AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

تقدير انموذج انحدار داربن المكاني شبه المعلمي باستعمال بعض الطرائق اللامعلمية

Estimating the Semi Parametric Durbin spatial regression model using some Non-Parametric methods

مريم حازم رزاق[1] أ.م.د وفاء جعفر حسين[2]

كلية الادارة والاقتصاد/جامعة واسط

المستخلص

تناول هذا البحث دراسة وتحليل البيانات المكانية ،حيث ركز على تقدير انموذج انحدار ديربن المكاني شبه المعلمي (SP-SDM) ،اذ تم تقدير معلمة الاعتماد المكاني (ρ) في الانموذج باستعمال طريقة الامكان الاعظم (MLE) أما دالة التمهيد $m(\tilde{X})$ فقد تم تقديرها باستعمال طريقتين لامعلمية، وهي (طريقة الأوراكل Oracle،وطريقة انحدار النواة من نوع الالتفاف ConK) ،كما اجريت تجارب المحاكاة باستعمال برنامج MATLAB اذ تم اخذ (9) أحجام من العينات المختلفة وبالاعتماد على معيار متوسط مربعات الخطأ MSE لمعرفة كفاءة الطرائق المستعملة . *من اهم الاستنتاجات التي توصل اليها* ان القيم التي تعبر عن المسافة الاقليدية تظهر انخفاضا ملحوظا حيث كلما كانت القيم أصغر، كانت الطريقة أفضل وأكثر دقة حيث نلاحظ كلما تغير حجم العينة تتغير هذه القيم، ونلاحظ ايضا أن أداء طريقة الأوراكل (Oracle) يظهر دائما نتائج متفوقة (قيم اقل) مقارنة بطريقة (Conk).

الكلمات الرئيسية: الاعتماد المكاني، مصفوفة الأوزان المكانية، مصفوفة الأوزان المعتمدة على المسافة، نموذج انحدار داربن المكاني شبه المعلمي (Sp-SDM).

Abstract

This research dealt with the study and analysis of spatial data, where it focused on estimating the semi-parametric spatial regression model (SP-SDM), where the spatial dependence parameter (p) in the model was estimated using the maximum likelihood method (MLE), while the bootstrap function (X) was estimated using two non-parametric methods, which are the Oracle method and the convolution kernel regression method (Conk). Simulation experiments were also conducted using MATLAB, where (9) sizes of different samples were taken and based on the mean square error (MSE) criterion to know the efficiency of the methods used. One of the most important conclusions reached is that the values expressing the Euclidean distance show a noticeable decrease, where the smaller the values, the better and more accurate the method is, where we notice that these values

and we also notice that the performance of the ،change whenever the sample size changes .(Oracle method always shows superior results (lower values compared to the Conk method

Keywords:

semi-، distance-dependent weight matrix، spatial weight matrix،spatial dependence parametric spatial Durbin regression model (SP-SDM)

1-المقدمة

تـؤدي الأسـاليب الحديثـة للقيـاس الاقتصـادي المكـاني دورا مهمـا فـي تحليـل العديـد مـن الظـواهر الاقتصـادية والاجتماعيـة والصـحية وغيرهـا ، خاصـة التـي تعتمـد علـى البيانـات المكانيـة ، ويعـد الانحـدار المكـاني احـد هـذه النمـاذج التـي تسـعى لفهـم الترابطـات المكانيـة بـين المشـاهدات المختلفـة للظـواهر المدروسـة .تظهـر اهميـة دراسـة التفـاعلات المكانيـة لأنهـا تمثـل عـاملا اساسـيا يـؤثر علـى دقـة النمـاذج المسـتعملة ،وان مـن اهـم النمـاذج الاحصـائية التـي تمثـل الظـواهر التـي تعتمـد علـى المكـان هـو نمـوذج الانحـدار المكـاني الـذي يقـوم بتحليـل العلاقـة بـين المتغيـر المعتمـد والمتغيـرات التوضـيحية فـي ظـل وجـود الاعتمـاد المكـاني ،حيـث يعتمـد بصـورة كبيـرة علـى البيانـات التـي يحصـل عليهـا الباحـث فـأن هـذه البيانـات ليسـت مسـتقلة بـذاتها وانمـا تعتمـد علـى المكـان الـذي تـم جمـع البيانـات منـه ، يعـد تحليـل تـأثير الموقـع علـى البيانـات محـورا اساسـيا فـي الدراسـات الحاليـة ، خاصـة عنـد التعامـل مـع نمـاذج شبه معلمية التي تتميز بمرونة عالية .

فـي عـام (2005)، قامـت الباحثـة (Daniela،Gumprecht)[8]، بدراسـة تنأولـت فيهـا اهميـة الانحـدار المكـاني وخصـائص البيانـات المكانيـة بالإضـافة إلـى الاسـاليب المسـتعملة فـي الكشـف عـن الاعتمـاد المكـاني والارتبـاط الـذاتي المكـاني فـي بيانـات البحـث والتطـوير المسـتعملة فـي البحـث لمعرفـة مـا اذا كـان الانفـاق المحلـي والاجنبـي علـى البحـث والتطـوير لـه تـأثير علـى اجمـالي انتاجيـة العوامـل فـي بلـد مـا . اذ قامـت الباحثـة بتقـدير نمـوذج الانحـدار الـذاتي المكـاني (SAR) ونمـوذج الخطـأ المكـاني (SEM) باسـتعمال طـريقتين همـا طريقـة المربعـات الصـغرى ذات المـرحلتين (2SLS) وطريقـة الامكـان الاعظـم (MLE) وبالاعتمـاد علـى اختبـار مـوران حيـث بينـت النتـائج ان الانفـاق المحلـي علـى البحـث والتطـوير لـه تـأثير ايجـابي علـى اجمـالي انتاجيـة العوامل

فـي عـام (2011) ، قـدم الباحـث (عطـره ،سـامي غنـي خضـير)[1] ، دراسـة حـول طرائـق بيـز فـي تحليـل نمـوذج القيـاس الاقتصـادي المكـاني (SAR) ونمـوذج ديـربن المكـاني (SDM) ونمـوذج الخطـأ المكـاني(SEM)، حيـث اسـتخدم الباحـث حجـم العينـة (61) مريضـا مـن الاطفـال المصـابين بفقـر الـدم الـذين يرقـدون فـي مستشـفيات جانـب الكـرخ فـي بغـداد وبواقـع خمـس منـاطق جغرافيـة ، واقتـرح حسـاب العناصـر فـي المصـفوفة المعدلـة بنـاء علـى نسـبة طـول الحـدود المشـتركة الفعليـة بـين المواقـع المختلفـة ، ممـا يـؤدي إلـى الحصـول علـى القيمـة التقديريـة الدقيقة لمعلمة الاعتماد المكاني .

2 - مشكلة البحث

فـي بعـض الاحيـان تكـون هنـاك حاجـة للتعامـل مـع علاقـات غيـر خطيـة أو معقـدة بـين المتغيـرات التـي لا يمكـن التعبيـر عنهـا بشـكل دقيـق باسـتعمال النمـاذج الخطيـة المعلميـة وخاصـة عنـدما يكـون النمـوذج المعلمـي غيـر كـاف لتعريـف العلاقـات فـي البيانـات، فعنـد دراسـة الظـواهر المختلفـة نلاحـظ عـدم الاخـذ بنظـر الاعتبـار التـأثير المكـاني واهمالـه ممـا يـؤدي إلـى فقـدان معلومـات مهمـة لهـا الاثـر الكبيـر علـى دقـة التقـدير لمعلمـات الانمـوذج والحصـول على نتائج غير دقيقة؛ وذلك لوجود الاعتماد المكاني بين مفردات الظاهرة.

3-هدف البحث

يهدف البحث إلى تقدير العلاقة بين المتغيرات دون افتراض شكل معين لهذه العلاقة مقدما وهذا يسمح للنموذج (نموذج انحدار ديربن الشبه معلمي Durbin Semi-Parametric Spatial Regression Model) التعرف على الانماط والاتجاهات في البيانات (التي تعاني من الاعتماد المكاني) بشكل أكثر مرونة ودقة وخاصة إذا كانت العلاقة بين المتغيرات معقدة وغير خطية ،وكذلك معالجة مشكلة العلاقات اللاخطية بين المتغير المستقل والمتغير التابع التي قد تختلف حسب الموقع الجغرافي .

4-الاعتماد المكاني: Spatial Dependent

عُرِفَ الاعتماد المكاني من لدن (Anselin) على أنه " وجود علاقةٍ دالية بين ما يحدث في نقطة معينة في الفضاء وبين ما يحدث في مكان اخر " [3] ففي التحليل الاحصائي يعني ان القيم المرتبطة بمواقع جغرافية قريبة ليست مستقلة عن بعضها عندها يجب اخذ هذا الاعتماد في الاعتبار عند بناء النماذج. إن الاعتماد المكاني في مجموعة من بيانات العينة يعني ان المشاهدة في الموقع (i) تعتمد على المشاهدة في الموقع (j) حيث ان($i \neq j$) وتكتب بالصيغة الآتية[3]:

$$y_i = f(y_j) \quad \dots(2-1)، i = 1، 2، \dots.، n \quad i \neq j \qquad (1)$$

ويلاحظ من الصيغة اعلاه ان يكون الاعتماد بين عدة مشاهدات اذ ان الموقع(i) يمكن ان يتخذ اي قيمة من قيم (n)وكذلك نلحظ أن بيانات العينة المشاهدة في موقع ما في الفضاء تعتمد على القيم التي شوهدت في مواقع اخرى [4].

5- مصفوفة الأوزان المكانية Spatial Weight Matrix

هي وسيلة رياضية تستعمل في تحليل البيانات والاحصاءات المكانية والانحدار المكاني حيث يتم من خلالها ادخال العلاقات المكانية بين المشاهدات ضمن مساحة جغرافية معينة في الانموذج المكاني وتدعى ايضا مصفوفة الارتباط المكاني ،وتحدد تأثير كل موقع على المواقع الاخرى المجأورة له، ومن اهم صفاتها عناصر قطرها الرئيسي اصفار، وموجبة (لا تحتوي قيم سالبة) [8]،مربعة ابعادها($n \times n$) وان n تمثل عدد المشاهدات المستعملة في الأنموذج المكاني ويرمز لمصفوفة التجأور المكاني بالرمزW ، حيث إن كل صف وعمود فيها يمثل وحدة جغرافية والقيم فيها تعكس العلاقة المكانية بين الوحدات .ويمكن كتابة الصيغة العامة لمصفوفة الأوزان المكانية كما يأتي[10] :

$$W = \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{nn} \end{bmatrix}$$

6 - الأوزان المعتمدة على المسافة :Weight Based on Distance

تُعدُّ الأوزان التي تعتمد على المسافة أحد الاساليب المهمة لتقييم العلاقات بين نقطتين أو خليتين في الفضاء الجغرافي أو الهندسي .حيث تعبر هذه الأوزان عن التأثير المتبادل بين الخليتين بناءً على المسافة بينهما حيث يرمز للمسافة عادة بالرمز d_{ij} ويتم حسابها باستعمال المعادلات الجغرافية المناسبة ومن اهمها المسافة الاقليدية (Euclidean Distance) .وتمثل اقصر مسافة مستقيمة بين نقطتين في الفضاء، ويمكن حساب المسافة الاقليدية بين نقطتين حسب الصيغة الآتية[2] :

$$d_{ij} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

ويتم بناء مصفوفة الأوزان المكانية هنا بناءً على حساب المسافة العكسية *Inverse Distance* حيث تستعمل هذه الطريقة كثيراً في تحليل البيانات المكانية *Spatial Analysis* ، ويعتمد الوزن أو التأثير بين نقطتين على المسافة بينهما بشكل عكسي ،اي بمعنى انه كلما ازدادت المسافة يقل الوزن أو التأثير والعكس صحيح والصيغة العامة لحساب الوزن باستعمال المسافة العكسية هي :

$$W_{ij} = \frac{1}{{d_{ij}}^{\alpha}}$$

حيث ان d_{ij} تمثل المسافة بين وحدتين، و α تمثل معامل القوة.

7-انموذج انحدار ديربن المكاني: Durbin Spatial regression model

يعد انموذج انحدار ديربن المكاني (SDM) احد نماذج الانحدار الذاتي المكاني (SAR) ويعتبر نوعا متقدما من نماذج الانحدار المكاني الذي يأخذ بالحسبان التأثيرات المكانية لكل من المتغير التابع والمتغيرات المستقلة .في نموذج الانحدار الذاتي المكاني (SAR) العلاقة المكانية تكون مقتصرة على تأثير المتغير التابع بين المواقع أما في نموذج انحدار داربن المكاني(SDM) فان العلاقة المكانية تكون اشمل حيث تمتد لتشمل تأثير المتغيرات المستقلة في المواقع المجأورة[1] ،وهذا يؤدي إلى اضافة مستوى اعمق من التفاعل المكاني .تكون مصفوفة المتغيرات المستقلة في انموذج (SAR) هي X أما في انموذج (SDM) فتكون [X WX] حيث يلحظ أن هناك متغيرا تفسيريا متأخر مكانيا، زيادة عن ذلك، متغير معتمد متأخر مكانيا في النموذج .وان الصيغة الرياضية لنموذج انحدار ديربن المكاني المعلمي تكون كالآتي[5] :

$$\underline{y} = \rho W\underline{y} + X\underline{\beta_1} + X\underline{\beta_2} + \underline{\varepsilon} \qquad (2)$$

$$\underline{\varepsilon} \sim N(0, \sigma^2 In)$$

حيث ان ρ تمثل معلمة الاعتماد المكاني للمتغير المعتمد، Wy يمثل متجه المتغير المعتمد المتخلف مكانيا، أما $\underline{\beta_1}$ تمثل موجه المعالم. ولو افترضنا

$\theta = [\beta_1' \ \beta_2' \ ]'$ ، $\tilde{X} = [X \ WX]$ فأن المعادلة (2) تصبح بالشكل الآتي [6]:

$$\underline{Y} = \rho W\underline{y} + \tilde{X}\theta + \underline{\varepsilon} \qquad (3)$$

إن نموذج ديربن المكاني شبه المعلمي(Sp-SDM) هو توسيع إلى أنموذج داربن المكاني المعلمي (SDM) بأشكال دالية مرنة للمتغيرات الخارجية حيث نستبدل الصيغة الخطية θ بالدالة m(X̃) حيث يصبح النموذج بالشكل الآتي[6]:

$$\underline{Y} = \rho W\underline{y} + m(\tilde{X}) + \underline{\varepsilon} \qquad (4)$$

8- طريقة أوراكل نوع عدم المساواة (Oracle-type Inequality)

وتعد هذه الطريقة احدى الطرائق المهمة لتحليل البيانات التي تتضمن تفاعلات مكانية ،اذ تقوم بتحديد المتغيرات التي تسهم بشكل كبير في التنبؤ سواء كانت معلمية، أو شبه معلمية ويتم هذا عبر خطوات انتقائية تعتمد على تحسين التقديرات وتقليل الاخطاء، ويعد الباحثان (Donoho and Johnstone) أول من قدموا مفهوم (Oracle-type Inequality) في عام (1994) .يتم في هذه الطريقة تحديد فضاء الاسقاط $(V_{a_1}، _\infty)$ يتم باختيار اساسا مناسبا (basis function) مثل $\varphi_k، \varphi_j$(دوال فورير أو دوال موجيه)،اي انه مجموعة دوال يتم انشاؤها باستعمال دوال اساس وحاصل الضرب التنسوري بينها حيث يتم استعمال هذه الفضاءات لتقريب الدالة $m(\tilde{X})$ وتحقيق التوازن بين الانحياز والتباين مع التحكم في تعقيد النموذج ،ويعد فضاء الاسقاط $(V_{a_1}، _\infty)$ فضاء متجهات(*vector space*) يتم انشاؤه باستعمال مجموعة من الدوال التي يتم اختيارها مسبقا والهدف منه هو بناء تمثيل تقريبي للدالة $m(\tilde{X})$ الذي نريد تقديرها ضمن هذا الفضاء:[7]

$$V_{a_1}، _\infty = Span\{\hat{\varphi}_k \otimes \hat{\varphi}_j \text{ ،} 1 \le j \le a_1 \text{،} a_2 \ge 1\} \qquad (5)$$

حيث ان :

a_1 : *تمثل عدد الدوال الاساسية في الاتجاه الأول*، a_2 تمثل عدد الدوال الاساسية في الاتجاه الثاني (غير محدد)، يتم اختيار a_2، a_1 بناء على مبدأ الانحياز والتباين

Span: يشير إلى الفضاء الخطي المتولد(linear span) بوساطة مجموعة من الدوال أما الضرب التنسوري ⊗ هو طريقة لبناء دوال جديدة باستعمال ضرب القيم كما يلي

$$(\varphi_k \otimes \varphi_j)(x، y) = \varphi_k(x) * \varphi_j(y) \qquad (6)$$

حيث ان y،x قد يمثلان متغيرين منفصلين في نفس المجال الناتج هو دالة في متغيرين ويستخدم لتقريب الدوال ذات البعدين أو اكثر .وتعد دوال الاساس $\varphi_j، \varphi_k$ هي المكونات التي تم استعمالها لبناء فضاء الاسقاط ويمكن ان تكون هذه الدوال مختلفة الانواع حسب المشكلة قيد الدراسة .وبما ان الدالة $m(\tilde{X})$ لا توجد لها معلومات مسبقة يمكن استعمال دوال عامة مثل دوال فورير (fourier basis) وتكون هذه الدوال مناسبة لتقريب الدوال الدورية .أما كيفية بناء الفضاء $V_{a_1}، _\infty$ يمكننا تعريف كيفية بناء هذا الفضاء حسب المثال الآتي :

$$a_1 = 1، a_2 = 2، \varphi_k = \sin(kX) \text{ ،} \varphi_j(y) = \cos(jy)$$

الضرب التنسوري للدوال الاساسية سيكون كالآتي:

$$\varphi_1 \otimes \varphi_1، \varphi_1 \otimes \varphi_2، \varphi_2 \otimes \varphi_1، \varphi_2 \otimes \varphi_2$$

وعندئذ سيكون الفضاء المتولد كما يأتي:

$$V_{a_1,\infty} = span\{\sin(x)\cos(y)\,،\sin(x)\cos(2y)\,،\sin(2x)\cos(y)\,،\sin(2x)\cos(2y)\}$$

وبصورة عامة عند ضرب $\varphi_k \otimes \varphi_j$ ستكون كل قيمة إلى $k،j$ تحدد دالة جديدة حيث ان الفضاء $V_{a_1,\infty}$ يتضمن جميع التوليفات الخطية لهذه الدوال:[7]

$$f(x،y) = \sum_{j=1}^{a_1}\sum_{k=1}^{a_2} c_{jk}(\varphi_k \otimes \varphi_j)(x،y) \qquad (7)$$

ان الغرض الاساسي من استعمال فضاء الاسقاط هو التقليل من تعقيد تقدير الدالة($m(\tilde{X})$ عبر تقريبها بدالة موجودة داخل الفضاء $V_{a_1,\infty}$ وحيث ان الاسقاط العمودي المستعمل $\pi^{\perp}_{a_1,\infty}$ يضمن ان الدالة المقدرة هي الاقرب إلى ($m(\tilde{X})$ ضمن هذا الفضاء. و يتم حساب المعيار التجريبي باستعمال البيانات x_1، .... ،x_n وهو معيار رياضي يستخدم لتقيم جودة التقدير في حالة استعمال الاسقاط حيث يقيم مدى دقة الدالة المقدرة ($\hat{\mathrm{m}}(\tilde{X})$ بالمقارنة مع الدالة الحقيقية $m(\tilde{X})$ وحسب البيانات المشاهدة ويعتبر من الادوات الاساسية في تقدير الدوال اللامعلمية باستعمال فضاء الاسقاط ويتم حسابه كالآتي[7]:

$$\|T\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}\|T(X_i)\|^2 \qquad (8)$$

وعند تقدير الدالة $m(\tilde{X})$ باستعمال الاسقاط $\hat{\mathrm{m}}(\tilde{X})_{a_1,\infty}$ سيكون المعيار التجريبي كالآتي :

$$\|m(\tilde{X}) - \hat{\mathrm{m}}(\tilde{X})\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}\|m(\tilde{X}) - \hat{\mathrm{m}}(\tilde{X})_{a_1,\infty}\|_n^2 \qquad (9)$$

ويتم اختيار حد الجزاء $pen(a_1)$ الذي يضاف إلى معيار الخطأ التجريبي بعناية ، بهدف تحسين جودة التقدير وتكمن اهمية هذا الحد في تحقيق التوازن بين التحيز والتباين مما يساعد على ضبط أداء النموذج ، ويعمل حد الجزاء *(penalty Term)* على تقليل خطر الافراط في التكييف *(overfitting)* وهي المشكلة التي تحدث عندما يكون النموذج معقد جدا على سبيل المثال يكون الفضاء الاسقاطي a_1 كبير جدا ومشكلة التبسيط المفرط (under fitting) التي تحدث عندما يكون النموذج بسيطا جدا اي ان a_1 صغير جدا حيث تكتب الدالة بعد اضافة حد الجزاء كالآتي :

$$E\|m(\tilde{X}) - \hat{\mathrm{m}}(\tilde{X})_{a_1,\infty}\|_n^2 + c(\zeta).pen(a_1) \qquad (10)$$

حيث ان $c(\zeta).pen(a_1)$ يمثل حد الجزاء الذي يتم تعديله بالمعامل $c(\zeta)$ ويتم اختياره حسب المقدار الآتي :

$$c(\zeta) = \frac{(2+\zeta)}{(1+\zeta)}، \qquad \zeta > 0$$

يتم اختيار حد الجزاء $pen(a_1)$ بحيث يكون مرتبط مع تعقيد النموذج حيث ان تعقيد النموذج يتناسب مع عدد الابعاد في فضاء الاسقاط حيث كلما زاد عدد الابعاد زاد حد الجزاء :

$$pen(a_1) \propto \frac{\dim V_{a_1,\infty}}{n}$$

نقوم بتحسين اختيار البعد a_1 وتُعدُّ هذه المرحلة من اهم المراحل لتقدير الدوال اللامعلمية باستعمال فضاءات الاسقاط حيث ان الهدف الاساسي من هذه المرحلة هو اختيار البعد الأفضل a_1 لفضاء الاسقاط $V_{a_1,\infty}$ الذي يحسن من جودة تقدير الدالة اللامعلمية ($\hat{\mathrm{m}}(\tilde{X})$ وكما هو معرف ان a_1 هو عدد دوال الاساس حيث كلما ازداد a_1 يزداد تعقيد النموذج أما اذا كان a_1 صغير جدا فان النموذج يكون خالياً من التعقيد حيث في هذه المرحلة سيتم اختيار القيمة المثلى a_1 بحيث يتم الموازنة بين التحيز والتباين من خلال تقليل الخطأ الكلي ويتم تحسين اختيار a_1 وفق الخطوات الآتية :

1- يتم اختيار مجموعة من القيم إلى a_1 وكالآتي مثلا $a_1 \in \{1,2,3, \dots, N\}$ حيث ان N هو الحد الاعلى لعدد الابعاد في فضاء الاسقاط $V_{a_1,\infty}$.

2-حساب دالة الكلفة الآتية لكل قيمة من قيم المجموعة اعلاه وحسب الآتي:

$$c(a_1) = \|m(\tilde{X}) - \hat{\mathrm{m}}(\tilde{X})_{a_1,\infty}\|_n^2 + c(\zeta).pen(a_1) \qquad (11)$$

3- يتم اختيار البعد a_1^* الذي يقلل دالة الكلفة اعلاه:

$$a_1^* = argmin_{a_1} c(a_1)$$

وبعد تحديد البعد الامثل a_1^* يتم تحديد الفضاء الامثل $V_{a_1^*,\infty}$ الذي يتكون من مجموعة من دوال الاساس φ_k

$$V_{a_1^*,\infty} = span\{\varphi_1(x), \varphi_2(x), \dots, \varphi_{a_1^*}(x)\} \qquad (12)$$

نجد التقدير الأفضل $\hat{m}(\tilde{X})$ للدالة $m(\tilde{X})$ باستعمال الفضاء الامثل الذي تم تحديده مسبقا $V_{a_1^*,\infty}$ حيث تم استعمال الاسقاط العمودي لتقريب $m(\tilde{X})$ ضمن هذا الفضاء بطريقة تقلل الخطأ حيث يتم بناء $\hat{m}(\tilde{X})$ بوصفه عنصراً في الفضاء الامثل $V_{a_1^*,\infty}$ الذي يكون اقرب تقدير ممكن للدالة $m(\tilde{X})$ في هذا الفضاء . ان الاسقاط العمودي يجعل الفرق بين الدالة $m(\tilde{X})$ والتقدير $\hat{m}(\tilde{X})$ عموديا على الفضاء $V_{a_1^*,\infty}$ ويمكن تمثيله بالمعادلة الآتية :

$$\langle\ m(\tilde{X}) - \hat{m}(\tilde{X}), \varphi_k\ \rangle = 0$$

$\forall \varphi_k \in V_{a_1^*,\infty}$

حيث ان المقدر النهائي للدالة $m(\tilde{X})$ يمكن كتابته كالآتي :

$$\hat{m}(\tilde{X}) = \sum_{k=1}^{a_1^*} \hat{a}_k\, \varphi_k(x) \qquad (13)$$

حيث ان a_k تمثل مقدار مساهمة كل دالة اساس $\varphi_k(x)$ في التقدير ويعبر عنها بمعاملات الاسقاط .ولإيجاد معاملات الاسقاط $\hat{a}_k$ يتم ذلك عن طريق تحقيق شرط الاسقاط العمودي

$$\langle\ m(\tilde{X}) - \hat{m}(\tilde{X}), \varphi_k\ \rangle = 0 \qquad \forall k = 1,2, \dots . , a_1^*$$

وهذا بدوره سوف يؤدي إلى نظام من المعادلات الخطية وكالآتي:

$$\sum_{j=1}^{a_1^*} \hat{a}_j \langle \varphi_j, \varphi_k \rangle = (m(\tilde{X}), \varphi_K) \ \forall k$$

ولضمان الأداء المتوقع للمقدر النهائي $\widehat{m}(\tilde{X})_{a_1^*,\infty}$ واختيار البعد a_1^* بطريقة التوازن بين خطأ التقريب وخطأ التباين ولفهم أداء المقدر النهائي وكيفية تحسينه بناء على a_1 وحجم العينة يتم تطبيق المعادلة الآتية :[7]

$$E\|m(\tilde{X}) - \hat{m}_{a_1,\infty}(\tilde{X})\|_n^2 \leq (1+\zeta) inf_{a_1 \in A_n} [E\|m(\tilde{X}) - \pi_{a_{1,\infty}}^{op} m(\tilde{X})\|_n^2$$

$$+ \ c(\zeta) Pen(a_1) + \frac{c_0}{n} \qquad \dots (14)$$

حيث ان $\widehat{m}_{a_1}$ ، $(\tilde{X})_{,\infty}$ تمثل المقدر النهائي للدالة $m(\tilde{X})$.

9- مقدر انحدار النواة من نوع الالتفاف_Convolution type kernel regression estimator

تعد طريقة التقدير باستعمال الالتفاف Convolution type kernel regression estimator من الطرائق المستعملة لتقدير دوال الانحدار اللامعلمية حيث تقوم بإدخال تعديلات على طريقة تقدير الانحدار التقليدية التي تعتمد على النواة (Kernel) والهدف منها هو تقليل التحيز والتباين في التقديرات .ومن مميزات هذه الطريقة هي مرونة التقدير حيث لا تحتاج إلى افتراض شكل معين للدالة اذ يمكنها التعامل مع العلاقات غير الخطية بسهولة و يمكن استعمالها عند حجوم العينات المختلفة حيث تعمل بشكل فعال مع احجام العينات الكبيرة والصغيرة. ان مقدر ناداريا واتسون (Nadaraya-Watson) هو من طرق التقدير الشائعة باستعمال النواة (Kernel) لتقدير دالة الانحدار دون افتراض شكل محدد للارتباط بين المتغيرات وهي من الطرائق البسيطة وسهلة الفهم ، لكنها قد لا تكون مثالية في البيانات ذات التوزيع الغير متجانس ويكون هذا المقدر معدل موزون ومحدد للقيم حيث ان دالة النواة (Kernel) تعطي أوزانا أعلى للنقاط الأقرب إلى النقطة (x) وأوزاناً اقل للنقاط الأبعد وتكون صيغته كما يأتي :

$$\widehat{m}(\tilde{X}) = \frac{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)} \qquad (15)$$

يكون الالتفاف بالتعديل على دالة النواة وهو ما يسمى بمقدر غاسر ميلر Gasser-Muller Estimator ويتميز هذا المقدر بأنه يعتمد على طريقة التفاف معدلة وذلك للحصول على تقديرات اكثر دقة مقارنة بمقدرات النواة التقليدية مثل ناداريا واتسون، وخاصة عند التعامل مع بيانات غير موزعة بشكل متساو ، وتكون صيغته الرياضية بالشكل التالي:[9]

$$\widehat{m}(\tilde{X}) = \sum_{i=1}^{n} c_i \int_{a_i}^{b_i} K\left(\frac{x-u}{h}\right) du \qquad (16)$$

حيث في مقدر غاسر- ميلر نحتاج إلى تحديد حدود التكامل a_i, b_i بشكل مناسب يمكن ان تكون هذه الحدود مرتبطة بنقاط البيانات x_i مثلا

$$a_i = 0.5(x_{i-1} + x_i)$$

$$b_i = a_i + 1$$

ولسهولة الحساب يمكن جعل $c_i = 1$ أما دالة النواة فيمكن استعمال نفس دالة النواة المستعملة في مقدر ناداريا - واتسون وبعد تعديل الحدود واختيار الأوزان يمكن كتابة مقدر غاسر – ميلر بالشكل التالي:[9]

$$\widehat{m}(\tilde{X}) = \sum_{i=1}^{n} \int_{x_i-h}^{x_i+h} \frac{1}{h} K\left(\frac{x-u}{h}\right) du\, Y_i \qquad (17)$$

ان هذا التحويل يساعد في تقليل التباين وتحسين دقة التقدير. ان مقدر ناداريا –واتسون يستخدم لتقدير القيم المتوقعة لدالة الانحدار بناءً على نقاط البيانات المعطاة وهو يعتمد على حساب المتوسط الموزون للقيم بناءً على المسافة من النقطة المستهدفة باستعمال دالة النواة، أما مقدر غاسر – ميلر فهو يعزز فكرة تقدير القيم المتوقعة عن طريق استعمال تكامل دالة النواة حول نقطة معينة مما يسمح بجمع المعلومات من نطاق أوسع من البيانات.

10 - المحاكاة

سوف يتم هنا استعمال اسلوب المحاكاة (Simulation Procedure) لغرض تطبيق طرائق التقدير اللامعلمية التي تم التطرق اليها في الجانب النظري وذلك من اجل الوصول إلى أفضل طريقة تقدير نموذج انحدار داربن المكاني شبه المعلمي (SPSDM) اذ تم استعمال معيار متوسط مربعات الخطأ (MSE). و من خلال استعمال برنامج (MATLAB) تم تنفيذ عدد من تجارب المحاكاة وبالاعتماد على عينة حجمها (9)، وهي وذلك في توليد البيانات التي تخص المتغيرات العشوائية المتضمنة في نموذج انحدار داربن المكاني شبه المعلمي(SP-SDM) ،واستعملت ثلاث قيم مختلفة لمصفوفة التباين والتباين المشترك والمتوسط العام وكما يأتي:

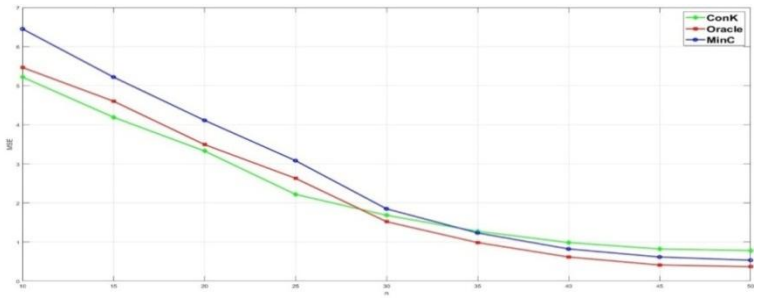
$\underline{\mu} = \underline{0}$ ،1 – $\Sigma = I$ ، $\underline{\mu} = \underline{0}$، 2- $\Sigma = 0.5^{|i-j|}$، $\underline{\mu} = \hat{\underline{\mu}}$، 3 – $\Sigma = \hat{\Sigma}$ وبتكرار التجربة (3000) مرة من اجل الوصول إلى نتائج اكثر دقة واتساق اذ تم الوصول إلى أفضل طريقة تقدير لأنموذج انحدار ديربن المكاني شبه المعلمي SP-SDM وكانت النتائج كما موضحة في الجداول التالية:

جدول (3-1) يمثل نتائج التقدير لجميع حجوم العينات وللمقارنة بين الطرائق اللامعلمية المستعملة في تقدير النموذج عند $\underline{\mu} = \underline{0}$ ، $\Sigma = I$

n	ConK	Oracle
10	7.8994	8.2726
15	6.3444	6.9664
20	5.0382	5.2872
25	3.3588	3.9808
30	2.5502	2.3014
35	1.9282	1.4928
40	1.4928	0.9334
45	1.2441	0.6226
50	1.1818	0.5598

جدول (3-2) يمثل المقارنة بين الطرائق الثلاثة المستعملة حسب معيار(MSE)

n	*ConK*	*Oracle*
10	*5.2197*	*5.4663*
15	*4.1922*	*4.6032*
20	*3.3291*	*3.4935*
25	*2.2194*	*2.6304*
30	*1.6851*	*1.5207*
35	*1.2741*	*0.9864*
40	*0.9864*	*0.6165*
45	*0.8221*	*0.4113*
50	*0.7809*	*0.3699*



شكل (3-1) يوضح مقارنة بين الطرائق باستعمال معيار (MSE)

من نتائج التحليل في الجدول رقم (3-1) و(3-2) يمكن ملاحظة الآتي:

1 - ان معيار المسافة الاقليدية Euclidean distance ينخفض تدريجيا بزيادة حجم العينة وللطريقتين (Oracle)، Conk وهذا يدل على ان دقة التقدير تتحسن مع زيادة عدد البيانات .

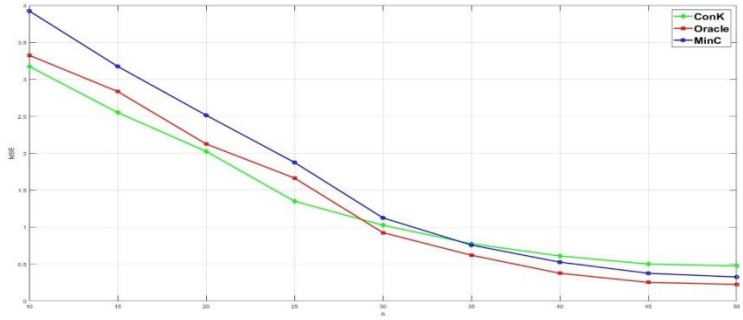
2- وكذلك هو الحال عند استعمال معيار متوسط مربعات الخطأ(MSE) يتضح عند زيادة حجم العينة تظهر جميع الطرائق انخفاضا في قيم (MSE) وهذا يدل على ان التقديرات تتحسن مع زيادة حجم العينة .

جدول(3-3)يمثل نتائج المقارنة بين الطرائق عند $\underline{\mu} = \underline{0}$ ، $\Sigma = 0.5^{|i-j|}$

n	*ConK*	*Oracle*
10	*3.8227*	*4.0033*
15	*3.0702*	*3.3712*
20	*2.4381*	*2.5585*
25	*1.6254*	*1.9264*
30	*1.2341*	*1.1137*
35	*0.9331*	*0.7224*
40	*0.7224*	*0.4515*
45	*0.6020*	*0.3017*
50	*0.5719*	*0.2709*

جدول (3-4) يمثل المقارنة بين الطريقتين حسب معيار المقارنة (MSE)

n	*ConK*	*Oracle*
10	*3.1751*	*3.3252*
15	*2.5508*	*2.8353*
20	*2.0256*	*2.1255*
25	*1.3500*	*1.6634*
30	*1.0259*	*0.9254*
35	*0.7754*	*0.6200*
40	*0.6100*	*0.3759*
45	*0.5001*	*0.2523*
50	*0.4751*	*0.2251*



شكل (2-3) يوضح مقارنة بين الطرائق باستعمال معيار (MSE)

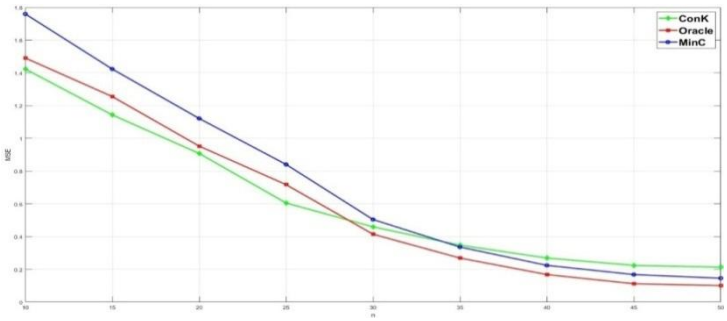
نلاحظ من نتائج التحليل للجدولين(3-4) و(3-5) عند $\underline{\mu} = \underline{0}$ ، $\Sigma = 0.5^{|i-j|}$ نلاحظ انه عند حجوم العينات الصغيرة فان طريقة نواة الالتفاف (Conk)هي الأفضل تليها طريقة الأوراكل (Oracle) بالمرتبة الثانية أما عند حجوم العينات الكبيرة فأن طريقة الأوراكل (Oracle) تتفوق في الأداء وهذا يدل على ان أداء طريقة الأوراكل يتحسن كلما يزداد حجم العينة ويصبح قريبا للأداء الامثل وكذلك بالنسبة لمعيار (MSE) نلاحظ عند زيادة حجم العينة تظهر جميع الطرائق انخفاضا في (MSE) حيث ان التقديرات تتحسن بزيادة حجم العينة .

جدول(3-5) يمثل نتائج المقارنة بين الطريقتين عند $\underline{\mu} = \underline{\hat{\mu}}$ ، $\Sigma = \hat{\Sigma}$

n	*ConK*	*Oracle*
10	*2.5425*	*2.6626*
15	*2.0420*	*2.2422*
20	*1.6216*	*1.7017*
25	*1.0810*	*1.2812*
30	*0.8208*	*0.7407*
35	*0.6206*	*0.4804*
40	*0.4804*	*0.3003*
45	*0.4004*	*0.2002*
50	*0.3803*	*0.1801*

جدول (3-6) يمثل المقارنة بين الطريقتين حسب معيار المقارنة (MSE)

n	*ConK*	*Oracle*
10	*1.4236*	*1.4909*
15	*1.1434*	*1.2555*
20	*0.9080*	*0.9528*
25	*0.6053*	*0.7174*
30	*0.4596*	*0.4147*
35	*0.3475*	*0.2690*
40	*0.2690*	*0.1681*
45	*0.2242*	*0.1121*
50	*0.2129*	*0.1008*



شكل(3-3) يوضح مقارنة بين الطرائق باستعمال معيار (MSE)

من ملاحظة الجدولين (3-5)،(3-6) عند $\underline{\mu} = \underline{\hat{\mu}}$ ، $\Sigma = \hat{\Sigma}$ نلاحظ ان القيم التي تعبر عن المسافة الاقليدية تظهر انخفاضا ملحوظا حيث كلما كانت القيم اصغر ،كانت الطريقة أفضل واكثر دقة حيث نلاحظ كلما تغير حجم العينة تتغير هذه القيم ،ونلاحظ ايضا أن أداء طريقة الأوراكل (Oracle) يظهر دائما نتائج متفوقة (قيم اقل) مقارنة بطريقة (Conk) وكذلك بالنسبة لمعيار (MSE) في الجدول (3-6)، فإن القيم تمثل معيار (MSE) لكل طريقة حيث كلما كانت القيمة اصغر، كان أداء الطريقة أفضل وبهذا نتوصل إلى أن طريقة الأوراكل (Oracle) تظهر دائما الأداء الأفضل عبر جميع احجام العينات مما يجعله المرجح المثالي. أما طريقة نواة الال تفاف(Conk) فأنها تتمتع بأداء جيد ويقترب من الأوراكل مع زيادة حجم العينة، ونجد أن زيادة حجم العينة يقلل من MSE ولجميع الطرائق مما يشير إلى ان زيادة حجم البيانات يحسن من دقة التقديرات.

الاستنتاجات

1- اظهرت نتائج الجداول (3-1)، (3-2) و(3-3) انه كلما ازداد حجم العينة (n) تكون القيم الناتجة للمسافة الاقليدية اصغر ولجميع الطرائق المستعملة مما يعني ان أداء كل طريقة يتحسن بزيادة حجم العينة .

2- من خلال النتائج يتضح ان قيم معيار متوسط مربعات الخطأ (MSE) تقل كلما ازداد حجم العينة ولجميع الطرائق المدروسة وهذا يدل على جميع الطرائق تمتلك خاصية الاتساق .

3- اظهرت النتائج من خلال استعمال معيار المقارنة (MSE) ان طريقة الأوراكل (Oracle) هي الأفضل وخصوصا عند حجوم العينات الكبيرة لأنها حققت اقل قيمة لمعيار (MSE) وتليها طريقة انحدار النواة من نوع الالتفاف(ConK)

التوصيات

1- اجراء دراسات مستقبلية لأستعمال طرائق لامعلمية حصينة لتقدير انموذج انحدار ديربن المكاني شبه المعلمي.

2- استعمال نماذج مكانية اخرى في تمثيل ظاهر ملوحة التربة في المحافظات.

3- اجراء دراسات مستقبلية تتناول مقارنة طرائق لامعلمية أخرى مع الطرائق التي تم اعتمادها في الرسالة .

4- توفير قاعدة بيانات متاحة للباحثين في وزارة الزراعة لدراسة المشاكل التي تواجه التربة في كافة محافظات العراق.

المصادر العربية:

1- عطره ، سامي غني خضير ، (2011) **"طرائق بيز في تحليل النموذج القياسي الاقتصادي المكاني مع تطبيق عملي"** اطروحة دكتوراه احصاء ، كلية الادارة والاقتصاد ، جامعة بغداد.

2-- عكار، أحمد عبيد علي ، (2019) **"تقدير الانحدار شبة المعلمي للبيانات المعتمدة مكانيا مع تطبيق عملي"**، أطروحة دكتوراه إحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد.

المصادر الاجنبية:

3- Anselin ،L. (1988)،**"Spatial Econometrics : Method and Models "**،Kluwer Academic ،Publishers ،Dordrecht ،The Netherlands.

4- Anselin، L .(1992)، **"Spatial Data Analysis With GIS:AN Introduction to Appalachian the Social Sciences "**، National Center for Geographic Information and Analysis ، University of California ،Santa Barbara.

5- Atikah، N.، Widodo، B.، Rahardjo، S.، Kholifia، N.، & Afifah، D. L. (2021، March**)." The efficiency of Spatial Durbin Model (SDM) parameters estimation on advertisement tax revenue in Malang City".** In *Journal of Physics: Conference Series* (Vol. 1821، No. 1، p. 012012). IOP Publishing.

6-Basile، R. (2008). **"Regional economic growth in Europe: A semiparametric spatial dependence approach"**. *Papers in Regional Science*، *87*(4)، 527-545.

7- Chagny، G.، Meynaoui، A.، & Roche، A. (2022). "**Adaptive nonparametric estimation in the functional linear model with functional output".**

8- Gumprecht، D. (2005). **" Spatial Methods in Econometrics. An Application to R&D Spillovers ".**

9- Herrmann، E. (1996). ***"On the convolution type kernel regression estimator"***. *Preprints، Dept. Math.، Tech. Univ. Darmstadt، 31 pp.*

10- Lesage، James P. (1999)،**"The Theory and Practice of Spatial Econometrics"**، Department of Economics University of Toledo .