الانحدار الضيق *Ridge Regression*: طريقة رقمية تتمثل في استبدال المصفوفة $(X'X)$ ب $(X'X + cI)$ أين c هو ثابت قيمته اختيارية أين يؤدي زيادة قيم القطر إلى تخفيف أثر تعدد الارتباط الخطي.

أمام هاته الحسابات الاصطناعية، تبقى الطريقة الأكثر فعالية لمعالجة تعدد الارتباط الخطي هي إعادة تشخيص النموذج من خلال اختيار متغير ممثل للمتغيرات التي تمثل نفس الظاهرة (المتغيرات المرتبطة فيما بينها).

في الفقرة التالية، نقدم طرق تسمح بتحديد "العدد الأمثل" للمتغيرات المفسرة في نموذج قياسي.

5.ث. طرق اختيار النموذج الأمثل:

في الواقع التطبيقي، يواجه القياسي دائما الاختيار بين مجموعة من المتغيرات المفسرة $x_1, x_2, \dots, x_k$ المرشحة لتفسير المتغير y حيث توجد طرق إحصائية تسمح بتحديد المتغيرات التي تدخل أو تحذف من النموذج. هاته الطرق يجب التعامل معها بحذر لأنها، على الرغم من قبولها إحصائيا، يكون أحيانا من الصعب تفسيرها اقتصاديا أو يكون تفسيرها بدون معنى اقتصادي.

يضاف إلى إشكالية اختيار المتغيرات المفسر إشكالية اختيار شكل النموذج اين تكون مجموعة من النماذج مقبولة ويبقى الاختيار بين أفضلها. نتحدث بصفة عامة عن اختيار النموذج الأمثل.

إن معيار تعظيم معامل التحديد R^2 يتمثل في اختيار النموذج ذو القيمة العظمى ل R^2. تكمن مساوئ هذا المعيار في عدم قدرته على الموازنة بين فقدان درجة الحرية وزيادة قيمته معامل. لهذا نستعمل معياري أكاييك أو شوارز (Akaike & Schwarz) للمقارنة أو المفاضلة بين النماذج الممكنة. إن طريقة حسابهما تعتمد على مجموع مربعات البواقي لذلك:

نختار النموذج الذي يقلل قيمة دالة أكاييك (Akaike):

$$AIC = Ln\left(\frac{SCR}{n}\right) + \frac{2k}{n}$$

أو يقلل قيمة دالة شوارز (Schwarz):

$$SC = Ln\left(\frac{SCR}{n}\right) + \frac{kLn(n)}{n}$$

مع:

Ln: اللوغاريتم النبيري،

SCR: مجموع مربعات البواقي للنموذج،

n: حجم العلنة (عدد مشاهدات العينة)،

k: عدد المتغيرات المفسرة.

بعد تحديد معايير المفاضلة بين النماذج، سنعرض فيما يلي 5 طرق ممكنة والتي ستسمح بناءا على المعايير باختيار النموذج الأمثل، الذي يتشكل من المتغيرات المفسرة:

- الأكثر ارتباطا مع المتغير التابع،
- الأقل ارتباطا فيما بينها.

5.ث.1. طريقة كل الانحدارات الممكنة:

تتمثل في أبسط الطرق، أين نقوم بتقدير جميع التقديرات أو النماذج الممكنة ($2^k - 1$ تقدير ممكن) ونختار النموذج الذي يعطي أقل قيمتين لمعياري **أكاييك** أو **شوارز** (**Akaike** & **Schwarz**). حدود استعمال هاته الطريقة مرتبط بقيمة k، عدد المتغيرات المفسرة، مثلا من أجل $k = 10$، عدد النماذج أو التوليفات الممكنة هو 1023.

5.ث.2. طريقة الإقصاء التدريجي:

تتمثل هاته الطريقة، بعد تقدير النموذج الكلي مع k متغير مفسر، في الإقصاء التدريجي (إعادة تقدير النموذج بعد كل عملية إقصاء) للمتغيرات التي قيمها لستودنت t أقل من العتبة المختارة. قابلية استعمال هاته الطريقة مرتبطة بالقدرة على تقدير النموذج الكلي، الأمر الذي قد لا يتحقق أحيانا. في الواقع، لما يحوي النموذج الكلي على عدد كبير من المتغيرات المفسرة يزداد احتمال تعدد الارتباط الخطي وبالتالي يصبح من غير الممكن تقديره.

5.ث.3. طريقة الاختيار (الإضافة) التدريجي:

من خلال هاته الطريقة، نختار، في مرحلة أولى، المتغير المفسر x_i الأكثر ارتباطا مع المتغير التابع y لتقدير النموذج الأولي. في مرحل ثانية نقوم بحساب معامل الارتباط الجزئي $(i \neq j)$ $r^2_{y\, x_i\, x_j}$ ونضيف للنموذج المتغير x_j الموافق لأكبر معامل ارتباط جزئي وهكذا دواليك. نتوقف عن الاختيار والاضافة عندما تصبح قيمة ستودنت t لآخر متغير مضاف أقل من العتبة.

5.ث.4. طريقة الانحدار خطوة بخطوة:

هاته الطريقة مماثلة للسابقة فقط لما نضيف أو ندخل متغير مفسر جديد للنموذج نتفحص قيمة ستودنت t لكل متغير مفسر تم اختياره مسبقا ونقصي من النموذج الذي قيمته أقل من العتبة (غير معنوي).

5.ث.5. طريقة الانحدار الطبقي:

هي طريقة اختيار للمتغيرات المفسرة تسمح بتقليل الارتباط البيني بينها من خلال دراسة الباقي، تتم عبر 3 مراحل:

المرحلة 1: يتم اختيار المتغير المفسر x_i الذي لدية أكبر ارتباط بسيط مع المتغير التابع y ،

المرحلة 2: حساب باقي انحدار y على المتغير x_i المختار في المرحلة 1، $e_1 = y - \hat{a}_0 - \hat{a}_1\, x_i$، ثم نحسب معاملات الارتباط البسيطة بين الباقي e_1 وباقي المتغيرات المفسرة ونختار المتغير، x_j ، الموافق لأكبر قيمة ارتباط.

المرحلة 3: نحسب باقي انحدار y على المتغيرين x_i و x_j ، $e_2 = y - \hat{a}_0 - \hat{a}_1\, x_i - \hat{a}_2\, x_j$، ثم نحسب مجددا معاملات الارتباط البسيطة بين الباقي e_2 وباقي المتغيرات المفسرة ونضيف للنموذج المتغير الموافق لأكبر قيمة ارتباط ونتحصل باقي جديد و معاملات ارتباط بسيطة جديدة. نكرر العملية بإضافة متغيرات مفسرة جديدة إلى النموذج ونتوقف لما تكون معاملات الارتباط البسيطة بين الباقي والمتغيرات المفسرة المتبقية غير معنوية (معدومة إحصائيا).

6. إشكالية الارتباط الذاتي للأخطاء:

إن فرضية استقلالية أخطاء التشخيص هي فرضية أساسية في بناء نموذج قياسي مثالي. عدم تحقق هاته الفرضية، أو ما يعرف بإشكالية الارتباط الذاتي للأخطاء ترهن بعض النتائج المتحصل عليها.

6.أ. تقديم الإشكالية:

نذكر أنه لحد الآن، أنه افترضنا لتقدير معالم النموذج تحقق الفرضيات التالية:

- H_1: النموذج خطي بالنسبة ل x_t،
- H_2: قيم المتغير المستقل x_t تمت مشاهدتها بجون خطأ (x_t غير عشوائي)،
- H_3: $E(\varepsilon_t) = 0$، الأمل الرياضي لخطأ التشخيص هو معدوم، أي أن النموذج هو مشخص جيدا في المتوسط ،
- H_4: $E(\varepsilon_t^2) = \sigma^2$، تباين الأخطاء هو ثابت $\forall t$، أي مجال تضاعفه هو نفسه مهما كانت الفترة أو الوحدة الإحصائية، تسمى هاته الفرضية بفرضية ثبات تباين الأخطاء،
- H_5: $E(\varepsilon_t, \varepsilon_{t'}) = 0$، لما $t \neq t'$، أخطاء التشخيص هي غير مرتبطة (أو حتى مستقلة)، أي الخطأ في الزمن t ليس له تأثير على خطأ آخر في الزمن t'،
- H_6: $COV(x_t, \varepsilon_t) = 0$، الخطأ هو مستقل عن المتغير المستقل.

ونذكر أيضا أن تشخيص مصفوفة التباين والتباين المشترك لخطأ التشخيص يعطى ب:

$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \begin{pmatrix} E(\varepsilon_1\varepsilon_1) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2\varepsilon_2) & \dots & E(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E(\varepsilon_n\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 I_n$$

عند عدم تحقق الفرضية H_5 (استقلالية أخطاء التشخيص) يكون $E(\varepsilon_t \varepsilon_{t'}) \neq 0$ من أجل $t \neq t'$ وتصبح مصفوفة التباين والتباين المشترك لخطأ التشخيص $\Omega_\varepsilon = E(\varepsilon\varepsilon') \neq \sigma_\varepsilon^2 I_n$ ($COV(\varepsilon_t, \varepsilon_{t'}) \neq 0$) وتصبح التقديرات المتحصل عليها بطريقة المربعات الصغرى (العلاقة 3.3) على الرغم من عدم تحيزها لكن ليس ذات أقل تباين، يكون لدينا:

$$\Omega_{\hat{a}} = E[(\hat{a} - a)(\hat{a} - a)'] = (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} = (X'X)^{-1}X'\Omega_\varepsilon X(X'X)^{-1}$$

أي نتحصل على تقدير $\hat{a}$ أين قيم قطر مصفوفة تباينه وتباينه المشترك أكبر من قيم قطر مصفوفة تباينه وتباينه المشترك في حالة تحقق الفرضية H_5 ($\sigma_\varepsilon^2 (X'X)^{-1}$).

إشكالية الارتباط الذاتي للأخطاء تجعلنا نطرح التساؤلات:

- ماهي أسباب وطرق الكشف عن وجود الارتباط الذاتي للأخطاء؟
- ماهي طريقة التقدير المناسبة؟
- كيف نتحصل على تقدير جديد ل a؟

6.ب. تقدير المربعات الصغرى المعممة GLS:

ليكن النموذج الخطي العام:

$$\underset{(n,1)}{Y} = \underset{(n,k+1)}{X}\ \underset{(k+1,1)}{a} + \underset{(n,1)}{\varepsilon}$$

أين

$$E(\varepsilon\varepsilon') = \Omega_\varepsilon \neq \sigma_\varepsilon^2 I_n$$

و Ω_ε مصفوفة مربعة ذات بعد n.

نرغب في تقدير مصفوفة المعالم a بنفس خصائص تقدير طريقة المربعات الصغرى، غير متحيز، دالة خطية بالنسبة ل Y وأقل تباين. يبرهن أن هذا التقدير يعطى ب:

$$\hat{a} = (X'\Omega_\varepsilon^{-1}X)^{-1}(X'\Omega_\varepsilon^{-1}Y) \,\ldots.. (6.1)$$

يسمى هذا التقدير ب"تقدير المربعات الصغرى المعممة" أو أيضا "تقدير أيتكان Aitken".

كما يبرهن أيضا أن تباينه يعطى بالعلاقة:

$$\Omega_{\hat{a}} = (X'\Omega_\varepsilon^{-1}X)^{-1} \,\ldots.. (6.2)$$

ملاحظة:

عند استعمال العلاقة (1.6) في حالة تحقق الفرضيات الكلاسيكية (من بينها الفرضية H_5) نتحصل على تقدير طريقة المربعات الصغرى OLS:

$$\hat{a} = (X'\Omega_\varepsilon^{-1}X)^{-1}(X'\Omega_\varepsilon^{-1}Y) = (X'(\sigma_\varepsilon^2 I_n)^{-1}X)^{-1}(X'(\sigma_\varepsilon^2 I_n)^{-1}Y)$$

$$= \left(X'\left(\frac{1}{\sigma_\varepsilon^2}\right)X\right)^{-1}\left(X'\left(\frac{1}{\sigma_\varepsilon^2}\right)Y\right)$$

$$= \sigma_\varepsilon^2\left(\frac{1}{\sigma_\varepsilon^2}\right)(X'X)^{-1}(X'Y)$$

$$= (X'X)^{-1}(X'Y)$$

في التطبيق الواقعي قيم المصفوفة Ω_ε مجهولة، العلاقتين السابقتين (1.6) و (2.6) لا يمكن استعمالهما، فقط في حالات استثنائية.

إذا يجب تقديم طريقة تقدير عملياتية.

6.ت. أسباب وطرق الكشف عن الارتباط الذاتي للأخطاء :

6.ت.1. أسباب الارتباط الذاتي للأخطاء: يكون ارتباط ذاتي للأخطاء عندما تكون الأخطاء مرتبطة بتسلسل (تراتب) تكراري حيث يمكننا أن نميز بين الارتباط الذاتي للأخطاء الموجب والارتباط الذاتي للأخطاء السالب. يمكن أن يشاهد أو يحدث الارتباط الذاتي للأخطاء لأسباب متعددة:

- غياب متغير مفسر (مستقل) مهم الذي يجعل من البواقي مرتبطة ذاتيا،
- تشخيص سيئ للنموذج، مثلا العلاقات بين المتغير التابع والمتغيرات المفسرة هي غير خطية وتكون من شكل رياضي آخر غير شكل النموذج المشخص (لوغاريتمية، نصف لوغاريتمية،...)،
- مسح البيانات بالمتوسطات المتحركة أو استقرائها يخلق ارتباط ذاتي بين الأخطاء نتيجة هاته التحويلات.

يصادف الارتباط الذاتي بين الأخطاء أساسا في النماذج ذو البيانات الزمنية أين يزيد احتمال تأثير خطأ فترة على خطأ فترة أخرى. في حالة نموذج البيانات المقطعية الآنية، يكون ارتباط ذاتي بين الأخطاء فقط في حالة ترتيب قيم المتغير التابع، تنازليا كان أو تصاعديا. في الواقع، جمع البيانات يجب أن يتم بطريقة عشوائية.

6.ت.2. الكشف عن الارتباط الذاتي للأخطاء:

الكشف عن إمكانية وجود ارتباط ذاتي للأخطاء يتم فقط من خلال تحليل بواقي النموذج، الوحيدة المعلومة.

أ- التحليل البياني للبواقي:

التحليل البياني للبواقي يسمح عادة بالكشف عن تسلسل متكرر أو منتظم للبواقي عندما:

- تكون قيم البواقي لفترة طويلة إما موجبة أو سالبة، ويكون هنا ارتباط ذاتي للأخطاء موجب،
- تكون قيم البواقي مرة موجبة ومرة سالبة، ويكون هنا ارتباط ذاتي للأخطاء سالب.

ب-اختبار داربن-واتسون (**Durbin-Watson**):

اختبار داربن-واتسون (DW) يسمح بالكشف عن احتمالية وجود ارتباط ذاتي للأخطاء من الدرجة 1 من الشكل:

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \quad v_t \sim N(0, \sigma_v^2)$$

ويكتب الاختبار:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$

أو $\rho < 0$، $\rho > 0$، **تنبيه**: العتبة تختلف في الاختبار أحادي الطرف).

من أجل اختبار الفرضية $\mathrm{H_0}$، نحسب إحصائية داربن–واتسون (DW):

$$\mathrm{DW} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \ldots\ldots (6.3)$$

أين e_t يمثل الباقي في الزمن t للنموذج المقدر.

من شكله البنائي، تتغير قيمة إحصائية داربن–واتسون (DW) من 0 إلى 4 ويكون $\mathrm{DW} = 2$ لما $\hat{\rho} = 0$ ($\hat{\rho}$ هو تقدير أو مشاهدة ρ). من أجل اختبار الفرضية $\mathrm{H_0}$، داربن–واتسون جدولا قيم حرجة عند العتبة 5% بدلالة حجم العينة n وعدد المتغيرات المستقلة k . قراءة قيم الجدول تسمح بتحديد قيمتين d_1 و d_2 محصورتين بين 0 و 2 وتقسمان المجال من 0 إلى 4 على النحو التالي:

حسب تموضع قيمة داربن–واتسون (DW) على هذا المجال، يمكننا وضع قواعد القرار على النحو:

- $d_2 < DW < 4 - d_2$، نقبل الفرضية $\mathrm{H_0}$ $(\rho = 0)$.
- $0 < DW < d_1$، نرفض الفرضية $\mathrm{H_0}$ $(\rho > 0)$.
- $4 - d_1 < DW < 4$، نرفض الفرضية $\mathrm{H_0}$ $(\rho < 0)$.
- $d_1 < DW < d_2$ أو $4 - d_2 < DW < 4 - d_1$، يمثلان مجال عدم التحديد، أي منطقتي شك اين لا يمكننا الجزم لا بقول الفرضية $\mathrm{H_0}$ أو رفضها.

شروط استعمال اختبار داربن–واتسون (DW):

- أن يحوي النموذج على الحد الثابت،
- أن لا يظهر المتغير التابع كمتغير مستقل (مفسر بتأخير زمني)، في هاته الحالة يمكننا استعمال إحصائية h لداربن أو اختبار بروش–غودفراي (Breusch-Godfrey)،

- من أجل نماذج البيانات المقطعية، يجب أن تكون البيانات مرتبة تصاعديا أو تنازليا بالنسبة للمتغير التابع أو بالنسبة للمتغير المفسر المشكوك في كونه سببا للارتباط الذاتي للأخطاء.
- عدد المشاهدات (حجم العينة) يجب أن يكون أكبر من 15.[2]

اختبار داربن-واتسون (DW) هو اختبار لاحتمال استقلالية الأخطاء كونه يستعمل لذلك قيم البواقي كما أنه يختبر فقط، كما أشرنا، الارتباط الذاتي من الدرجة 1.

ت-اختبار بروش-غودفراي (Breusch-Godfrey):

هذا الاختبار، الذي يعتمد على اختبار فيشر (Fisher) لمعدومية المعالم أو مضاعف لاغرونج (Lagrange)، يسمح باختبار ارتباط ذاتي للأخطاء من درجة أكبر من أو يساوي 1 ويبقى صحيح في حالة ظهور المتغير التابع كمتغير مستقل بتأخير زمني. الفكرة العامة لهذا الاختبار تكمن في البحث عن علاقة معنوية بين الباقي والباقي نفسه بتأخير زمني.

ارتباط ذاتي بين الأخطاء من الدرجة p يكتب:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \cdots + \rho_p\varepsilon_{t-p} + v_t, \quad v_t \sim N(0, \sigma_v^2)$$

ليكن النموذج العام ذو ارتباط ذاتي بين الأخطاء من الدرجة p:

$$y_t = a_0 + a_1x_{1t} + a_2x_{2t} + \cdots + a_kx_{kt} + \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \cdots \rho_p\varepsilon_{t-p} + v_t$$

هذا الاختبار يتم ضمن 3 مراحل:

[2] جدول اختبار داربن-واتسون تبدأ قيمه من أجا العينات ذات الحجم 15.

المرحلة 1: تقدير النموذج باستعمال طريقة المربعات الصغرى OLS وحساب الباقي e_t، كون الأخطاء مجهولة سيتم الاختبار على الباقي،

المرحلة 2: التقدير باستعمال طريقة المربعات الصغرى OLS للعلاقة البينية:

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \cdots + a_k x_{kt} + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \cdots \rho_p e_{t-p} + v_t$$

ليكن n حجم المشاهدات المتوفرة (حذار لأن كل تأخير يفقد العينة مشاهدة) لتقدير النموذج وحساب معامل التحديد R^2.

أحيانا بعض القياسيين يوصون بإعطاء قيم البواقي التأخيرية "0" حتى لا يتقلص عدد المشاهدات أين اختلاف النتائج يكون ملاحظا فقط في حالة العينات الصغيرة.

المرحلة 3: اختبار الفرضيات على العلاقة البينية.

الفرضية H_0 المختبرة لعدم وجود ارتباط ذاتي بين الأخطاء تكتب:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_p = 0$$

عند رفض الفرضية الصفرية، يكون هناك احتمال وجود ارتباط ذاتي بين الأخطاء.

للقيام عمليا بالاختبار نذكر أنه لدينا خيارين: إما القيام باختبار فيشر الكلاسيكي لمعدومية مجموعة من المعالم، أو اللجوء إلى إحصائية LM التي تتبع توزيع χ^2 ذو درجة حرية p بحيث لما $n \times R^2 > \chi_p^{2\,(\alpha)}$ فإننا نرفض الفرضي الصفرية H_0.

6.ث. طريقة التقدير في حالة الارتباط الذاتي للأخطاء:

6.ث.1. مبادئ عامة:

لو نأخذ بفرضية ارتباط ذاتي للأخطاء من الدرجة 1، يكتب النموذج الخطي من الشكل:

$$Y = Xa + \varepsilon$$

مع

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \quad |\rho| < 1 \ldots\ldots (6.4)$$

(تسلسل ذاتي الارتباط من الدرجة 1: $AR(1)$ أين $v_t \sim N(0, \sigma_v^2)$ و $E(v_t v_{t'})$ من أجل $t \neq t'$).

بالتعويض المتتالي في المعادلة (4.6)، نتحصل على:

$$\varepsilon_t = \rho(\rho\varepsilon_{t-2} + v_{t-1}) + v_t = \rho^2\varepsilon_{t-2} + (\rho v_{t-1} + v_t)$$

$$\varepsilon_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \cdots \quad \ldots\ldots (6.5)$$

يؤول هذا التسلسل إلى 0 كون $(|\rho| < 1)$.

ندرس الآن خصائص ε_t.

$$E(\varepsilon_t) = E(\rho\varepsilon_{t-1} + v_t) = \rho E(\varepsilon_{t-1}) + E(v_t) = 0$$

و

$$E(\varepsilon_t^2) = E(\rho\varepsilon_{t-1} + v_t)^2 = \rho^2 E(\varepsilon_{t-1})^2 + E(v_t)^2$$

كون v_t مستقل عن ε_t.

تباين ε_t، كون حتى الآن فرضيا متجانس، يكون لدينا $E(\varepsilon_t)^2 = E(\varepsilon_{t-1})^2 = \sigma_\varepsilon^2$.

ليصبح:

$$\sigma_\varepsilon^2(1 - \rho^2) = \sigma_v^2 \rightarrow \sigma_\varepsilon^2 = \frac{\sigma_v^2}{(1 - \rho^2)} \ldots\ldots (6.6)$$

أيضا:

$$Cov(\varepsilon_t, \varepsilon_{t+1}) = E(\varepsilon_t \varepsilon_{t+1}) = E\big(\varepsilon_t(\rho\varepsilon_t + v_{t+1})\big) = \rho\sigma_\varepsilon^2$$

$$Cov(\varepsilon_t, \varepsilon_{t+2}) = E(\varepsilon_t \varepsilon_{t+2}) = E\big(\varepsilon_t(\rho^2\varepsilon_t + \rho v_t + v_{t+1})\big) = \rho^2\sigma_\varepsilon^2$$

$$Cov(\varepsilon_t, \varepsilon_{t+i}) = E(\varepsilon_t \varepsilon_{t+i}) = E\left(\varepsilon_t(\rho^i\varepsilon_t + \rho^{i-1} v_{t+1} + \cdots + v_{t+i})\right) = \rho^i\sigma_\varepsilon^2 = \frac{\rho^i\sigma_v^2}{(1-\rho^2)}$$

في هاته الحالة تكتب مصفوفة التباين والتباين المشترك لخطأ التشخيص ε، Ω، على الشكل:

$$\Omega_\varepsilon = \frac{\sigma_v^2}{(1-\rho^2)}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

مع $\rho \neq 1$.

حسب العلاقة (1.6)، يعطى تقدير المربعات الصغرى المعممة MCG ب:

$$\hat{a} = (X'\Omega_\varepsilon^{-1}X)^{-1}(X'\Omega_\varepsilon^{-1}Y)$$

مع

$$\Omega_\varepsilon^{-1} = \frac{1}{\sigma_v^2}\begin{pmatrix} 1 & -\rho & 0 & \cdots & & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{pmatrix}$$

في الواقع، لا نعلم لا ρ و لا σ_v^2، إذا نبحث عن تحويل مصفوفي M بحيث يكون النموذج $MY = MXa + M\varepsilon$ أخطاؤه التشخيصية مستقلة ومتجانسة التباين.

ليكن:

$$E(M\varepsilon(M\varepsilon)') = E(M\varepsilon\varepsilon'M') = ME(\varepsilon\varepsilon')M' = M\Omega_\varepsilon M' = \sigma_\varepsilon^2 I$$

في هاته الحالة يمكننا تحديد التقدير غير متحيز والمقارب لشعاع المعالم a بطريقة المربعات الصغرى العادية OLS:

$$\hat{a} = ((MX)'MX)^{-1}(MX)'MY = (X'M'MX)^{-1}X'M'MY \ldots\ldots (6.7)$$

بالمقارنة عنصر بعنصر بين العلاقتين (1) و (7) نتحصل على:

$$M'M = \lambda\Omega_\varepsilon^{-1} = \sigma_\varepsilon^2\Omega_\varepsilon^{-1}$$

لأنه إذا تحقق $(6.7) = (6.1)$ من أجل $M'M = \Omega_\varepsilon^{-1}$ فإنه كذلك من أجل $\lambda\Omega_\varepsilon^{-1}$ مهما تكن λ.

المصفوفة

$$\underset{(n-1)\times n}{M} = \begin{pmatrix} -\rho & 1 & 0 & \cdots & & 0 \\ & -\rho & 1 & 0 & \cdots & 0 \\ & 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 1 & 0 \\ & & & & & -\rho & 1 \end{pmatrix}$$

تحقق هذا الشرط لأنه يمكن التأكد أن:

$$M'M = \begin{pmatrix} \rho^2 & -\rho & 0 & \cdots & & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 1+\rho^2 & -\rho \\ & & & & & -\rho & 1 \end{pmatrix}$$

و $\sigma_v^2\Omega_\varepsilon^{-1}$ هما متماثلتين تقريبا فقط العنصر ρ^2 بدل 1.

كذلك، يمكننا استبدال طريقة GLS ب OLS (لما تكون n كبيرة كفاية) مطبقة على النموذج الخطي

$$MY = MXa + M\varepsilon$$

الذي ما هو إلا النموذج الأولي أين متغيراته مستبدلة **بشبه فروق** من الدرجة الأولى على النحو:

$$MX_x = \begin{pmatrix} y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_t - \rho y_{t-1} \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix} \text{ و } MY = \begin{pmatrix} x_{k2} - \rho x_{k1} \\ x_{k3} - \rho x_{k2} \\ \vdots \\ x_{kt} - \rho x_{t-1} \\ \vdots \\ x_{kn} - \rho x_{kn-1} \end{pmatrix}$$

حالة خاصة: عندما يكون ارتباط ذاتي للأخطاء تام، $\rho = 1$، تصبح العملية السابقة تطبيق طريقة المربعات الصغرى العادية OLS على نموذج متغيراته هي **فروق** من الدرجة الأولى.

نبين ذلك من خلال مثال. ليكن النموذج ذو متغيرين مفسرين:

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t \quad t = \overline{1, n} \ldots\ldots (6.8)$$

مع تحقق العلاقة (4) $(\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \quad |\rho| < 1)$ أين $v_t \sim N(0, \sigma_v^2)$ و $E(v_t v_{t'})$ من أجل $t \neq t'$.

يكتب النموذج (8.6) في الزمن $(t-1)$:

$$y_{t-1} = a_0 + a_1 x_{1t-1} + a_2 x_{2t-1} + \varepsilon_{t-1} \quad t = \overline{2, n} \ldots\ldots (6.9)$$

حساب $\big((6.8) - \rho(6.9)\big)$ يعطي:

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_{1t} - \rho x_{1t-1}) + a_2(x_{2t} - \rho x_{2t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

$$d_t \quad = \quad b_0 \quad + a_1 dx_{1t} \quad + \quad a_2 dx_{2t} \quad + v_t$$

خطأ التشخيص العشوائي v_t يحقق شروط أو فرضيات تطبيق طريقة المربعات الصغرى، OLS، إذا يمكننا تطبيق هاته الأخيرة على المتغيرات المحولة. تقديرات المعالم المحسوبة من النموذج المحول (شبه فروق) تفسر مباشرة على أنها معالم النموذج الأصلي باستثناء معلمة الحد الثابت أين يظهر بالمقارنة أن:

$$a_0(1-\rho)=b_0$$

ومنه

$$a_0=\frac{b_0}{(1-\rho)}$$

بمعرفتنا لحد الآن طريقة تقدير تسمح لنا بتجاوز إشكالية الارتباط الذاتي للأخطاء من الدرجة 1، يكفي لجعلها عملية أن نعرف كيفية تقدير المعلمة ρ.

6.ث.2. طريقة تقدير ρ:

طرق التقدير التالية هي صالحة فقط لما يكون الارتباط الذاتي للأخطاء من الشكل $\varepsilon_t=\rho\varepsilon_{t-1}+v_t$،

أ– التقدير المباشر ل ρ من خلال بواقي تقدير النموذج الأولي:

المرحلة 1: تقدير ρ بطريقتين:

- بالانحدار المباشر ل e_t على e_{t-1}

$$\hat{\rho}=\frac{\sum_{t=2}^{n}e_te_{t-1}}{\sum_t e_t^2}\dots\dots(6.10)$$

- أو من خلال إحصائية داربين–واتسون (DW)

$$\hat{\rho}=1-\frac{DW}{2}\dots\dots(6.11)$$

المرحلة 2: تحويل المتغيرات والانحدار على شبه الفروق:

$$y_t - \hat{\rho} y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho} x_{1t-1}) + \cdots + a_k(x_{kt} - \hat{\rho} x_{kt-1}) + v_t$$

وتكون المعالم المقدرة ب طريقة المربعات الصغرى العادية OLS: $\hat{a}_0 = \frac{\hat{b}_0}{(1-\hat{\rho})}$، $\hat{a}_1$،....، $\hat{a}_k$.

ب- التقدير المتكرر لشعاع المعالم a و ρ (طريقة كوران-أركوت ***Cochrane-Orcutt***):

يتم ذلك من خلال 4 مراحل.

المرحلة 1: إعطاء قيمة ابتدائية ل ρ .

من خلال تقدير مباشر للعلاقة (10.6)، نثبت قيمة ابتدائية ل ρ، لتكن $\hat{\rho}_0 = \hat{\rho}$.

المرحلة 2: الانحدار على شبه الفروق:

$$y_t - \hat{\rho}_0 y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho}_0 x_{1t-1}) + \cdots + a_k(x_{kt} - \hat{\rho}_0 x_{kt-1}) + v_t$$

وتكون المعالم المقدرة : $\hat{a}_0 = \frac{\hat{b}_0}{(1-\hat{\rho}_0)}$، $\hat{a}_1$،....، $\hat{a}_k$.

المرحلة 3: تقدير ρ

من خلال بواقي النموذج المقدر في المرحلة 2 (e_t^1) نحسب قيمة جديدة ل ρ، ولتكن $\hat{\rho}_1$:

$$e_t^1 = y_t - \hat{a}_0 - \hat{a}_1 x_{1t} - \cdots - \hat{a}_k x_{kt}$$

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{n} e_t^1 e_{t-1}^1}{\sum_t (e_t^1)^2}$$

المرحلة 4: الانحدار على شبه الفروق:

$$y_t - \hat{\rho}_1 y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho}_1 x_{1t-1}) + \cdots + a_k(x_{kt} - \hat{\rho}_1 x_{kt-1}) + v_t$$

ثم نحسب منها بواقي جديدة (e_t^2) الأمر الذي يسمح بحساب تقدير جديد $\hat{\rho}_2$.

نكرر العملية حتى استقرار المعالم المقدرة ل a_k (غالبا بعدد 3 أو 4 تكرارات).

ت-طريقة المسح ل هيلدرتش-لي ***Hildretch-Lu***

تتم هاته الطريقة عبر مرحلتين:

المرحلة 1: تحديد نوع الارتباط الذاتي للأخطاء.

من خلال إحصائية داربن-واتسون (DW)، نحدد إن كان هناك ارتباط ذاتي **موجب** أو **سالب** $(\rho < 0$ أو $\rho > 0)$.

المرحلة 1: الانحدار من أجل مجال قيم ممكنة ل ρ.

نعلم مثلا أن $\rho \in [0;1]$، نقدر من أجل جميع قيم المجال بالأخذ بخطوة ثابتة، مثلا $(0,1)$، العلاقة:

$$y_t - \hat{\rho}_i y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho}_i x_{1t-1}) + \cdots + a_k(x_{kt} - \hat{\rho}_i x_{kt-1}) + v_t$$

ونأخذ ب $(\hat{\rho} = \hat{\rho}_i)$ التي تعطي حسب معيار مجموع مربعات بواقي $\left(\sum_t v_{it}^2\right)$ أقل قيمة ممكنة.

بالإمكان ضبط أكثر قيمة ρ من خلال استعمال نفس الطريقة بمجال مضيق وخطوة أصغر، مثلا $(0,01)$.

نشير إلى أن هاته الطريقة هي المثلى بالنسبة لمعيار مربعات بواقي لأننا نختار تقدير ρ الذي يقلل أو يصغر مجموعها.

ث-طريقة المعقولية العظمى:

أكثر تعقيدا بالنسبة للطرق السابقة، حيث تتمثل هاته الطريقة في التقدير المشترك (في آن واحد) لشعاع المعالم a و قيمة ρ من خلال تعظيم دالة المعقولية.

7. إشكالية عدم تجانس تباين الأخطاء (Heterscedasticity):

7.أ. تقديم الإشكالية:

ليكن النموذج $Y = Xa + \varepsilon$ الذي لا يحقق الفرضية H_4، إذا مصفوفة التباين والتباين المشترك للأخطاء تكتب:

$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \begin{pmatrix} E(\varepsilon_1\varepsilon_1) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2\varepsilon_2) & \dots & E(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E(\varepsilon_n\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma^2_{\varepsilon_1} & 0 & \dots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2_{\varepsilon_n} \end{pmatrix} \neq \sigma^2_\varepsilon I_n$$

أين تباينات أخطاء التشخيص، المعبر عنها بالقيم القطرية، هي غير متجانسة (مختلفة أو غير ثابتة). تصادف هاته الإشكالية عادة في النماذج ذات البيانات الآنية أو عندما تكون مشاهدات العينة هي قيم متوسطة. يكون تباين الخطأ مرتبط بقيم المتغير المفسر.

نتائج أو تأثيرات إشكالية عدم تجانس تباين الأخطاء هي مماثلة المتعلقة بوجود ارتباط ذاتي بين الأخطاء، أي:

- التقديرات غير متحيزة،
- تقديرات المربعات الصغرى (OLS) ليست التقديرات ذات أقل تباين.

أسباب حدوث عدم تجانس تباين الأخطاء هي متعددة:

- مثلما أشرنا، عندما تكون مشاهدات العينة تمثل قيم متوسطة محسوبة من عينات ذات أحجام مختلفة،
- تكرار نفس القيمة للمتغير التابع من أجل قيم مختلفة لمتغير مستقل،
- عندما تكون الأخطاء مرتبطة بقيم متغير مستقل، مثلا في نموذج ذو بيانات آنية تباين الاستهلاك يرتفع مع تغير الدخل المتاح.

في حالة تحقق عدم تجانس تباين الأخطاء، يكون اختبار (Chow) متحيزا، حيث يمكننا أن نرفض بالخطأ الفرضية H_0 لاستقرار معالم النموذج. إذا قبلنا، أثناء اختبار (Chow)، الفرضية H_0 فإن الإشكالية لا تطرح، بينما في حالة رفض H_0، فإنه يتعين أولا قبل القيام بالاختبار تصحيح إشكالية عدم تجانس تباين الأخطاء وتطبيق الاختبار على النموذج المصحح.

7.ب. تصحيح إشكالية عدم تجانس تباين الأخطاء:

يبرهن أن تقديرات المعالم غير متحيزة، ذات أقل تباين والمتقاربة لنموذج ذو أخطاء غير متجانسة التباين هي تقدير المربعات الصغرى المعممة (MCG) التي تعطى بالعلاقة:

$$\hat{a} = (X'\Omega_\varepsilon^{-1}X)^{-1}(X'\Omega_\varepsilon^{-1}Y) \dots\dots (7.1)$$

مع

$$\Omega_{\hat{a}} = (X'\Omega_\varepsilon^{-1}X)^{-1} \dots\dots (7.2)$$

لا توجد طريقة تصحيح واحدة معينة لمعالجة إشكالية عدم تجانس تباين الأخطاء على عكس تصحيح إشكالية الارتباط الذاتي بين الأخطاء من الدرجة 1 (مثلما رأينا سابقا)، لكن توجد طرق متعددة يمكن تطبيقها وفقا للسبب المفترض للإشكالية. تتمثل القاعدة العامة لتصحيح النموذج في إيجاد تحويل لبيانات العينة، المتغيرات التابع والمتغيرات والمستقلة، بحيث نتحصل على نموذج أخطاؤه ذات تباين متجانس.

سوف نستعرض فيما يلي كيفية تصحيح إشكالية عدم تجانس تباين الأخطاء في حالة بيانات ممثلة لمتوسطات.

7.ب.1. تصحيح إشكالية عدم تجانس تباين الأخطاء عندما تكون البيانات (مشاهدات العينة) ممثلة لمتوسطات:

نبين ذلك من خلال نموذج خطي بسيط. ليكن النموذج المراد تقديره:

$$y_i = a_0 + a_1 x_i + \varepsilon_i \quad i = \overline{1, n}$$

في حالة بيانات ممثلة لمتوسطات نعلم في الواقع النموذج بين متوسطات المتغير التابع والمستقل التي تكتب من الشكل:

$$\bar{y}_i = a_0 + a_1 \bar{x}_i + \bar{\varepsilon}_i \quad i = \overline{1, n}$$

بحيث

$$V(\bar{\varepsilon}_i) = \frac{\sigma_\varepsilon^2}{n_i}$$

بحيث n_i تمثل عدد المشاهدات الذي متوسطها يمثل القيمة i لمتغيرات النموذج المراد تقديره.

إذا مصفوفة التباين والتباين المشترك للخطأ هي من الشكل:

$$V(\bar{\varepsilon}_i) = \Omega_\varepsilon = \sigma_\varepsilon^2 V = \sigma_\varepsilon^2 \begin{pmatrix} 1/n_1 & 0 & \cdots & 0 & 0 \\ 0 & 1/n_2 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 \quad 0 & & \cdots & 1/n_{n-1} & 0 \\ 0 \quad 0 & & & 0 & 1/n_n \end{pmatrix}$$

كون النموذج تباين أخطائه هي غير متجانسة، نستعمل طريقة المربعات المعممة (GLS) لتقديره لنتحصل على:

$$\hat{a} = (X'\Omega_\varepsilon^{-1}X)^{-1}(X'\Omega_\varepsilon^{-1}Y) = \hat{a} = (X'V^{-1}X)^{-1}(X'V^{-1}Y)$$

مع

$$V^{-1} = \begin{pmatrix} n_1 & 0 & \cdots & 0 & 0 \\ 0 & n_2 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & n_n \end{pmatrix}$$

التي تكون نفس معالم النموذج المراد تقديره.

7.ت. الكشف عن عدم تجانس تباين الأخطاء:

هناك العديد من اختبارات الكشف عن عدم تجانس تباين الأخطاء، نذكر في ما يلي البعض منها.

7.ت.1. اختبار تساوي التباينات: يعبر عنه باختبار الفرضية:

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_2^2$$

يتم هذا الاختبار عبر 3 مراحل:

المرحلة 1: حساب التباين المقدر من أجل كل مجموعة

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2}{n_i - 1}$$

المرحلة 2: حساب التباين الكلي

$$\hat{\sigma}_T^2 = \frac{\sum_{i=1}^{n}(n_i - 1)\,\hat{\sigma}_i^2}{\sum_{i=1}^{n}(n_i - 1)} = \frac{\sum_{i=1}^{n} v_i\,\hat{\sigma}_i^2}{v}$$

مع

$$v = \sum_{i=1}^{n} v_i = \sum_{i=1}^{n}(n_i - 1) \text{ و } v_i = (n_i - 1)$$

المرحلة 3: حساب قيمة كي-دو المحسوبة χ_c^2:

$$\chi_c^2 = Q' = v\,Ln(\hat{\sigma}_T^2) - \sum_{i=1}^{n} v_i\,Ln\left(\hat{\sigma}_i^2\right)$$

التي تتبع توزيع كي-دو بدرجة حرية $(n-1)$، $\chi_{(n-1)}^{2\,(\alpha)}$، ويكون القرار على النحو التالي:

- إذا كان $\chi_c^2 \geq \chi_{(n-1)}^{2\,(\alpha)}$: عند مستوى خطأ (α) نرفض الفرضية H_0، هناك عدم تجانس لتباين الأخطاء (Heteroscedasticity)،
- إذا كان $\chi_c^2 < \chi_{(n-1)}^{2\,(\alpha)}$: بمستوى خطأ (α) نقبل الفرضية H_0، هناك تجانس لتباين الأخطاء (Homoscedasticity).

ملاحظة: يمكن تحسين تقدير Q' بقسمتها على ثابت سلمي C:

$$Q = \frac{Q'}{C} \rightarrow \chi_{(n-1)}^{2\,(\alpha)}$$

أين:

$$C = 1 + \frac{1}{3(n-1)}\left(\left(\sum_{i=1}^{n} \frac{1}{v_i}\right) - \frac{1}{v}\right)$$

7.ت.2. اختبار غولدفالد–كوانت Goldfeld-Quandt: هذا الاختبار يستعمل فقط لما يكون أحد المتغيرات هو سبب عدم تجانس تباين الأخطاء وحجم العين كبير كفاية. يتم عبر 3 مراحل.

المرحلة 1: ترتيب المشاهدات تصاعديا أو تنازليا بالنسبة للمتغير التابع أو المتغير سبب عدم تجانس تباين الأخطاء.

المرحلة 2: إهمال c مشاهدة مركزية. تختار قيمتها عشوائيا لكن من يجب أن تقترب من ربع حجم العينة، $c \approx \frac{n}{4}$. يبقى لدينا عينتين (فترتين) جزئيتين.

المرحلة 3: نقدر نموذجين للفترتين 1 و 2 ونحسب:

$$F^* = \frac{SCR_2/ddl_2}{SCR_1/ddl_1} \sim F^{\alpha}_{(ddl_2, ddl_1)}$$

بحيث نرفض فرضية عدم تجانس تباين الأخطاء لما $F^* \geq F^{\alpha}_{(ddl_2, ddl_1)}$ وذلك عند هامش خطأ α.

7.ت.3. اختبار غليسجر Gleisjer:

اختبار غليسجر لا يسمح فقط بالكشف عن وجود عدم تجانس تباين الأخطاء ولكن أيضا عن شكله. يرتكز هذا الاختبار على العلاقة بين باقي تقدير النموذج الأولي (OLS) e_i، والمتغير المستقل الذي يعتقد أنه سبب عدم تجانس تباين الأخطاء X_i.

يتم الاختبار عبر مرحلتين:

المرحلة 1: التقدير بطريقة المربعات الصغرى العادية OLS للنموذج

$$Y_i = \hat{a}_0 + \hat{a}_1 X_i + e_i \quad i = \overline{1, n}$$

المرحلة 2: انحدار القيمة المطلة للباقي، $|e_i|$، على X_i.

غليسجر يقترح اختبار مختلف أشكال عدم تجانس تباين الأخطاء، مثلا:

- $|e_i| = \hat{a}_0 + \hat{a}_1 X_i + v_i$ ويكون هناك عدم تجانس تباين الأخطاء من الشكل $\sigma_{ui}^2 = k^2 X_i^2$ (k أي عدد ثابت) عندما يكون a_1 معنويا.
- $|e_i| = \hat{a}_0 + \hat{a}_1 X_i^{1/2} + v_i$ ويكون هناك عدم تجانس تباين الأخطاء من الشكل $\sigma_{ui}^2 = k^2 X_i$ (k أي عدد ثابت) عندما يكون a_1 معنويا.
- $|e_i| = \hat{a}_0 + \hat{a}_1 X_i^{-1} + v_i$ ويكون هناك عدم تجانس تباين الأخطاء من الشكل $\sigma_{ui}^2 = k^2 X_i^{-1}$ (k أي عدد ثابت) عندما يكون a_1 معنويا.

7.ت.4. اختبار وايت White:

اختبار وايت قريب من الاختبار السابق، يرتكز على العلاقة المعنوية بين من جهة مربعات البواقي ومن جهة متغير أو عدة متغيرات مستقلة (مفسرة) ومربعاتها في نفس النموذج:

$$e_t^2 = a_1 x_{1t} + b_1 x_{1t}^2 + \cdots + a_k x_{kt} + b_k x_{kt}^2 + a_0 + v_t$$

ليكن n حجم العينة بتقدير معالم النموذج و R^2 معامل التحديد. إذا كان في النموذج المقدر معلمة معنوية (تختلف عن 0)، نقبل فرضية عدم تجانس تباين الأخطاء. يمكننا القيام بالاختبار إما باختبار فيشر التقليدي لمعدومية المعالم:

$$H_0: a_1 = a_1 = \cdots = a_k = b_k = 0$$

ويكون احتمال وجود عدم تجانس تباين الأخطاء عند رفض الفرضية H_0.

وإما بحساب الإحصائية $LM = nR^2$ التي تتبع توزيع $\chi_{P=2k}^2$ (k عدد معالم نموذج وايت المقدرة عدى الحد الثابت). نرفض فرية تجانس تباين الأخطاء لما يكون $nR^2 > \chi_{P=2k}^2$ وذلك عند العتبة المختارة α.

7.ت.5. اختبارات أخرى: اختبار ARCH

النماذج من شكل الانحدار الذاتي الشرطي غير متجانس التباين ARCH (AutoRegressif Conditional heteroscedasticity)، تسمح بنمذجة السلاسل (خاصة السلاسل الزمنية المالية) التي لها تغير أو تشتت آني مرتبطة بالزمن. يمكن من خلال هذه النماذج القيام بتنبؤ ديناميكي للسلسلة بالنسبة لمتوسطها (اتجاهها العام) وتباينها. يجسد اختبار ARCH إما باختبار فيشر التقليدي أو باختبار LM (Lagrange Multiplicateur).

عمليا، يتم الاختبار عبر المراحل التالية:

المرحلة 1: نحسب e_t، باقي نموذج الانحدار الأولي،

المرحلة 2: نحسب e_t^2،

المرحلة 3: تقدير الانحدار الذاتي لتأخير من الدرجة p أين نحتفظ فقط بالتأخيرات المعنوية:

$$e_t^2 = \alpha_1 + \sum_{i=1}^{p} \alpha_i \, e_{t-i}^2$$

أو اختبار الفرضية:

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$$

المرحلة 4: حساب إحصائية مضاعف لاغرونج $LM = nR^2$ مع n: عدد المشاهدات عند التقدير في المرحلة 3، R^2: معامل تحديد التقدير في المرحلة 3.

ويكون القرار كالتالي: إذا كان $nR^2 > \chi_p^{2(\alpha)}$، نرفض الفرضية H_0 عند العتبة α، ونقول أن النموذج يقبل تمثيل ARCH من الدرجة P $(ARCH(p))$. إن اختبار معنوية المعالم α_i لانحدار e_t^2 على e_{t-i}^2 هو الذي يحدد درجة التسلسل ARCH أين تكون غالبا درجة التأخير القصوى هي $p = 3$.

نشير في النهاية أنه في حالة تشخيص عدم تجانس تباين الأخطاء للنموذج من شكل ARCH (أو مشتقاته) فإن طريقة تقدير فإننا نستعمل لطريقة التقدير طرق تقدير النماذج غير خطية المشار إليها في محور "النماذج غير خطية".

8. إشكالية التوزيع غير الطبيعي للأخطاء:

مثلما أشرنا في محور الانحدار الخطي البسيط، تحقق فرضية التوزيع الطبيعي لخطأ التشخيص ليس بالضروري للحصول على مقدرات غير متحيزة ومقاربة لكن عدم تحققها يجعل من الصعوبة بناء اختبارات إحصائية ووضع مجالات ثقة. كما أن ربط أخطاء التشخيص بتوزيع آخر غير التوزيع الطبيعي يحد من هاته الآثار لكن يبقى من الصعوبة البالغة القيام بجميع الاختبارات الإحصائية ووضع مختلف مجالات ثقة.

8.أ. طرق اختبار فرضية طبيعية الأخطاء:

هنا العديد من الطرق للتأكد من صحة فرضية التوزيع الطبيعي للأخطاء. نذكر بعضها في النقاط التالية.

8.أ.1. طريقة تفحص المدرج التكراري للبواقي:

من خلال التفحص العيني للتمثيل البياني للمدرج التكراري لبواقي النموذج المقدر يمكننا أخذ فكرة جيدة عن تحقق فرضية التوزيع الطبيعي. نعلم أن التمثيل البياني للتوزيع الطبيعي يأخذ شكل الجرس المتناظر حيث كلما ابتعد شكل المدرج التكراري عن شكل الجرس المتناظر كلما زاد احتمال عدم تحقق فرضية التوزيع الطبيعي للأخطاء.

8.أ.2. اختبار الطبيعية لجاك-بيرا Jarque-Bera:

يعتمد هذا الاختبار على حساب الإحصائية J.B التي تعطى بالعلاقة:

$$J.B = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right) \dots\dots\dots (8.1)$$

أين S يمثل معامل التناظر *(Skewnes)* و K يمثل معامل التفلطح أو الشكل *(Kurtosis)* اللذان يعطيان على التوالي بالعلاقين العامتين التاليتين:

$$S = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\hat{\sigma}}\right)^3 \dots\dots (8.2)$$

و

$$K = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\hat{\sigma}}\right)^4 \dots\dots\dots (8.3)$$

بحيث $\hat{\sigma}$ هو تقدير الانحراف المعياري. في اختبارنا هذا، X_i يمثل البواقي e_t و $\hat{\sigma}$ يمثل σ_{e_t}.

تحت فرضية التوزيع الطبيعي للبواقي، القيمة $J.B$ تتبع توزيع كي-دو بدرجة حرية 2، $J.B \sim \chi_2^{2(\alpha)}$، بحيث نقبل بفرضية التوزيع الطبيعي للبواقي في حالة $J.B < \chi_2^{2(\alpha)}$.

8.ب. كيفية التعامل مع عدم طبيعية الأخطاء:

سوف نكتفي فقط هنا بذكر بعض الطرق للتعامل مع عدم تحقق فرضية طبيعية الأخطاء دون ذكر تفاصيلها.

1. زيادة حجم العينة: إن زيادة حجم العينة بشكل كاف يسمح بتقارب توزيع الأخطاء نحو التوزيع الطبيعي وفق نظرية النهايات المركزية،
2. استعمال طرق تقدير التي تقلل من تأثير فرضيات التوزيع،
3. ادخال اللوغاريتم، الجذر التربيعي أو تحويل Box-Cox على بيانات المتغير التابع،
4. في حالة تحقيق فرضية طبيعية الأخطاء، يمكن استعمال بعض الاختبارات غير معلميه لتشخيص معالم النموذج كاختبار Mann-Whitney أو اختبار Kruskal-Wallis.

9. النماذج غير خطية:

قمنا لحد الساعة بتحليل النماذج الخطية، لكن غالبا ما تستوجب النظرية الاقتصادية صياغة علاقات ونماذج غير خطية. نميز بين نوعين من النماذج غير الخطية، النماذج غير خطية لكن يمكنها جعلها خطية بإحدى التحويلات الرياضية والنماذج غير خطية مطلقا.

9.أ. النماذج غير خطية لكن يمكن جعلها خطيا بإحدى التحويلات الرياضية:

يشمل هذا النوع من النماذج غير خطية بعض النماذج غير خطية بالنسبة للمتغيرات المستقلة والنماذج غير خطية بالنسبة للمعالم لكن التي بالإمكان جعلها خطية بإحدى طرق التحويلات الرياضية. نذكر بعضها فيما يلي.

9.أ.1. النماذج الأسية:

نبين ذلك من خلال دالة الإنتاج من نوع كوب دوغلاس *Cobb-Douglas* التي تأخذ الشكل الكلاسيكي الأسي التالي:

$$Q = \alpha_0 K^{\alpha_1} L^{\alpha_2}$$

بحيث:

Q، K و L هم على التوالي متغيرات الإنتاج، راس المال والعمل.

α_0، α_1 و α_2 هي معالم الدالة أو النموذج.

يتم كتابة الدالة على شكل خطي بإدخال اللوغاريتم (النبيري مثلا) على طرفيها لتصبح:

$$Ln(Q) = b_0 + \alpha_1 Ln(K) + \alpha_2 Ln(L)$$

إذا تحصلنا على نموذج انحدار خطي متعدد متغيراته هي لوغاريتم متغيرات دالة الإنتاج ومعالمه هي معالمها باستثناء الحد الثابت $\alpha_0 = e^{b_0}$.

9.أ.2. كثيرات الحدود والنماذج العكسية:

تأخذ دالة كثير حدود من الدرجة q الشكل:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k$$

يتم جعلها خطيا بوضع $\left(Z_2 = x^2, \ldots, Z_k = x^k\right)$ لنتحصل على النموذج الخطي:

$$y = \beta_0 + \beta_1 x + \beta_2 Z_2 + \cdots + \beta_k Z_k$$

نشير أنه على الرغم من عدم وجود علاقة خطية بين x، x^2،...، x^k، إلا أنه هناك زيادة في احتمال خطر ظهور إشكالية تعدد الارتباط الخطي. يزداد هذا الخطر كلما نقص حجم العينة. إضافة على ذلك، يجب حد درجة الدالة إلى $q = 4$، وإلا سيكون من الصعب إيجاد تفسير اقتصادي للمعالم.

نذكر بعض النماذج الاقتصادية الممثلة بدالة كثير حدود:

- تقدير الاتجاه العام لسلسلة زمنية مع معرة، مثلا، نقطتي رجوع (عودة):

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

- تقدير دالة التكلفة الكلية:

$$y_t = \beta_0 + \beta_1 Q + \beta_2 Q^2$$ (Q: كمية الإنتاج)

الدالة العكسية تأخذ الشكل:

$$y = \beta_0 + \beta_1 \left(\frac{1}{x}\right)$$

حيث تؤول قيمتها إلى β_0 لما x يؤول إلى ∞.

يستعمل هذا النوع من الدوال (النماذج) في تقدير العلاقة بين التضخم والبطالة (منحنيات فيليبس *Phillips*).

لإرجاعها خطية نضع $Z = \frac{1}{x}$ لتصبح:

$$y = \beta_0 + \beta_1 Z$$

ملاحظات مهمة:

- النماذج غير خطية لكن يمكن جعلها خطيا بإحدى التحويلات الرياضية، بالإمكان تقدير معالمها دون تحويلها إلى شكل خطي باستعمال طريق الانحدار المتعرف عليها لكن بصيغ مختلفة لمقدراتها. غالبا ما تهدف الكتابة الخطية لها إلى توحيد العلاقات المستعملة لتقدير معالمها مع علاقات النماذج الخطية.
- في الواقع يطلق مصطلح النماذج غير خطية فقط على تلك النماذج غير خطية مطلقا والتي لا يمكن جعلها خطيا بإحدى التحويلات الرياضية والت سنشير إليها في النقطة الموالية.

9.ب. النماذج غير خطية مطلقا:

هذا النوع من النماذج غير خطية لا يمكن جعلها خطية بإحدى التحويلات الرياضية. نذكر بعض منها والتي تندرج ضمن نماذج تسمى بنماذج الانتشار (كنمذجة مثلا انتشار (حياة) منتوج ما).

9.ب.1. النموذج اللوجستيكي: هذا النموذج معرف بالعلاقة:

$$y_t = \frac{y_{max}}{1 + br^t}$$

أين b و r هما معلمتا النموذج مع $(0 < \mathrm{r} < 1)$ ، r تمثل سرعة الانتشار، b تمثل قيمة الترتيب بالنسبة للمبدأ و y_{max} تسمى عتبة التشبع. من خصائص هذا النموذج:

لما $t \to \infty$ فإن $y_t \to y_{max}$ ولما $t \to -\infty$ فإن $y_t \to 0$.

ملاحظة: عندما تكون y_{max} معلومة فإن النموذج اللوجستيكي يمكن جعله خطي بإحدى التحويلات الرياضية.

9.ب.2. نموذج غومبارتز Gompertz: هذا النموذج معرف بالعلاقة:

$$y_t = e^{br^t+a} \rightarrow Ln(y_t) = br^t + a$$

أين b و r هما معلمتا النموذج مع $(b < 0)$ و $(0 < \mathrm{r} < 1)$ و $y_{max} = e^a$ تمثل عتبة التشبع. من خصائص هذا النموذج أيضا:

لما $t \rightarrow \infty$ فإن $y_t \rightarrow y_{max}$ ولما $t \rightarrow -\infty$ فإن $y_t \rightarrow 0$.

9.ب.3. طرق تقدير النماذج غير خطية مطلقا:

بالنسبة للنماذج غير خطية مطلقا يكون من المستحيل استعمال طرق تقدير مباشرة لها كطريقة المربعات الصغرى (OLS) حيث يستعمل لذلك طرق رياضية رقمية. الفكرة العامة لطرق التقدير هاته ترتكز على خوارزميات تكرارية أين الأشكال غير خطية للنماذج تكتب على شكل خطي تقريبي باستعمال صيغة تايلور (Taylor) مع إعطاء قيم ابتدائية للمعالم المراد تقديرها. تستعمل طريقة المربعات الصغرى لتقدير معالم النموذج الخطي التقريبي لتقدير معالم جديدة التي بدورها تسمح بإيجاد، دائما باستعمال صيغة تايلور (Taylor)، شكل خطي تقريبي آخر. تتكرر العملية إلى أن تستقر المعالم المقدرة من مرحلة لأخرى (يحدد مؤشر للاستقرارية). نذكر أهم هاته الخوارزميات على سبيل المثال لا على سبيل الحصر ودون تفصيلها: خوارزمية غوس-نيوتن Gauss-Newton، خوارزمية نيوتن-رافسن Newton-Raphson و خوارزمية Berndt, Hall, Hall et Hausman (BHHH).

10. إشكالية البيانات وأخطاء القياس:

عندما قدمنا النموذج الخطي، فرضنا أو اشترطنا لتقديره أن المتغير الداخلي والمتغيرات الخارجية تم مشاهدتها بدون خطأ قياس. في الواقع هاته الفرضية نادرا ما تتحقق لكن يمكن أن نقبل عموما بأن خطأ قياس المشاهدات هو مهمل مقارنة بخطأ تشخيص النموذج. لكن في بعض النماذج، المتغيرات الاقتصادية المتناولة يمكن أن تتميز بأخطاء قياس مهمة ومعتبرة نسبيا كما هو الحال بالنسبة للمتغيرات المتحصل على مشاهداتها من عملية سبر آراء بدلا من قياس مباشر.

فيما يلي يجب أن نميز المتغيرات الصحيحة (والمجهولة)، $y^*, x_1^*, x_2^*, \dots, x_k^*$ عن القيم المشاهدة $y, x_1, x_2, \dots, x_k$.

10.أ. آثار المتغيرات ذات أخطاء قياس:

ليكن النموذج:

$$Y^* = X^* a + \varepsilon$$

بحيث ε تحقق الفرضيات العادية.

نضع:

$$Y = Y^* + v \text{ و } X = X^* + \mu$$

مع الفرضيات: $E(\mu) = 0$، $E(v) = 0$، $E(X^{*\prime}\mu) = 0$، $E(Y^{*\prime}v) = 0$، $E(X^{*\prime}v) = 0$، $E(Y^{*\prime}\mu) = 0$.

لدينا إذا:

$$E(\varepsilon'\mu) = E[(Y^* - X^* a)'\mu] = E(Y^{*\prime}\mu) - a'E(X^{*\prime}\mu) = 0$$
$$E(\varepsilon' v) = E[(Y^* - X^* a)' v] = E(Y^{*\prime} v) - a'E(X^{*\prime} v) = 0$$

إذا لدينا فرضيا استقلالية بين أخطاء قياس المتغيرات، μ و v، وخطأ تشخيص النموذج ε.

العلاقة بين المتغيرات المشاهدة X و Y هي كالآتي:

$$Y^* = Y - v = X^* a + \varepsilon = (X - \mu)a + \varepsilon$$

$$\left.\begin{matrix} Y^* = Y - v \\ X^* = X - \mu \end{matrix}\right\}, Y^* = X^* a + \varepsilon \Rightarrow Y - v = (X - \mu)a + \varepsilon$$

$$\Rightarrow Y = Xa + v - \mu a + \varepsilon$$

$$\Rightarrow Y = Xa + \eta$$

مع

$$\eta = v - \mu a + \varepsilon$$

الخصائص الإحصائية ل η هي:

- $E(\eta) = E(v - \mu a + \varepsilon) = E(v) - E(\mu)a + E(\varepsilon) = 0$
- $E(X'\eta) = E((X^* + \mu)'\eta) = E(X^{*\prime}\eta) + E(\mu'\eta) = E(\mu'\eta)$

لدينا:

$$E(X^{*\prime}\eta) = E(X^{*\prime}v) - E(X^{*\prime}\mu)a + E(X^{*\prime}\varepsilon) = 0$$

ومنه:

$$E(X'\eta) = E(\mu'\eta) = E(\mu'v) - E(\mu'\mu)a + E(\mu'\varepsilon) = -E(\mu'\mu)a \neq 0$$

إذا في نموذج متغيراته بها خطأ قياس، الفرضية H_6 هي غير محققة، $COV(X'\eta) \neq 0$، أي أن X و η هما مرتبطين، طريقة المربعات الصغرى GLS إذا تعطي مقدرات متحيزة.

10.ب. طريقة المتغيرات المساعدة *Instrumental Variables (IV)* :

عندما نجد أنفسنا أمام نموذج ذو أخطاء في متغيراته، $Y = Xa + \eta$، الفرضية H_6 مثلما أشرنا هي غير محققة والتقدير $\hat{a}$ لا يتقارب نحو a. الفرضيات الأخرى فرضيا تبقى محققة.

هدف تقنية أو طريقة المتغيرات المساعدة هي تحديد k متغير $z_1, z_2, \dots, z_k$ بحيث:

$$Z = (z_1, z_2, \dots, z_k) \text{ أين } E(Z'\eta) = 0$$

$COV(Z'X) \neq 0$ و

أي لا توجد أي توليفة خطية للمتغيرات z_k عمودية (مستقلة) على المتغيرات x_k أو بعبارة أخرى المتغيرات Z و X هي مرتبطة.

لدينا إذا:

$$E(Z'Y) = E[Z'(Xa + \eta)] = E(Z'X)a + E(Z'\eta) = E(Z'X)a$$

ليكن:

$$\hat{a} = (Z'X)^{-1}(Z'Y) \ldots\ldots (10.1)$$

يبرهن أن $\hat{a}$ هو تقدير مقارب ل a وتباينه يزداد صغرا كلما زاد الارتباط بين Z و X، تعطى مصفوفة التباين والتباين المشترك للتقدير $\hat{a}$ ب:

$$\widehat{\Omega}_{\hat{a}} = \hat{\sigma}_{\varepsilon}^{2}(Z'X)^{-1}(Z'Z)(X'Z)^{-1} \ldots \ldots (10.2)$$

تكمن صعوبة تطبيق طريقة التقدير هاته في اختيار أو إيجاد المتغيرات المساعدة Z غير مرتبطة ب η وقوية الارتباط مع X. في بعض الحالات، يمكننا ببساطة أخذ المتغير الخارجي مع تأخير زمني بفترة كمتغير مساعد.

10.ت. اختبار أوسمان ***Hausman***:

يسمح اختبار أوسمان (1978) بالكشف عن ارتباط محتمل بين خطا التشخيص ε_t ومتغير أو مجموعة متغيرات خارجية x_{it}. في هاته الحالة لا يمكننا استعمال تقدير المربعات الصغرى غير مقارب، اين يجب اللجوء إلى طريقة المتغيرات المساعدة للتقدير أو طريقة العزوم المعممة (GMM).

لتكن الفرضية $H_0: COV(x_t, \varepsilon_t) = 0$ (المتغير x_t هو متغير خارجي) مقابل الفرضية البديلة $H_1: COV(x_t, \varepsilon_t) \neq 0$ (المتغير x_t هو متغير داخلي).

تحت الفرضية H_0 تقديرات المربعات الصغرى وتقديرات المتغيرات المساعدة هي متقاربة لكن تحت الفرضية H_0 التباين المشترك غير معدوم وتقدير المربعات الصغرى هو متحيز وغير مقارب. يمكن القيام باختبار أوسمان بطريقتين، إما اختبار **الفرق** بين تقيدرات المربعات الصغرى (GLS) وتقديرات المتغيرات المساعدة (IV)، إما عبر اختبار **انحدار متزايد**.

10.ت.1. اختبار الفرق:

نحسب الإحصائية:

$$H = (\hat{a}_{IV} - \hat{a}_{GLS})'[V(\hat{a}_{IV}) - V(\hat{a}_{GLS})]^{-1}(\hat{a}_{IV} - \hat{a}_{GLS})$$

الإحصائي H تتبع توزيع كي-دو بدرجة حرية k، $H \sim \chi_k^2$.

إذا كان $H < \chi_k^{2(\alpha)}$ فإننا نقبل الفرضية H_0 عند المستوى α ، (المتغير x_t هو متغير خارجي)، ونقبل الفرضية البديلة في الحالة العكسية.

10.ت.2. اختبار الانحدار المتزايد:

الطريقة المقترحة من طرف أوسمان لهذا الاختبار تتم عبر 4 مراحل:

1. تقدير النموذج بطريقة المربعات الصغرى اين يكون المتغير التابع هو المتغير الذي نرغب باختبار خارجيته والمتغيرات المستقلة (المفسرة) هي المتغير أو المتغيرات المساعدة، التي تكون غالبا مثلما أشرنا المتغيرات المفسرة بتأخير زمني،
2. تقدير المتغير أو المتغيرات $\hat{x}_{it}$ من خلال النموذج المقدر في المرحلة الأولى،
3. تقدير النموذج المدعم (النموذج الأولي مضاف إليه المتغيرات $\hat{x}_{it}$)،
4. اختبار معنوية (بالنسبة ل 0) المعلمة أو المعالم المتغيرات $\hat{x}_{it}$ في النموذج المقدر في المرحلة 3. إذا كانت هاته المعلمة أو المعالم غير معنوية (اختبار ستودنت أو فيشر) نأخذ بصحة الفرضية H_0 ($COV(x_t, \varepsilon_t) = 0$) أي خارجية المتغير أو المتغيرات المختبرة.

10.ث. طريقة العزوم المعممة:

تستعمل طريقة العزوم المعممة (*Generalized Method of Moments (GMM*) عندما يكون المتغير أو المتغيرات المفسرة (المستقلة) مفترضة خارجية ($COV(x_t, \varepsilon_t) = 0$) و تكون مصفوفة التباين والتباين المشترك لخطأ التشخيص عادية أي $\Omega_\varepsilon \neq \sigma_\varepsilon^2 I_n$. طريقة تقدير العزوم المعممة (GMM) تمزج بين طريقتي المربعات الصغرى المعممة وطريقة المتغيرات المساعدة ويعطى التقدير ب:

$$\hat{a} = \left(X'Z\left(Z'\widehat{\Omega}Z\right)^{-1}Z'X\right)^{-1}\left(X'Z\left(Z'\widehat{\Omega}Z\right)^{-1}Z'Y\right) \ldots\ldots (10.3)$$

مع:

Y: المتغير التابع (المفسر)،

X: المتغيرات المستقلة (المفسرة)،

Z: المتغيرات المساعدة،

$\widehat{\Omega}$: مصفوفة التباين والتباين المشترك للخطأ المقدر (الباقي) في مرحلة أولى بطريقة المتغيرات المساعدة (IV).

يشار إلى أنه في حالة تحقق الفرضية الكلاسيكية أو الأساسية، $\Omega_\varepsilon = \sigma_\varepsilon^2 I_n$، تقدير طريقة العزوم المعممة (GMM)، العلاقة (3.10) تصبح نفسها العلاقة (1.10)، علاقة تقدير طريقة المتغيرات المساعدة.

11. المعادلات الآنية:

هذا المحور هو تمهيد للنماذج التي يتطلب تشخيصها كتابة مجموعة من المعادلات المرتبطة فيما بينها بمجموعة من المتغيرات التي تظهر في مجموعة المعادلات. لا يمكننا، عدى في بعض الحالات، تقدير مجموعة المعادلات (معادلات آنية) باستعمال طريقة المربعات الصغرى العادية OLS على كل معادلة على حدى كما لو أن كل معادلة مستقلة عن الأخرى.

سوف نحاول فيما يلي توضيح كيفية كتابة المعادلات الآنية على شكل مصفوفات وعلى شكلها المختزل (المتغيرات الداخلية معبر عنها بدلالة فقط بدلالة المتغيرات الخارجية). نتناول بعد ذلك إشكالية التعيين، أي إمكانية تقدير المعادلات الآنية لما تكتب على شكلها المختزل قبل أن نقدم طرق التقدير الخاصة بالمعادلات الآنية.

11.أ. المعادلات الهيكلية والمعادلات المختزلة:

عندما نكون أمام نموذج خطي ذو معادلات متعددة، قد يظهر متغير كمتغير خارجي في معادلة وكمتغير داخلي (مفسر) في معادلة أخرى. هذه الصفة المزدوجة لبعض المتغيرات تخلق تحيز لتقديرات المعالم عند استعمال طريقة المربعات الصغرى العادية OLS لتقديرها معادلة بمعادلة. إذا نبحث عن تحويل للنموذج الابتدائي إلى نموذج تكون فيه المتغيرات الداخلية معبر عنها بدلالة فقط بدلالة متغيرات الخارجية.

11.أ.1. النموذج العام:

ليكن النموذج العام ذو g معادلة هيكلية التي تربط بين g متغير داخلي ب k متغير محدد مسبقا:

$$\begin{cases} b_{11}y_{1t} + b_{12}y_{2t} + \cdots + b_{1g}y_{gt} + c_{11}x_{1t} + c_{12}x_{2t} + \cdots + c_{1k}x_{kt} = \varepsilon_{1t} \\ b_{21}y_{1t} + b_{22}y_{2t} + \cdots + b_{2g}y_{gt} + c_{21}x_{1t} + c_{22}x_{2t} + \cdots + c_{2k}x_{kt} = \varepsilon_{2t} \\ \vdots \\ b_{g1}y_{1t} + b_{g2}y_{2t} + \cdots + b_{gg}y_{gt} + c_{g1}x_{1t} + c_{g2}x_{2t} + \cdots + c_{gk}x_{kt} = \varepsilon_{gt} \end{cases}$$

يكتب النموذج العام على شكل مصفوفات:

$$\underset{(g,g)}{B}\ \underset{(g,1)}{Y} + \underset{(g,k)}{C}\ \underset{(k,1)}{X} = \underset{(g,1)}{\varepsilon} \ldots\ldots (11.1)$$

في هاته الكتابة، كل معادلة تكون بها بعض المعالم معدومة والمتغيرات التي معالمها تساوي 1 هي المتغيرات التابعة. إذا كانت بمعادلة كل معاملاتها غير معدومة، مساوية ل 1 ولا تحوي الحد العشوائي هذا يعني أنها معدلة تعريفية (لا يوجد بها معلمة للتقدير).

إذا كانت المصفوفة B منتظمة (قابلة للعكس)، ننتقل من الصيغة الهيكلية إلى الصيغة المختزلة من خلال كتابة Y بدلالة X:

$$Y = -B^{-1}CX + B^{-1}\varepsilon \ldots\ldots (11.2)$$

ويمكننا تطبيق طريقة المربعات الصغرى OLS لتقديرها (نشير إلى أن $B^{-1}\varepsilon$ هي مستقلة عن X).

إن كتابة الصيغة يبدو سهلا لكن تقدير المعالم هو من الصعوبة. في الواقع معرفة العناصر $(g \times k)$ للمصفوفة $B^{-1}C$ لا يسمح بتحديد (تعيين) عناصر المصفوفة B (المكونة من$(g \times g)$ عنصر) وعناصر المصفوفة C (المكونة من$(g \times k)$ عنصر). يكون لدينا هنا $(g \times k)$ معادلة مع $(g \times g) + (g \times k)$ مجهول أين يكون، من دون قيود إضافية، من المستحيل حلها، يتعلق الأمر بإشكالية التعيين.

سنتناول لاحقا قواعد بسيطة تسمح، من خلال وضع قيود على المعالم، تحديد شروط التعيين.

11.أ.2. حالة خاصة: النماذج التراجعية

نظام أو مجموعة معادلات تسمى "تراجعية" إذا كان كل متغير داخلي (تابع) بها يمكن أن يحدد بطريقة تسلسلية:

$$y_{1t} = f(x_{1t}, x_{2t}, \ldots, x_{kt};\ \varepsilon_{1t})$$

$$y_{2t} = f(x_{1t}, x_{2t}, \ldots, x_{kt}; y_{1t};\ \varepsilon_{2t})$$

$$y_{3t} = f(x_{1t}, x_{2t}, \ldots, x_{kt}; y_{1t}; y_{2t};\ \varepsilon_{3t})$$

المتغيرات العشوائية ε_{it} هي مستقلة فرضيا.

نلاحظ أن المعادلة الأولى لا تحوي أي متغير داخلي كمتغير خارجي. المعادلة الثانية يظهر فيها فقط المتغير الداخلي للمعادلة الأولى كمتغير خارجي وهكذا دواليك. عندما تستجيب مجموعة معادلات آنية لحالة أو خاصية التراجعية يكون بالإمكان استعمال طريقة المربعات الصغرى العادية OLS لتقديرها معادلة معادلة. في الواقع في هاته الحالة الخاصة، توجد استقلالية بين المتغيرات الداخلية والأخطاء. مثلا في المعادلة الثانية، y_{1t} مرتبط ب ε_{1t} لكن ليس ب ε_{2t}.

النماذج التراجعية تسمى أيضا النماذج المثلثية كون معالم المتغيرات الداخلية تشكل مثلث في المصفوفة B.

11.ب. إشكالية التعيين:

11.ب.1. قيود على المعالم:

يوجد قيد على معلمة في الصيغة الهيكلية للنموذج عندما نضبطه بقيمة محددة. نميز بين نوعين من القيود.

- **قيود الإقصاء**: يمكننا اعتبار أنه كلما لا يظهر متغير داخلي أو خارجي في معادلة هيكلية (النموذج العام) فإنه مرفق بمعلمة محددة معدومة.
- **قيود خطية**: بعض أشكال تشخيصات النماذج تفرض على بعض المتغيرات أن ترفق بنفس المعلمة، نقول أن هناك قيد خطي مسبق على معالم النموذج.

11.ب.2. شروط التعيين:

تتحدد شروط التعيين معادلة بمعادلة. يمكننا التمييز بين ثلاث حالات تعيين:

- النموذج أقل تعيين أو غير معين إذا كانت معادلة منه غير معينة (يوجد معادلات أقل من المعالم المراد تعيينها في الصيغة الهيكلية، النموذج غير قابل للحل)،
- النموذج معين تماما إذا كانت جميع معادلاته معينة،
- النموذج أكثر تعيين إذا كانت معادلاته معينة تماما أو أكثر تعيينا (يوجد معادلات أكثر من المعالم المراد تعيينها).

إذا كان النموذج أقل تعيين، لا توجد أية إمكانية لتقدير معالمه اين يجب إعادة تشخيصه.

شروط التعيين يمكن أن تكون موضوع أكثر تعقيدا، نكتفي في ما يلي فقط بعرض القواعد البسيطة.

ليكن:

g: عدد المتغيرات الداخلية للنموذج (أو عدد معادلات النموذج).

k: عدد المتغيرات الخارجية للنموذج،

g': عدد المتغيرات الداخلية التي تظهر في معادلة،

k': عدد المتغيرات الخارجية التي تظهر في معادلة.

عندما تكون القيود فقط قيود إقصاء، حالات التعيين للمعادلات تكون كالآتي:

- $g-1 > g-g'+k-k'$: المعادلة أقل تعيين،
- $g-1 = g-g'+k-k'$: المعادلة معينة تماما،
- $g-1 < g-g'+k-k'$: المعادلة هي أكثر تعيين.

الذي يمكن أ يلخص، لكي لا تكون معادلة أقل تعيينا، عدد المتغيرات المقصاة (المعالم) يجب أن يكون على الأقل مساوي لعدد معادلات النموذج ناقص 1.

عندما يكون لدينا r قيد غير الاقصاء فيما يتعلق بمعادلة، حالات التعيين السابقة تصبح:

- $g-1 > g-g'+k-k'+r$: المعادلة أقل تعيين،
- $g-1 = g-g'+k-k'+r$: المعادلة معينة تماما،
- $g-1 < g-g'+k-k'+r$: المعادلة هي أكثر تعيين.

هاته الشروط الأساسية تسمى شروط التعيين.

11.ت. طرق التقدير:

طرق التقدير التي يمكن أن تستعمل في المعادلات الآنية هي مرتبطة أو تابعة لشروط التعيين للنموذج.

- إذ كان النموذج غير معين، ليس هناك إمكانية تقدير.

في حالة نموذج معين تماما أو أكثر تعيين، يمكننا أن نميز بين طريقة التقدير حسب شروط التعيين للمعادلة:

- إذا كانت المعادلة معينة تماما، طريقة المربعات الصغرى غير مباشرة أو طريقة المربعات الصغرى المزدوجة،
- إذا كانت المعادلة أكثر تعينا: طريقة المربعات الصغرى المزدوجة.

11.ت.1. طريقة المربعات الصغرى غير مباشرة (ILS):

طريقة المربعات الصغرى غير مباشرة (Indirect Least Squares) تتمثل في تطبيق أو استعمال طريقة المربعات الصغرى على معادلات النموذج المختصر المعينة تماما، تنقسم إلى 3 مراحل:

1. كتابة النموذج الهيكلي على الشكل المختزل،
2. التقدير بطريقة المربعات الصغرى لمعالم كل معادلة،
3. تحديد معالم المعادلات الهيكلية من خلال علاقات جبرية بين المعالم المختزلة والهيكلية.

تقدير المربعات الصغرى غير مباشرة (ILS) للنموذج المختزل هو غير متحيز، مقارب وذو أقل تباين. لكن، تقدير معالم النموذج الهيكلي المتحصل عليها من خلال تقدير المربعات الصغرى غير مباشرة (ILS) للنموذج المختزل هي متحيزة من أجل العينات الصغيرة. الخصائص التماثلية تجعل التقارب غير متحيز عند زيادة حجم العينة.

تقدير المربعات الصغرى غير مباشرة (ILS) نادرا ما يستعمل لصعوبة تحديد الشكل المختزل للنموذج، نفضل عليه تقدير طريقة المربعات الصغرى المزدوجة (DLS) لبساطة استعماله ويعطي نتائج مماثلة من أجل المعادلات المعينة تماما.

11.ت.2. طريقة المربعات الصغرى المزدوجة (DLS):

طريقة المربعات الصغرى المزدوجة (Double Least Squares) هي الأكثر استعمالا في الواقع. تستعمل من أجل النماذج المعينة تماما أو أكثر تعيينا. ترتكز، مثلما تشير تسميتها، على تطبيق طريقة المربعات الصغرى العادية (OLS) على مرحلتين.

ليكن نموذج المعادلات الآنية ذو g متغير داخلي و k متغير خارجي:

$$\begin{cases} b_{11}y_{1t} + b_{12}y_{2t} + \cdots + b_{1g}y_{gt} + c_{11}x_{1t} + c_{12}x_{2t} + \cdots + c_{1k}x_{kt} = \varepsilon_{1t} \\ b_{21}y_{1t} + b_{22}y_{2t} + \cdots + b_{2g}y_{gt} + c_{21}x_{1t} + c_{22}x_{2t} + \cdots + c_{2k}x_{kt} = \varepsilon_{2t} \\ \vdots \\ b_{g1}y_{1t} + b_{g2}y_{2t} + \cdots + b_{gg}y_{gt} + c_{g1}x_{1t} + c_{g2}x_{2t} + \cdots + c_{gk}x_{kt} = \varepsilon_{gt} \end{cases}$$

تتمثل المرحلة الأولى لطريقة المربعات الصغرى المزدوجة (DLS) في القيام بانحدار لكل متغير داخلي على حدى على جميع المتغيرات الخارجية:

$$\begin{cases} y_{1t} = \alpha_{11}x_{1t} + \alpha_{12}x_{2t} + \cdots + \alpha_{1k}x_{kt} + u_{1t} \\ y_{2t} = \alpha_{21}x_{1t} + \alpha_{22}x_{2t} + \cdots + \alpha_{2k}x_{kt} + u_{2t} \\ \vdots \\ y_{gt} = \alpha_{g1}x_{1t} + \alpha_{g2}x_{2t} + \cdots + \alpha_{gk}x_{kt} + u_{gt} \end{cases}$$

وفي مرحلة ثانية، تعوض المتغيرات الداخلية التي تظهر على يمين المعادلات الهيكلية بقيمها المقدرة بنماذج في المرحلة الأولى على النحو التالي:

$$\begin{cases} y_{1t} = \beta_{11}\hat{y}_{2t} + \cdots + \beta_{1g}\hat{y}_{gt} + c_{11}x_{1t} + c_{12}x_{2t} + \cdots + c_{1k}x_{kt} + \varepsilon_{1t} \\ y_{2t} = \beta_{21}\hat{y}_{1t} + \cdots + \beta_{2g}\hat{y}_{gt} + c_{21}x_{1t} + c_{22}x_{2t} + \cdots + c_{2k}x_{kt} + \varepsilon_{2t} \\ \vdots \\ y_{gt} = \beta_{g1}\hat{y}_{1t} + \beta_{g2}\hat{y}_{2t} \ldots + c_{g1}x_{1t} + c_{g2}x_{2t} + \cdots + c_{gk}x_{kt} + \varepsilon_{gt} \end{cases}$$

خصائص تقدير طريقة المربعات الصغرى المزدوجة (DLS) هي مماثلة تقريبا لخصائص التقدير الكلاسيكي، أي انها بالنسبة للعينات الصغيرة تكون متحيزة.

يشار إلى أن تقدير طريقة المربعات الصغرى المزدوجة (DLS) يمكن أن يفسر على أنه تقدير طريقة المتغيرات المساعدة (IV) أين المتغيرات الخارجية للمعادلات الأخرى هي المتغيرات المساعدة.

11.ت.3. طرق أخرى للتقدير:

يمكننا ذكر 3 طرق أخرى لتقدير المعادلات الآنية:

4. طريقة المربعات الصغرى الثلاثية هي ملائمة في حالة الارتباط الذاتي للأخطاء وحالة عدم تجانس تباين الأخطاء. تتمثل في تحديد تقدير المربعات المزدوجة ثم حساب تقدير المربعات الصغرى المعممة (GLS).
5. طريقة المعقولية العظمى ذات المعلومة الكاملة (MVCI) المتمثلة في تعظيم دالة لوغاريتم-المعقولية للنموذج.
6. طريقة العزوم المعمم (GM) تستعمل في حالة عدم تجانس تباين الأخطاء.

خـــــــاتمة

في ختام هذه المطبوعة، نأمل أن تكون المادة العلمي للمحاورة المقدمة قد أسهمت في ترسيخ المبادئ الأساسية للاقتصاد القياسي لدى الطلبة، ومهدت لهم الطريق لفهم هذا الميدان الحيوي الذي أصبح ضرورة لكل باحث وممارس اقتصادي يسعى إلى تفسير الظواهر الاقتصادية بشكل علمي دقيق.

لقد حرصنا من خلال هذه المحاضرات على تقديم عرض نظري مبسط وممنهج للمفاهيم الأساسية في الاقتصاد القياسي، مكتفين بإطار نظري عام يتناسب مع المستوى التكويني للطلبة. ومع ذلك، تجدر الإشارة إلى أن بعض المحاور التي تناولناها قد تحتاج بل تحتاج إلى تفصيلات أوسع وتعمق إضافي لا يتسع له إطار هذه المطبوعة. ولهذا الغرض، يمكن للطلبة الكرام الرجوع إلى قائمة المراجع المعتمدة والمقترحة في نهاية المطبوعة، لمزيد من الاستيضاح والاطلا ع الموسع على المواضيع ذات الصلة.

كما ننوّه إلى أن هناك العديد من المحاور المتقدمة والمتخصصة، على غرار نماذج السلاسل الزمنية، النماذج العشوائية، نماذج الانحدار الذاتي والمتجهات الذاتية (VAR) ، وأساليب القياس الاقتصادي المعمق، لم نتناولها في هذه المطبوعة. نخصص لها أعمالا علميًا مستقلة تتناول هذه المواضيع بشيء من التفصيل والدقة الأكاديمية اللازمة.

في الختام، نؤكد على أهمية الجمع بين الجانب النظري والجانب التطبيقي في دراسة الاقتصاد القياسي، وندعو الطلبة إلى الاستفادة من المطبوعة المكملة الخاصة بالتمارين التطبيقية لتعزيز مهاراتهم التطبيقية وترسيخ مكتسباتهم النظرية.

قائـــــمة المراجع:

1. Anil K. BERA, Matthew L. Higgins, “ARCH Models: Propertie, Estimation and Testing”, University of Wisconsin-Milwaukee”,
2. Bollerslev, T. (1986), «GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY”, Journal of Econometrics, University of California
3. Breusch, T.S and Pagan, A.R (1979) « A simple test for heteroscedastisity and random coefficient variation », Econometrica,
4. Bruce E. Hansen (2014), “Econometrics”, University of Wisconsin, USA.
5. Christian Francq, JM Zaköan (2010), “GARCH Models”, WILEY Engle, R.F (1982), “Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. Inflation.”, Econometrica
6. Engle, R. (1982); "Autoregressive Conditional Heteroskedasticity with Estimates of United Kingdom Inflation", Econometrica;
7. Eric Zivot (2009), « Maximum Likelihood Estimation », (livre électronique)
8. Emmanuel Duguet (2008), « Econométrie des Variables Qualitatives », (livre électronique)
9. George Casella, Roger L. Berger (2001), “Statistical Inference”, 2nd Edition, DUXBURY.
10. Roberts Perrilli, (2001), « Introduction to ARCH and GARCH models », University of Illinois,
11. Régis Bourbonnais (2009), « Econométrie », 7ème édition, DUNOD,
12. Régis Bourbonnais et Michel Terraza. (2008) « Analyse des séries temporelles », DUNOD,
13. Régis Bourbonnais. (2009) « Econométrie», DUNOD,
14. Samuelson, (1965), « Proof that Properly Anticipated Prices Fluctuate Randomly »,