المحور الرابع:

المشاكل القياسية (اختراق فرضيات النموذج)

(اعتدالية توزيع البواقي، استقلالية البواقي، تجانس البواقي، التعدد الخطي بين المتغيرات المستقلة)

إعداد الدكتور: إبراهيم رحيم-2023-2024

4-المشاكل القياسية (اختراق فرضيات النموذج)

يغطي هذا المحور مشاكل قياسية تواجه الباحث، ويتعلق كل مشكل منها بإسقاط إحدى الفرضيات الكلاسيكية لطريقة المربعات الصغرى العادية (OLS) **وهي** أشهر الطرق المستخدمة، وتتمثل أهم هذه المشاكل في الآتي:

1|4 اعتدالية التوزيع الاحتمالي للبواقي Normality test:

حتى يمكن استخدام كل من (**F-test**) و (**t-test**)، عند اختبار المعنوية الكلية أو الجزئية لنموذج الانحدار، يلزم توفر شرط اعتدالية التوزيع الاحتمالي للبواقي. وللإشارة فإن التقيد بهذا الشرط يكون في حالة العينات الصغيرة، أما في حالة العينات الكبيرة ($n \geq 30$) فيمكن التخلي عنه وفقا لنظرية النهاية المركزية حيث تؤول التوزيعات الاحتمالية إلى التوزيع الطبيعي.

1|1|4 الفروض الإحصائية:

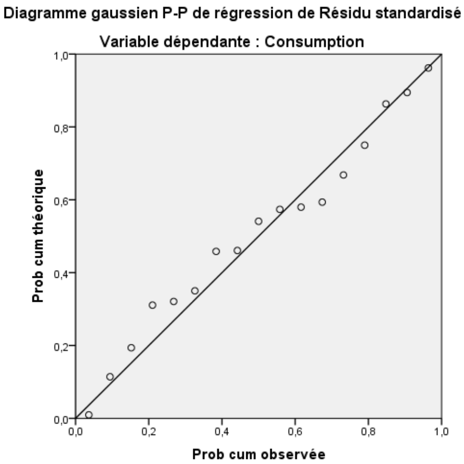
الفرضية الصفرية(H_0) : تتوزع البواقي توزيعا طبيعيا

الفرضية البديلة (H_1) : لا تتوزع البواقي توزيعا طبيعيا

وتتم دراسة اعتدالية التوزيع الاحتمالي للبواقي بطريقتين (حسب برنامج SPSS):

2|1|4 الطريقة البيانية: من خلال الشكل البياني للعلاقة بين الاحتمال التجميعي المشاهد والاحتمال التجميعي المتوقع للبواقي المعيارية بحيث:

- إذا كانت النقط تقع بشكل متقارب جدا من الخط الواصل بين الركن الأيمن العلوي والركن الأيسر السفلي، أو تتوزع بشكل عشوائي على جانبي هذا الخط يقال أن الأخطاء تتوزع توزيعا طبيعيا.
- أما إذا تم رصد نمط معين لتوزيع هذه النقط، فيقال أن الأخطاء لا تتوزع توزيعا طبيعيا."



3|1|4 الطريقة الحسابية: يوفر برنامج SPSS اختبارين لفحص النموذج من خلال تحليل البواقي وهما:

- اختبار " كولموغوروف- سميرنوف " Kolmogorov- Smirnov
- اختبار " شابيرو- وايلك " Shapiro- Wilk

حيث يمكن إنجاز الاختبارين في آن واحد أو الاكتفاء بأحدهما.

2|4 عدم الارتباط الذاتي بين البواقي:

تعود أهمية دراسة الارتباط الذاتي للبواقي في تحليل الانحدار، إلى أن وجود هذا الارتباط من شأنه أن يجعل قيمة تباين الخطأ مقدرة بأقل من قيمته الحقيقية. وبالتالي فإن قيم إحصاءات الاختبارات التي تعتمد على هذا التباين، مثل (t) و (F) و (R^2) تكون أكبر من قيمها الحقيقية، مما يجعل القرار الخاص بجودة توفيق النموذج قرار مشكوك في صحته.

نفترض أن الفرضية الثالثة للنموذج غير محققة، بمعنى: $\exists i \neq j: Cov(\varepsilon_i,\ \varepsilon_j) \neq 0$

أي هناك ارتباط ذاتي بين الأخطاء، ونطرح الأسئلة الآتية:

1) ما هي الأسباب التي تؤدي إلى وجود ارتباط ذاتي بين الأخطاء؟
2) كيف يمكن الكشف عن وجود ارتباط ذاتي بين الأخطاء؟
3) كيفية النمذجة في ظل اختلال فرضية عدم الارتباط الذاتي بين البواقي؟

الإجابة:

4|2|1 الأسباب المؤدية إلى وجود ارتباط ذاتي بين الأخطاء، منها:

- غياب متغير مستقل في النموذج قد يؤدي إلى اختلال الفرضية المذكورة؛
- احتمال أن تكون النمذجة غير صحيحة، كوجود علاقة غير خطية بين المتغيرين التابع والمستقل.

ومن نتائج اختلال هذه الفرضية هو: الحصول على مقدرات متحيزة وغير متسقة.

4|2|2 الكشف عن وجود ارتباط ذاتي بين الأخطاء: يمكن الاعتماد على الأسلوب الإحصائي التالي:

اختبار دربن–واتسونDurbin–Watson

اقترح $(D.W)$ اختبارا يعتمد على سيرورة الانحدار الذاتي من الرتبة الأولى حيث: $\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \end{cases}$

ρ : هو معامل الارتباط الذاتي بين الأخطاء، حيث: $|\rho| \le 1$; $\hat{\rho} = \frac{\sum \hat{\varepsilon}_t \cdot \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_{t-1}^2}$

η_t : هو حد الخطأ العشوائي في نموذج الانحدار الذاتي $AR(1)$.

4|2|2أ) <u>**الاختبار**</u>: تعتمد اختبار دربن–واتسون على الإحصاءة المقدرة $\widehat{DW}$ التي تختبر المعنوية الإحصائية لمعامل الارتباط الذاتي.

- **الفروض الإحصائية:**

الفرضية الصفرية: للاستقلال (عدم الارتباط) الذاتي بين الأخطاء، بمعنى: $H_0 : \rho = 0$

الفرضية البديلة: الارتباط الذاتي بين الأخطاء $H_1 : \rho \neq 0$

حيث: $\widehat{DW}_c = \frac{\sum^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum \hat{\varepsilon}_{t-1}^2} \sim DW(T, k, \alpha)$

T : حجم العينة، k عدد المتغيرات المستقلة في النموذج، α: نسبة الدلالة أو المخاطرة.

- **العلاقة بين $\widehat{DW}_c$ و $\hat{\rho}$:** $\boxed{\hat{\rho} = 1 - \frac{\widehat{DW}}{2} \ : \ \widehat{DW} \in [0,\ 4]}$

<u>**ملاحظة:**</u>

تحتاج الإحصاءة $\widehat{DW}_c$ إلى قيمتين حرجتين dl و du، اللتان تعتمدان على حجم العينة T، وعدد المتغيرات المستقلة في النموذج k، ونسبة الدلالة α. ويتم الحصول على القيمتين الحرجتين من خلال " جدول دربن–واتسون".

4|2|2ب) اتخاذ القرار:

0 — dl	dl — du	du — 2 — $4-du$	$4-du$ — $4-dl$	$4-dl$ — 4
$\hat{\rho} > 0$ ارتباط ذاتي موجب نرفض H_0	منطقة شك (فشل الاختبار)	$\hat{\rho} = 0$ استقلالية الأخطاء نقبل H_0	منطقة شك (فشل الاختبار)	$\hat{\rho} < 0$ ارتباط ذاتي سالب نرفض H_0

1) **4|2|3 النمذجة في ظل اختلال فرضية الاستقلال الذاتي بين الأخطاء**

نستخدم شبه الفروقات للمتغيرين X و Y والمتمثلة في: $X_i - \rho X_{i-1}$ و $Y_i - \rho Y_{i-1}$، وبالتالي نفقد المشاهدة الأولى لكل متغير. ولتجنب ضياعها نضع: $Y_1^* = Y_1\sqrt{1-\rho^2}$ ، $X_1^* = X_1\sqrt{1-\rho^2}$

نحصل على النموذج المصحح التالي:

$$\mathbf{Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(1-\rho)(X_t - \rho X_{1-t}) + (\varepsilon_t - \rho\varepsilon_{t-1})}$$

إذن يأخذ النموذج المصحح الشكل التالي: $Y_t^* = \beta_0{}^* + \beta_1{}^* X_t^* + u_t$

ويكون النموذج المقدر بطريقة (OLS) هو: $\widehat{Y}_t^* = E(Y_t^*) = \widehat{\beta_0}^* + \widehat{\beta_1}^* X_t^*$

حيث: $\widehat{\beta_0}^* = \boldsymbol{\widehat{\beta_0}(1-\widehat{\rho})}$ و $\widehat{\beta_1}^* = \boldsymbol{\widehat{\beta_1}(1-\widehat{\rho})}$

<u>سؤال:</u> كيف يتم تقدير معامل الارتباط الذاتي (ρ)؟

<u>الجواب:</u> يتم تقدير معامل الارتباط الذاتي (ρ) بعدة طرق، منها طريقة $D.W$ المباشرة: $\hat{\rho} = 1 - \widehat{DW}/2$

(محمد شيخي، 2011، ص ص 104-106)

أما $\widehat{\beta_0}$ ، $\widehat{\beta_1}$ ، $\widehat{DW}$ فهي تحدد من النموذج الأصلي المقدر: $\widehat{Y}_t = \widehat{\beta_0} + \widehat{\beta_1} X_t$

4|3 اختبار التجانس (ثبات التباين) Homoscedasticity:

يترتب عن عدم ثبات التباين في نموذج الانحدار نفس الآثار في حالة وجود ارتباط ذاتي بين البواقي، حيث تكون الأخطاء المعيارية مقدرة بأقل من قيمتها الحقيقية. وبالتالي تصبح هذه المقدرات متحيزة (**Biased**)، مما يجعل نتائج الاستدلال الإحصائي مشكوك في صحتها.

نفترض أن الفرضية الثانية للنموذج غير محققة، بمعنى: $\exists i: Var(\varepsilon_i) \neq \sigma_\varepsilon^2$

أي أن تباين الأخطاء غير ثابت (يتغير بتغير الزمن)، ونطرح الأسئلة الآتية:

1) ما هي الأسباب التي تؤدي إلى عدم التجانس (عدم ثبات التباين)؟
2) كيف يمكن الكشف عن عدم التجانس؟
3) كيفية النمذجة في ظل اختلال فرضية عدم التجانس؟

الإجابة:

4|3|1 الأسباب المؤدية إلى عدم تجانس البواقي، منها:

- مشكل في النمذجة (احتواء النموذج على متغير مبطأ بالنسبة للمتغير التابع)؛ فقد تختل الفرضية المذكورة؛
- تحسن أساليب تجميع البيانات، وهذه يقلل من الأخطاء المرتكبة في القياس، ومن ثم سوف يقل تباين حد الخطأ.

ومن نتائج اختلال هذه الفرضية هو: الحصول على مقدرات متحيزة وغير متسقة.

4|3|2 الكشف عن مدى التجانس:

4|3|2|أ) الفروض الإحصائية:

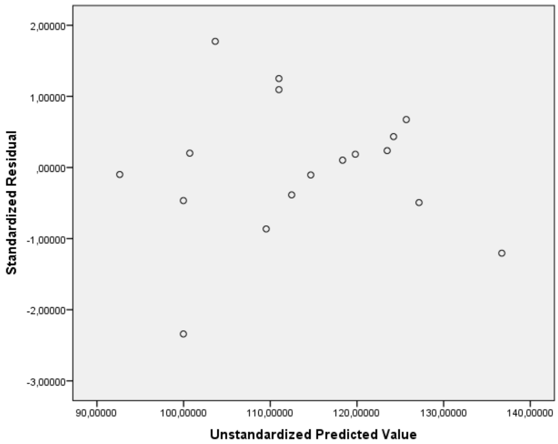
الفرضية الصفرية: تجانس البواقي (ثبات التباين)

الفرضية البديلة: عدم تجانس البواقي (عدم تساوي التباينات)

يمكن الاعتماد على الأسلوب البياني أولا ثم على الأسلوب الإحصائي التي يوفرها برنامج SPSSعلى النحو التالي:

4|3|2|ب) طريقة الرسم البياني لاختبار ثبات التباين

يلاحظ أن انتشار وتوزيع البواقي يأخذ شكلا عشوائيا على جانبي الخط الذي يمثل الصفر (الخط الفاصل بين البواقي السالبة والبواقي الموجبة)، حيث لا يمكننا رصد شكل معين لتباين هذه البواقي، وهو ما يعني أن هناك تجانس أوثبات في تباين الأخطاء.



4|3|2|ت) الطريقة الحسابية (Goldfield– Quandt)

لاختبار ثبات التباين: وتتم وفق الخطوات التالية:

1) ترتيب المشاهدات– **تنازليا**– وفقا للمتغير المستقل
2) استبعاد 20% من المشاهدات في المنتصف، لكل من المتغيرين(X) و (Y) فنحصل على سلسلتين؛
3) حساب مجموع مربعات $(\mathrm{SSE})_{(1)}$ و $(\mathrm{SSE})_{(2)}$ للسلسلتين (1) و(2)، من خلال جدول تحليل التباين لمعادلة الانحدار لكل سلسلة.
4) تنفيذ خطوات إيجاد نموذجي انحدار (Y)على$(X'S)$ للسلسلتين (1) و(2). ومن مخرجات SPSS نهتم بجدول تحليل التباين للحصول على كل من $(\mathrm{SSE})_{(1)}$ و $(\mathrm{SSE})_{(2)}$.
5) حساب النسبة التالية: $F_c=\frac{(\mathrm{SSE})_{(2)}}{(\mathrm{SSE})_{(1)}}$
6) المقارنة بين F_c المحسوبة و $F(df_1,df_2,\alpha)$ المجدولة، لرفض أو قبول الفرضية H_0.

حيث: يمكن حساب درجتي الحرية كما يلي: $\boldsymbol{df_1=\mathrm{df}_2=\frac{n-2-m}{2}}$ ، m : عدد المشاهدات المستبعدة.

4|3|3 النمذجة في ظل اختلال فرضية التجانس

للتخلص من مشكل عدم التجانس، نقوم بقسمة طرفي النموذج على $\hat{\rho}_\varepsilon$ فنحصل على النموذج المصحح الذي ينبغي تقديره بطريقة (OLS)، والمتمثل فيما يلي: $\frac{\mathrm{Y_t}}{\hat{\sigma}_\varepsilon}=\frac{\alpha}{\hat{\sigma}_\varepsilon}+\frac{\beta}{\hat{\sigma}_\varepsilon}X_t+\frac{1}{\hat{\sigma}_\varepsilon}\varepsilon_t$

والهدف من عملية القسمة على " $\hat{\sigma}_\varepsilon$ " هو إقصاء المتغير المستقل المتسبب في عدم التجانس.

وبصفة عامة: يمكن قسمة طرفي النموذج على مقادير أخرى، (محمد شيخي،2011، ص ص 117–118).

4|4 التعدد الخطي بين المتغيرات المستقلة Multicollinearity

بافتراض أن النموذج هو: $$Y_i=\beta_0+\sum_{j=1}^{k}\beta_j X_{ij}+\varepsilon_i : i=\overline{1,n}$$

فإنه في أغلب الأحيان نواجه مشكلة التعدد الخطي بين المتغيرات المستقلة، بمعنى: $\exists i\neq j: Cov(X_i,\ X_j)\neq 0$، مما يؤدي إلى عدم استقرار تباين المقدرات ويكون هناك فائض في التقدير، وهو ما يجعل المصفوفة $X'X$ غير قابلة للقلب.

1|4|4 أسباب وجود تعدد خطي بين المتغيرات المستقلة

- تقل قيمة R، والسبب أن المتغيرات المستقلة تتشارك في نفس تباين المتغير التابع؛
- صعوبة تحديد الأهمية النسبية لكل متغير من المتغيرات المستقلة بالنسبة للمتغير التابع؛
- زيادة تباين المعاملات في الانحدار المتعدد، فكلما زاد تباين المعاملات، قل ثبات معادلة التنبؤ.

2|4|4 الكشف عن وجود التعدد الخطي بين المتغيرات المستقلة

لتشخيص مشكلة الارتباطات المتداخلة بين المتغيرات المستقلة بشكل دقيق، يمكن استخدام برنامج SPSS بأكثر من طريقة (أو مقياس) منها:

2|4|4أ) مصفوفة الارتباط$(\boldsymbol{Correlation\ Matrix})$: فحص مصفوفة الارتباط بين المتغيرات المستقلة، بحيث يمكن الحكم بعدم وجود تعدد خطي في الحالة التي تتراوح قيم معاملات الارتباط بين $(\boldsymbol{-0.7})$ و $(\boldsymbol{+0.7})$.

2|4|4ب) مؤشر الحالة (Condition Index): ويستخرج مؤشر الحالة من قيم الجذور الكامنة (القيم الذاتية) المدرجة في العمود الثاني من الجدول " Collinearity Diagnostic " وذلك وفق العلاقة الرياضية: $\boldsymbol{C} = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$

حيث: $\boldsymbol{\lambda_{max}}$ الجذر الكامن الأكبر، $\boldsymbol{\lambda_{min}}$ الجذر الكامن الأصغر

- يرى جونسون أنه إذا كان: $\boldsymbol{C > 20}$، فهذا دليل على وجود تعدد خطي مرتفع؛
- ويرى بيلسلي وزملاؤه أنه إذا كان: $\boldsymbol{C > 30}$ فهو دلالة على وجود تعدد خطي مرتفع جدا بين المتغيرات المستقلة.

2|4|4ت) معامل التحميل$(\boldsymbol{Tolerance})$: يعد هذا المعامل من المؤشرات التي تكشف عن مدى وجود تعدد خطي بين المتغيرات المستقلة، ويمكن حسابه من خلال الصيغة الرياضية: $(\boldsymbol{Tolerance}) = \boldsymbol{1 - R^2}$

حيث: $(\boldsymbol{R^2})$ هو مربع معامل الارتباط بين كل متغيرين مستقلين.

فإذا كانت قيمة معامل التحمل أصغر من $(\boldsymbol{0.20})$ فإنها تشير إلى احتدام مشكلة التعدد الخطي بين المتغيرات المستقلة. ووفقا للمعادلة أعلاه فإنه كلما ارتفعت قيمة $(\boldsymbol{R^2})$، اقتربت قيمة معامل التحمل من الصفر.

2|4|4ث) معامل تضخم التباين $(\boldsymbol{Variance\ Inflation\ Factor})$: يعد هذا المعامل من المؤشرات التي تكشف عن مدى وجود تعدد خطي بين المتغيرات المستقلة. ففي حالة وجود تعدد خطي تام، يأخذ $\boldsymbol{VIF}$ قيمة لانهائية. وفي حالة عدم وجود تعدد خطي فإن $\boldsymbol{VIF = 1}$.

أما إذا تجاوزت قيمة $\boldsymbol{VIF}$ الأربعة (4) دل ذلك على احتدام مشكلة التعدد الخطي، فكلما ارتفعت قيمته تسبب في عدم استقرار قيمة المعامل (**بيتا**) من جراء ارتفاع الخطأ المعياري. ويبدي بعض المختصين تسامحا في هذا الأمر حيث يرون أن نقطة القطع (المحك) هي خمسة (5).

ويمكن حساب معامل تضخم التباين من خلال الصيغة الرياضية: $(\boldsymbol{VIF}) = \frac{\boldsymbol{1}}{\boldsymbol{Tolerance}}$

3|4|4 النمذجة في ظل وجود تعدد خطي بين المتغيرات المستقلة

من الإجراءات التي يمكن اتخاذها في هذا الإطار ما يلي:

- دمج المتغيرات المستقلة المرتبطة مع بعضها ارتباطا قويا؛
- استخدام التحليل العاملي في حال وجود عدد كبير من المتغيرات للتخلص من بعضها؛
- استخدام طريقة الاختبار $(\boldsymbol{Stepwise})$ التي يتيحها برنامج SPSS؛
- استخدام القيم المعيارية للمتغيرات المستقلة بدلا من قيمها العادية