مجلة البحوث المالية والتجارية

المجلد (24) – العدد الأول – يناير 2023

المقارنة بين أنواع نماذج البيانات الطولية

(دراسة تطبيقية)

Comparing The types of Panel data Models
(An Applied Study)

الباحث/ رامي محمد طايع محمد

مدرس مساعد الاحصاء بالمعهد العالي للعلوم الادارية و التكنولوجيا بالمنزلة

مرشح للدكتوراه

كلية التجارة ـ جامعة بورسعيد ـ قسم الاحصاء والرياضيات والتأمين

إشــراف

الأستاذ الدكتور

محمد المهدي محمد علي

أستاذ الرياضيات و الاحصاء الاكتواري

كلية التجارة ـ جامعة بورسعيد

الدكتورة

سمر أحمد حلمي

مدرس الاحصاء التطبيقي

كلية التجارة ـ جامعة بورسعيد

تاريخ الإرسال	2023-01-23
تاريخ القبول	2023-03-01
رابط المجلة:	https://jsst.journals.ekb.eg/

الملخص :

اكتسبت البيانات المتعلقة بالظواهر الاقتصادية القياسية أهمية كبيرة في الفترة الأخيرة و ذلك لأهمية هذا النوع من الدراسات و تأثيرها على الواقع الاقتصادي للدول و حجم التبادل التجاري المتعلق بتحليل عدة متغيرات و العلاقة بينها في ظل توافر البيانات عن فترات زمنية لتلك الدول تشكل بيانات طولية , و يستهدف البحث العرض و التعريف بالبيانات الطولية و الفرق بينها و بين كل من البيانات المقطعية و بيانات السلاسل الزمنية , و كذلك يعرض البحث المميزات و العيوب المتعلقة باستخدام البيانات الطولية و أفضلية استخدامها عن البيانات المقطعية و بيانات السلاسل الزمنية و المميزات و الصعوبات الخاصة باستخدام البيانات الطولية في الدراسات الاقتصادية القياسية , و كذلك عرض النماذج و الطرق الاحصائية المستخدمة في تقدير معلمات العلاقات في حالة استخدام البيانات الطولية , و المقارنة بين الطرق و النماذج المستخدمة في البيانات الطولية من حيث الافتراضات التي قدمتها هذه النماذج ، ونقاط القوة والضعف المرتبطة بها ، مما يساعد على فهم أفضل طريقة للتطبيق على كل دراسة على حدة بعد ثورة المعلومات وتوافر البيانات بين الباحثين , و أيضا من حيث أفضلية استخدام كل نموذج حسب افتراضات معينة متعلقة بالنموذج و البيانات المستخدمة كحجم العينة و طول السلسلة الزمنية , ثم يقدم البحث دراسة تطبيقية على بيانات حقيقية لإحدى الظواهر الاقتصادية القياسية و الخاصة بتغير قيمة الواردات بالدولار الأمريكي في مصر وبعض الدول الأفريقية قريبة الشبه في الظروف الاقتصادية و ذلك بناءا على تغير سعر صرف الدولار بالعملة المحلة لكل دولة و كذلك الأسعار القياسية للسلع الاستهلاكية لكل دولة و تحديد أفضلية استخدام النماذج المختلفة حسب طبيعة بيانات الدراسة ثم تطبيق النموذج الأكثر ملائمة لبيانات الدراسة و استخراج النتائج و تحليلها و التعليق عليها .

الكلمات المفتاحية : البيانات الطولية – التأثيرات الفردية – نماذج التأثيرات الثابتة – نماذج التأثيرات العشوائية – التباين غير المشاهد .

Abstract :

The Economic data have gained great importance in the recent period due to the importance of this type of studies and their impact on the economic reality of countries and the volume of trade exchange related to the analysis of several variables and the relationship between them in light of the availability of data for time periods for those countries that constitute Panel data. The research presents and defines Panel data and the difference between them and each of the cross-sectional data and the time series data. The research also presents the advantages and disadvantages related to the use of Panel data and the preference for using them over cross-sectional data and time series data, and the advantages and difficulties of using Panel data in Econometric studies, as well as presenting the models and statistical methods used in estimating relationship parameters in the case of using Panel data, and comparing the methods and models used in Panel data in terms of the assumptions made by these models, and the strengths and weaknesses associated with them, which helps to understand The best way to apply to each study separately after the information revolution and the availability of data among researchers, and also in terms of The preference of using each model according to certain assumptions related to the model and the data used, such as the sample size and the length of the time series, then the research presents an applied study on real data for one of the econometric phenomena related to the change in the value of imports in US dollars in Egypt and some African countries that are similar in economic conditions. Based on the change in the exchange rate of the dollar in the local currency of each country, as well as the standard prices of consumer goods for each country, and determining the preference for using different models according to the nature of the study data, then applying the most appropriate model for the study data, extracting the results, analyzing them, and commenting on them.

Keywords : Panel data - Longitudinal data - Individual effects - Fixed effects models - Random effects models - Unobserved variance .

مقدمة :

مع تزايد الاهتمام بعلوم البيانات وأنواعها على مدار العقود القليلة الماضية ، أصبحت البيانات الطولية من أهم أنواع البيانات في البحث التطبيقي لما لها من مزايا عديدة من حيث التحليل والتنبؤ ودقة النتائج وكفاءتها ، وهذا يتضح بشكل خاص في العديد من دراسات الاقتصاد القياسي.

و يهدف هذا البحث إلى دراسة البيانات الطولية والتعرف على كيفية اختلافها عن بيانات المقطع العرضي وبيانات السلاسل الزمنية ، وتوضيح النماذج المستخدمة في حالة البيانات الطولية من حيث الافتراضات التي قدمتها هذه النماذج ، ونقاط القوة والضعف المرتبطة بها ، مما يساعد على فهم أفضل طريقة للتطبيق على كل دراسة على حدة بعد ثورة المعلومات وتوافر البيانات بين الباحثين من أجل فهم أعمق للتحليل والتنبؤ وتقدير أفضل للظواهر الاقتصادية القياسية و التي هي موضوع أحدث الأبحاث في جميع المجالات.

حيث يتناول هذا البحث التعريف بالبيانات الطولية , و مميزات استخدامها عن غيرها من البيانات , و النماذج المستخدمة في حالة البيانات الطولية مثل نموذج التأثيرات الثابتة و نموذج التأثيرات العشوائية و المقارنة بين النموذجين .

هدف البحث :

يهدف البحث إلى إلقاء الضوء على النماذج المستخدمة في حالة البيانات الطولية و المميزات و العيوب المتعلقة بكل نموذج و الفروق الجوهرية بين تلك النماذج و حالات استخدام كل نموذج من تلك النماذج .

أهمية البحث :

إن استخدام النموذج المناسب لطبيعة بيانات الدراسة يؤدي إلى الحصول على مقدرات تتسم بالاتساق و عدم التحيز و انخفاض أخطاء التقدير المتعلقة بالمقدرات مما يؤدي إلى زيادة دقة التحليل و التنبؤ , و خاصة في حالة النماذج الاقتصادية القياسية و التي لها أهمية كبيرة في دراسة المؤشرات و العوامل الاقتصادية المؤثرة في دراسة و تحليل و علاج المشكلات الاقتصادية بشكل عام .

ما هي البيانات الطولية Panel Data ؟ :

من أهم تعريفات الاقتصاد القياسي أنه " تحليل كمي للظواهر الاقتصادية الفعلية ، هذا التحليل مبني علي النمو المتزامن للنظرية والمشاهدة المرتبطة بالطرق الملائمة للاستدلال " (Samuelson , 1954) .

ويتضح من التعريف السابق أن ملائمة و توافر البيانات اللازمة لتحليل الظواهر الاقتصادية هو من أهم العوامل التي يتوقف عليها نجاح التحليل الاقتصادي ، وبالتالي يعتمد نجاح التحليل علي نوعية البيانات المستخدمة في التحليل والقياس وطريقة الحصول عليها و طبيعة تلك البيانات .

يمكن تقسيم البيانات الي ثلاث انواع هي :-

1- بيانات السلاسل الزمنية Time Series Data

السلسلة الزمنية هي مجموعة من الملاحظات أو المشاهدات حول متغير معين خلال فترة زمنية متواصلة (يومية – أسبوعية – شهرية – موسمية – سنوية) وتستخدم على نطاق واسع في البحث الاقتصادي القياسي ، ولكن إحدى المشكلات التي يواجهها الاقتصاديون في التعامل مع السلاسل الزمنية أن معظم الأعمال التكرارية تستند إلى افتراض أن السلاسل الزمنية ساكنة ، مما يعني أن السلوك الإحصائي للمتغير قيد الدراسة لا يتغير بمرور الوقت.

2-بيانات القطاع العرضي Cross Section Data

تعرف بيانات القطاع العرضي بأنها بيانات لمتغير أو أكثر تم جمعها في نفس الفترة الزمنية لمفردات عديدة مختلفة و هي التي تمثل عينة الدراسة , وتعاني البيانات في القطاع العرضي من الكثير من المشكلات مثل السلاسل الزمنية وبصفة خاصة مشكلة التغاير Heterogeneity .

3-البيانات الطولية Panel Data

وهي جمع أو مزيج من بيانات السلاسل الزمنية وبيانات القطاع العرضي ، و فيها تكون وحدة البيانات عبارة عن عينة قطاع عرضي (أسرة – مؤسسة – دولة – فرد) يتم جمع بياناتها خلال فترة زمنية ما , و في هذه الحالة يصبح للبيانات الطولية بعد زمني و بعد مكاني أيضا .
و في حالة جمع بيانات عن متغيرات الدراسة لعدة وحدات , و كل وحده يتم جمع بياناتها خلال فترة زمنية محددة , ثم يتم تجميع تلك البيانات في تجميعه واحدة يطلق عليها بيانات طولية أو تجميعية او جدولية Panel Data , فهي تعتبر خليط من بيانات السلاسل الزمنية والبيانات

المقطعية و سوف يتم استخدام مصطلح البيانات الطولية للتعبير عن تلك البيانات . بينما تعرف نماذج الانحدار التي تعتمد علي تلك البيانات بنماذج انحدار البيانات الطولية أو الجدولية .

مميزات استخدام البيانات الطولية عن غيرها من البيانات :

الفائدة الرئيسية من استخدام البيانات الطولية تتضح من خلال ارتفاع دقة التنبؤ من خلال كثرة عدد المشاهدات و الذي يتحقق من خلال ربط عدد المشاهدات المقطعية بعدد من الفترات الزمنية ، حيث تعرف البيانات الطولية على أنها مشاهدات مقطعية مقاسة في فترات زمنية معينة (الجمال , 2012) .

و تتحقق الفائدة السابق ذكرها من خلال المميزات الخاصة بالبيانات الطولية و التي تميزها عن بيانات السلاسل الزمنية أو البيانات المقطعية و تتمثل هذه المميزات فيما يلي : (D.Gujarati , 2009)

1- الأساليب المستخدمة في تقدير معالم البيانات الطولية تأخذ في الاعتبار مسألة عدم التجانس بين المفردات أو الوحدات الجزئية عن طريق السماح بوجود متغيرات مجددة للمفردات .
2- تعطي البيانات الطولية معلومات أكثر عن البيانات بتباين أكثر و ارتباط داخلي (تعددية خطية) بين المتغيرات أقل و درجات حرية أكثر و كفاءة أكبر .
3- البيانات الطولية مناسبة أكثر لدراسة حركية التغير (ديناميكية التغير) خاصة في البيانات الخاصة بمعدلات البطالة و التحول الوظيفي و تحرك قوى العمل و بيانات الاستهلاك و الادخار و هكذا من الظواهر الاقتصادية .
4- البيانات الطولية من الممكن أن تتنبأ و تقيس التأثيرات التي لا تستطيع ببساطة مشاهدتها من خلال بيانات مقطعية فقط أو من خلال بيانات السلاسل الزمنية فقط .
5- البيانات الطولية تجعل من الممكن دراسة النماذج السلوكية الأكثر تعقيدا .
6- كثرة البيانات تقلل من التحيز الذي قد يتواجد في النتائج إذا قمنا بتجميع المفردات في تجميعه واحدة .

مما سبق يتضح زيادة جودة التحليل الاختباري بطريقة قد لا تكون ممكنة عند استخدام بيانات مقطعية فقط أو سلاسل زمنية فقط , إلا ان هناك بعض المشكلات التي تواجه البيانات الطولية , على سبيل المثال أنها قد تجعل التحليل أكثر تعقيدا ، كما انها قد لا تناسب جميع الحالات ولابد للباحث ان يستخدم أحكاما عملية تتغير حسب الحالة التطبيقية محل الدراسة .

النماذج المستخدمة في حالة البيانات الطولية Panel Data Model :

تتحدد النماذج المستخدمة في حالة البيانات الطولية بناء علي طبيعة البيانات محل الدراسة ويستخدم عادة ثلاثة أشكال رئيسة لتلك النماذج و هي :

1- نموذج الانحدار التجميعي Pooled Regression Model :

حيث يتم مزج أو تجميع المفردات كلها في تجميعة واحدة ومن ثم استخدام طرق التقدير المعتادة لتقدير معالم نموذج الانحدار , مثل طريقة المربعات الصغرى العادية OLS , إلا ان هذه الطريقة تغفل التأثيرات الفردية المحددة غير المشاهدة المختلفة للمفردات المقطعية , تلك الطريقة هي الأبسط ولكنها تعتبر غير واقعية حيث لا تأخذ في الاعتبار اختلاف المشاهدات أو اختلاف الزمن للبيانات المجمعة ويكون النموذج علي صورة .

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + u \qquad (1)$$

حيث يفترض فيها أن جميع المعاملات ثابتة بالنسبة للزمن وكذلك بالنسبة للمشاهدات .

تلك الطريقة قد يشوبها خطأ في التوصيف حيث يفترض أن الجزء المقطوع من المحور الصادي متساوي لجميع المفردات المقطعية بالعينة كما يفترض أن معاملات الميل لكل متغير تفسيري متساوية تماما لكل المفردات المقطعية وهذا الفرض عادة ما يكون غير واقعي وشديد التعقيد ، ويتم الاستدلال علي وجود ذلك الخطأ في التوصيف من خلال قيمة احصاء Durbin-watson المنخفضة عادة مع تقدير ذلك النموذج بهذه الطريقة , بالإضافة إلي أن التقدير بهذا الأسلوب يفتقد لإمكانية دراسة ديناميكية التغير في النماذج السلوكية .

2- نموذج التأثيرات الثابتة Fixed Effects Model :

أو ما يسمي بنموذج انحدار المربعات الصغرى باستخدام المتغيرات الوهمية " LSDV " , حيث يفترض ثبات معاملات الميل ولكن الجزء المقطوع من المحور الصادي يختلف باختلاف المشاهدات المقطعية , حيث أن اختلاف الجزء المقطوع في المحور الصادي باختلاف المفردات يجعلنا نضع في الاعتبار الاختلافات الفردية لكل وحدة في البيانات المقطعية .

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots\ldots + u_{it} \qquad (2)$$

حيث الترميز (i) للجزء المقطوع من المحور الصادي يشير للتعبير عن الاختلاف بين المفردات المقطعية التي قد ترجع إلى صفات خاصة بكل مفرده تختلف عن صفات المفردات الأخرى .

هذا النموذج السابق يعرف بنموذج التأثيرات الثابتة FEM ويعود مصطلح التأثيرات الثابتة الي أنه علي الرغم من اختلاف الجزء المقطوع من المحور الصادي بين المشاهدات إلا أنه لا يختلف باختلاف الزمن , وبالتالي فهو ثابت زمنيا .

وللسماح للجزء المقطوع من المحور الصادي بالاختلاف بين المشاهدات المقطعية يستخدم لذلك الاسلوب المتغيرات الوهمية كالتالي :

$$y_{it} = \alpha_1 + \alpha_2 D_2 + \alpha_3 D_3 + ... + \beta_1 x_{1it} + \beta_2 x_{2it} + ... + u_{it} \qquad (3)$$

إن اتساق تقديرات نماذج التأثيرات الثابتة لا يتطلب افتراضات حول التأثيرات الفردية المحددة ولكن يتطلب افتراض توزيع اضافي لها " افتراض السكون علي سبيل المثال " (Honore ,1998) .

لاحظ أنه من الممكن افتراض أن معاملات الميل ثابته بينما الجزء المقطوع من المحور الصادي يختلف باختلاف الزمن أو باختلاف المشاهدات والزمن , وكذلك يمكن افتراض أن جميع المعاملات تختلف باختلاف المشاهدات .

إن كل افتراض علي حده مما سبق يمكن التعبير عنه باستخدام المتغيرات الوهمية , إلا أن تلك الافتراضات تجعل عدد المتغيرات الوهمية يزداد بشكل كبير مما يجعل هناك مشكلة في درجات الحرية , كما أن وجود متغيرات عديده في النموذج يجعل هناك دائما احتمال ظهور مشكلة الارتباط الخطي المتعدد والذي يحدث صعوبة في تقدير واحدا أو أكثر من معلمات النموذج .

هناك مشكلة أخري في نماذج التأثيرات الثابتة عند وجود متغيرات ثابتة علي الزمن مثل النوع أو العرق , فإن تلك الطريقة قد تكون غير قادرة علي توضيح تأثير مثل هذه العوامل غير المتغيرة مع الزمن .

من الممكن افتراض ثبات تباين الخطأ بالنسبة للمفردات المقطعية أو افتراض عدم وجود ارتباط ذاتي مع الزمن أو انه من النوع (AR1) مثلا , من الممكن افتراض واحدا أو أكثر من

الافتراضات السابق ذكرها في مقدار الخطأ U_{it} ولكن ذلك سيجعل التحليل أكثر تعقيدا في هذه الطريقة .

لذلك يعتبر نموذج التأثيرات الثابتة نموذج غير محدد بالكامل وقد يكون اللجوء للمتغيرات الوهمية هو نتيجة الفشل في إيجاد متغيرات مفسرة مناسبة لا تتغير مع الزمن , أو تتغير مع الزمن ولكن لا تتغير مع الوحدات المقطعية , ولكن علي أي حال يفضل استخدام تلك النماذج عندما يرتبط عنصر الخطأ بواحد أو أكثر من المتغيرات التفسيرية وهو ما يحدث عادة في النماذج الاقتصادية القياسية خاصة السلوكية منها , حيث المقدرات الناتجة تكون غير متحيزة .

كما أنه يستخدم عندما تكون الوحدات المقطعية تمثل مفردات المجتمع ككل وليست جزء منه , وبالتالي يكون الاستدلال الاحصائي مشروط علي الوحدات الموجودة في العينة فقط , كما تستخدم في حالة قلة عدد الوحدات المقطعية وزيادة السلسلة الزمنية لأنها الطريقة الاسهل عن غيرها .

3- نموذج التأثيرات العشوائية Random Effects Model :-

اتضح فيما سبق أن طريقة التأثيرات الثابتة LSDV يتبعها خسارة كبيرة في درجات الحرية بسبب اللجوء إلي المتغيرات الوهمية , والذي يعكس الفشل في ايجاد متغيرات مفسره مناسبة للنموذج لا تتغير مع الزمن (Kmenta , 1986) .

فإذا كان اللجوء للمتغيرات الوهمية ما هو إلا تمثيل لعدم المعرفة بالنموذج الحقيقي ، فقد يكون التعبير عن ذلك من خلال مقدار الخطأ العشوائي Uit هو الأسهل والأنسب , و يتم ذلك في الاقتراح المسمى بنموذج عناصر الخطأ (ECM) أو نموذج التأثيرات العشوائية (REM) ، ويتضح ذلك من خلال المعادلة التالية :

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots\ldots + u_{it} \qquad (4)$$

ويتم التعامل مع المقدار α_i علي أنه متغير عشوائي له توقع ثابت وهو α حيث القيمة المقطوعة من المحور الصادي لوحده مقطعية ما هي :

$$\alpha_i = \alpha + \varepsilon_i \qquad (5)$$

i= 1,2,3,...,N

حيث ε_i هو مقدار الخطأ العشوائي وله توقع يساوي الصفر وتباين يساوي (σ_ε^2) , حيث تعتبر كل وحدة مقطعية موجودة في العينة مسحوبة من مجتمع أكبر يحتوي علي عدد أكبر من الوحدات المقطعية المماثلة والتي لها جميعا توقع مشترك للجزء المقطوع من المحور الصادي هو α , والفرق بين الوحدات في الجزء المقطوع من المحور الصادي يتم تمثيله في مقدار الخطأ (ε_i) , ويكون النموذج علي الصورة التالية :

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots\ldots + \varepsilon_i + u_{it}$$
$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots\ldots + w_{it}$$
$$w_{it} = \varepsilon_i + u_{it} \qquad (6)$$

ويسمي مقدار الخطأ المركب المكون من جزئين حيث ε_i يمثل عنصر الخطأ الخاص بالبيان المقطعي أو المفردات , والمقدار u_{it} يمثل عنصر الخطأ الناتج من دمج السلاسل الزمنية مع البيانات المقطعية وذلك تحت الفروض التالية :

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$
$$u_{it} = N(0, \sigma_u^2)$$
$$E(\varepsilon_i, u_{it}) = 0$$
$$E(\varepsilon_i, \varepsilon_j) = 0 \ldots\ldots i \neq j$$
$$E(u_{it}, u_{is}) = E(u_{it}, u_{jt}) = E(u_{it}, u_{js}) = 0 \ldots . i \neq j, t \neq s \qquad (7)$$

وهو ما يعبر عن استقلال عناصر الخطأ و لا يوجد ارتباط ذاتي بالنسبة لوحدات البيانات المقطعية او لوحدات السلاسل الزمنية .

ويطلق هنا علي المقدار ε_i المتغير غير المشاهد أو الخفي , حيث أنه انحراف عشوائي للأجزاء المقطوعة من المحور الصادي للمفردات المقطعية عن القيمة المتوقعة , وهو لا يلاحظ مباشرة .

وكنتيجة للفروض السابقة تحصل علي ما يلي :

$$E(w_{it}) = 0$$
$$Var(w_{it}) = \sigma_{\varepsilon}^2 + \sigma_u^2 \qquad (8)$$

لاحظ أنه اذا كانت $\sigma_{\varepsilon}^2 = 0$, فإنه يمكن تجميع كل المشاهدات والقيام بانحدار تجميعي .

كذلك من الممكن اثبات ان W_{is} , W_{it} عندما ($t \neq s$) عند أي نقطتان زمنيتان مرتبطان ويكون معامل الارتباط هو :

$$corr(w_{it}, w_{is}) = \frac{\sigma_{\varepsilon}^2}{\sigma_{\varepsilon}^2 + \sigma_u^2} \qquad (9)$$

هذا المعامل للارتباط بين مقادير الأخطاء يظل كما هو عند أي قيمتين مختلفتين في الزمن , وهذا ما يخالف ما هو معروف في أسلوب السلاسل الزمنية حيث يقل هذا الارتباط بمرور الزمن , في حين أن شكل هذا الارتباط يظل كما هو بالنسبة لجميع الوحدات المقطعية في هذه الطريقة.

الملاحظ أنه عند عدم وضع هذا الارتباط في الاعتبار , والتقدير بطريقة OLS سوف نحصل علي نتائج للمقدرات لا تتسم بالكفاءة , وأفضل طريقة هنا للتقدير هي استخدام أسلوب المربعات الصغرى المعممة (GLS) (Wooldridge ,2003) .

الفرق بين النموذجين (FEM) ، (ECM) :

الأمر يرجع إلى الفروض التي يضعها الباحث عن الارتباط المحتمل بين المفردات المقطعية ومقدار الخطأ ε_i والمتغيرات المفسرة X's .

1- اذا كان هناك افتراض بأن ε_i و X's غير مرتبطين فأن نموذج ECM يكون أفضل ، فمثلا بافتراض أن مكاسب الشخص داله في التعليم وخبرة العمل وما إلى ذلك من عوامل مفسرة اخري , وأن المقدار ε_i يتعلق بالقدرات الفطرية أو الخلفية الأسرية للشخص

, بالتالي يرتبط هذا المقدار مع التعليم , وهذا هو السبب الرئيسي لاستخدام البيانات الطولية حيث السماح لمقدار التأثير غير الملاحظ ε_i بالارتباط مع المتغيرات المفسرة (D.Gujarati , 2009) , ولهذا فإنه عند ارتباط ε_i مع X's فإن من الافضل استخدام (FEM) .

2- الفروض الخاصة بنموذج ECM تعتمد علي أن المقدار ε_i مسحوب عشوائيا من مجتمع أكبر من المفردات المقطعية , ولكن أحيانا لا يكون الأمر كذلك .

3- إذا كانت t كبيره و N صغيره سيكون الفرق بين قيم المقدرات بكلتا الطريقتين بسيط جدا وبالتالي يتم استخدام الاسلوب الأسهل حسابيا و هو FEM .

4- اذا كان N كبيره و t صغيرة ستكون الفروق بين مقدرات النموذجين معنوية , وبالتالي فإن استخدام FEM ومعاملة المقدار α_{it} علي أنه ثابتا وليس عشوائيا يجعل الاستدلال الاحصائي مشروطا علي الوحدات المقطعية الموجودة في العينة فقط ، بينما إذا كانت الوحدات المقطعية مسحوبة عشوائيا من مجتمع أكبر فأن ECM أفضل ويكون الاستدلال الاحصائي غير مشروط .

5- إذا كان هناك ارتباط بين مقدار الخطأ ε_i وأحد المتغيرات المفسرة فإن مقدرات ECM تكون متحيزة علي عكس FEM .

6- إذا كانت N كبيرة و t صغيرة وفروض ECM متحققة فإن مقدرات ECM تكون أكثر كفاءة من مقدرات FEM .

7- هناك اختبار Housman والذي يفترض أن مقدرات الطريقتان غير مختلفتان , و له احصاء اختبار تتبع تقاربيا توزيع χ^2 , و إذا تم رفض هذا الفرض فإن هذا دليل علي أن استخدام ECM غير مناسب , ويكون من الأفضل استخدام FEM وأن يكون الاستدلال الاحصائي في هذه الحالة مشروطا علي ε_i الموجودة في العينة (Taylor , 1980) .

و يجب ملاحظة أنه لا توجد قاعدة بسيطة تساعد في الاختيار بين النموذجين (Johnson , 1997) .

الدراسة التطبيقية :

من خلال دراسة بيانات ستة دول أفريقية تمثل مفردات أو مشاهدات أو عينة من الدول التي تعتمد على الوارد ات لسداد احتياجاتها من السلع الأساسية , تم دراسة نموذج انحدار المتغير الخاص بقيمة واردات تلك الدول بالدولار الأمريكي كمتغير تابع على كل من متغير سعر صرف الدولار أمام عملة الدولة و كذلك متغير الأسعار القياسية للسلع الاستهلاكية في تلك الدول كمتغيرات مفسرة باعتبار سنة 2010 سنة أساس و ذلك في الفترة من 2000 إلى 2019 , بحيث تكون هناك بيانات طولية تمثل ستة دول كمفردات الدراسة هي مصر و تونس و الرأس الأخضر و كينيا و المغرب و موريشيوس , و لكل مفردة سلسلة زمنية من عشرين سنة هي فترة الدراسة , و ذلك باستخدام البيانات السنوية المنشورة على موقع البنك الدولي حيث :

X1 : يمثل سعر صرف الدولار الأمريكي أمام عملة الدولة

X2 : الأسعار القياسية للسلع الاستهلاكية في كل دولة باعتبار 2010 سنة الأساس

X3 : قيمة الواردات من السلع لكل دولة بالدولار الأمريكي

و يكون النموذج على الصورة التالية :

$$x_{3it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + u_{it}$$

1- بتطبيق طريقة المربعات الصغرى العادية OLS على نموذج الدراسة و تجميع البيانات في تجميعه واحدة وعدم الاعتبار لاختلاف المشاهدات كانت النتائج كما يلي :

$$x_{3it} = 8.06 - 0.312x_{1it} + 0.242x_{2it}$$

Se : (3.81) (0.034) (0.032)

R2 =0.549 F=71.32 prop=0.00 RMSE=14.16 D.W= 0.18

و قد اتضح أن كل معاملات الانحدار لها معنوية منفردة و لها الاشارات المتوقعة و قيمة معامل تحديد مرتفعة إلى حد ما , بينما قيمة احصاء ديربن واتسون منخفضة نوعا ما و هو ما قد يشير إلى وجود خطأ في توصيف النموذج , حيث يفترض النموذج أن الجزء المقطوع من المحور الصادي واحدا لكل مفردات الدراسة و هو بالتالي

مقيد بهذا الفرض و هو فرض غير سليم نظرا للتباين و الاختلاف غير المشاهد بين مفردات الدراسة .

2- باستخدام نموذج التأثيرات الثابتة و ذلك باستخدام اسلوب المتغيرات الوهمية (LSDV) و بافتراض أن معاملات الميل ثابتة و لكن الجزء المقطوع من المحور الصادي يختلف باختلاف المشاهدات , يكون النموذج على الصورة التالية :

$$x_{3it} = 27.79 - 0.209x_{1it} + 0.205x_{2it} - 28.8D_1 - 27.4D_2 - 18.7D_3 - 9.08D_4 - 35.4D_5$$

Se : (2.94) (0.111) (0.02) (2.65) (10.34) (9.36) (2.7) (4.09)

R^2 = 0.85 F=91.09 prop=0.000 RMSE= 8.33

و يلاحظ معنوية جميع المقدرات منفردة و زيادة قيمة معامل التحديد للنموذج بسبب ادخال متغيرات جديدة لمراعاة اختلاف الجز المقطوع من المحور الصادي لكل مفردة بسبب الصفات الخاصة بكل دولة .

بزيادة قيمة معامل التحديد و قيمة احصاءه ديربن واتسون , و باستخدام اختبار F المقيد للمقارنة بين النموذجين باعتبار النموذج الأول مقيدا , يتضح أن النموذج الأول غير مناسب حيث قيمة احصاء الاختبار لها معنوية احصائية عالية حيث كانت :

$$F = \frac{(0.85 - 0.549)/3}{(1-0.85)/112} = 74.9$$

3- باستخدام نموذج التأثيرات العشوائية باعتبار أن المفردات تمثل عينة من مجتمع أكبر و لها توقع مشترك للجزء المقطوع من المحور الصادي , تكون المعادلة على الصورة :

$$x_{3it} = 9.25 - 0.254x_{1it} + 0.209x_{2it}$$

Se : (5.99) (0.085) (0.02)

R2 = 0.549 Wald=109.55 prop=0.000 RMSE=8.33

نجد بوجه عام عدم اختلاف قيم معاملات الانحدار بشكل كبير عن طريقة التأثيرات الثابتة , مع وجود معنوية احصائية فردية أفضل لمعاملات الانحدار و أخطاء معيارية أقل , و هو النموذج الأنسب في حالة أن المفردات مسحوبة من مجتمع أكبر , مع الوضع في الاعتبار عدم ارتباط التأثيرات الفردية غير المشاهدة للمفردات مع المتغيرات التفسيرية , بينما يفترض نموذج التأثيرات الثابتة أن القياس يقتصر فقط على مشاهدات المفردات محل الدراسة فقط .

الاستنتاجات :

1- استخدام الانحدار التجميعي بطريقة OLS قد ينتج مقدرات تتسم بالمعنوية الفردية و معامل تحديد مرتفع , و لكن ذلك لا يجعلنا نغفل مشكلة خطأ التوصيف و الانحدار الزائف , و الذي من الممكن اكتشافه بسهولة من خلال قيمة إحصاءه ديربن واتسون أو المقارنة مع نماذج البيانات الطولية و ملاحظة النتائج و اجراء الاختبارات المناسبة .

2- عند افتراض اختلاف تباين الخطأ والارتباط الذاتي فلا يصح استخدام نموذج التأثيرات الثابتة , كذلك من عيوب نموذج التأثيرات الثابتة ارتباط حد الخطأ مع التأثيرات الفردية وكثرة المتغيرات الوهمية ونقص درجات الحرية وتزايد فرصة وجود التعدد الخطي , و إذا كانت التأثيرات الفردية لا ترتبط بالمتغيرات المفسرة فمن الأفضل استخدام النموذج العشوائي .

3- من الأفضل استخدام النموذج العشوائي عندما تكون المفردات مسحوبة من مجتمع أكبر , وحتي لو كانت N صغيرة و t غير صغيرة فإن النموذج العشوائي سوف لا يكون مشروطا علي العينة .

4- النموذج العشوائي يطبق بطريقة (GLS) والتي تعالج ارتباط الخطأ مع البيانات سواء زمنية أو مقطعية , وكذلك مع نفسه عبر الزمن وكذلك مع المتغيرات التفسيرية .

قائمة المراجع

أولا المراجع باللغة العربية :

الجمال ، زكريا يحيي ,"اختيار النموذج في نماذج البيانات الطولية الثابتة و العشوائية" ,المجلة العراقية للعلوم الاحصائية ,عدد 21 ,العراق ,2012.

جوجاراتي ، دامودار, "الاقتصاد القياسي" , تعريب : هند عبد الغفار عوده , دار المريخ للنشر ,المملكة العربية السعودية , 2009 .

ثانياً : المراجع باللغة الاجنبية :

Honore , Bo E. ,"Estimation Of Panel Data Tobit Models With Normal Errors" , Department Of Economics ,NJ 08544-1021 ,Princeton University ,Princeton ,1998.

Johnson, J. and Di Nardo , J. , "Econometric Methods" , 4th Ed , McCraw , Hill, 1997.

Kmenta , J. , "Elements Of Econometrics" , 2nd Ed , Macmillan Company , New York , 1986.

Samuelson ,P. A. and Other , "Report of The Evaluative Committee for Econometrica" , Econometrica , Vol.22 , No.2 , 1954.

Taylor, W .E . , "Small Sample Consideration In Estimating From Panel Data" , Journal of Econometrics , Vol. 13 , 1980.

Wooldridge , Jeffrey M. , "Introductory Econometrics : A Modern Approach" , 2nd edition , South-Western Publishing Co. , Michigan State University , USA , 2003 .