مقارنة بين طريقتي النواة لمتعدد الحدود الموضعي والشريحه الجزائية في تقدير أنموذج المعاملات المتغيرة*

أ.د. ظافر حسين رشيد
كلية الادارة والاقتصاد/ جامعة بغداد

م.م. حسام عبد الرزاق رشيد
كلية الادارة والاقتصاد/ الجامعة المستنصرية

المستخلص

هنالك العديد من الظواهر الاقتصادية والمالية والبيئية وغيرها تحتاج الى بناء الأنموذج المناسب الذي يمثل العلاقات السببيه. وعملية بناء النموذج تعتمد على العوامل المحيطه بالظاهرة، ولأن اعتماد الانحدار الخطي التقليدي يكون غير كفوء بسبب العلاقة بين المتغيرات التوضيحية والاستجابة معلومة وللتوسع في تمثيل العلاقات بين المتغيرات التي تمثل الظاهرة قيد البحث تم استخدام انموذج المعاملات المتغيرة (Varying Coefficient Models (VCM)) والتي تفترض تأثيرات المتغيرات التوضيحية متغيرة باعتماد متغير توضيحي آخر. أن هذه الهيكلية تجنبنا ما يعرف بمشكلة البعدية (تعدد الابعاد) (Curse of Dimensionality) والتي تظهر عند استخدام الطرائق اللامعلمية في التقدير. وقد تم تقدير معاملات أنموذج المعاملات المتغيرة بأعتماد الطرائق اللامعلمية والتي هي طريقة النواة لمتعدد الحدود الموضعي (LPK)) Local Polynomial Kernel) وطريقة الشريحة الجزائية (Penalized Spline (PS)) اذ تم استخدام اسلوب المحاكاة لاجل المقارنه ووجد ان طريقة LPK هي الافضل.

المصطلحات الرئيسية للبحث/ أنموذج المعاملات المتغيرة- طريقة النواة لمتعدد الحدود الموضعي- طريقة الشريحه الجزائية- معيار العبور الشرعي العام.

مجلة العلوم
الاقتصادية والإدارية
المجلد 20
العدد ٧٨
لسنة ٢٠١٤
الصفحات ٣٢٥-٣٣٨

* بحث مستل من اطروحة الدكتوراه

١. المقدمة

ان أنموذجات الانحدار الخطي المعممة (Generalized Linear Regression) تعد من الأنموذجات المعلميه التي تستند في تحليل العلاقة بين المتغيرات على عدة افتراضات ومن بين هذه الافتراضات تلك التي تنص على خطية العلاقة بين المتغيرات، ومع ذلك فأن هذا الافتراض يتصف بعدم المرونة الامر الذي يؤدي الى تقديرات متحيزة في حالة عدم تحقق هذا الافتراض، مما يجعل تلك الأنموذجات الخطية غير واقعية من الناحية التطبيقية وللحصول على أنموذجات أكثر واقعية اقترحت عدة أنموذجات معلمية تهدف الى جعل تلك الافتراضات اكثر مرونة مع الاخذ بنظر العناية امكانية تطبيق تلك الأنموذجات لتفادي وضع قيود عند نمذجة العلاقة بين المتغيرات. اعتمد الباحثون على النمذجة اللامعلمية التي لا تضع اية قيود عند صياغة الأنموذج غير ان الأنموذجات اللامعلمية قد تفشل في تضمين بعض المعلومات السابقة (Prior Information) المتوفره في العينة مما يؤدي الى تقدير ذي تباين كبير لدالة الانحدار غير المعلومة. المشكلة الاخرى التي تعاني منها الطرائق اللامعلمية هي البعدية (الابعاد) Curse of (Dimensionality) تلك المشكلة تجعل الطرائق اللامعلمية ضعيفة عملياً وذلك عندما يكون هنالك اكثر من متغير توضيحي، ولتفادي هذه المشكلة سنعتمد الأنموذجات التجمعية (Additive) أو ما يعرف بأنموذجات المعاملات المتغيرة (VCM) Varying Coefficient Models))

والتي تعد من الأنموذجات الشبه معلمية (Semiparametric) من ناحية الهيكلية [5]. تم اعتماد الطرائق اللامعلمية لتقدير معاملات أنموذج المعاملات المتغيرة اذ يهدف البحث الى تناول طريقة النواة لمتعدد الحدود الموضعي (LPK) Local Polynomial Kernel)) وبتوظيف دالة نواة مقترحة ومقارنتها مع طريقة الشريحة الجزائية (PS) Penalized Spline)). تمت المقارنةبأستخدام المحاكاة فيما يخص تقدير الدالة التي بموجبها يعتمد معامل المتغير التوضيحي على متغير مستقل آخر.

٢. الجانب النظري

٢-١ أنموذجات المعاملات المتغيرة (VCM) Varying Coefficient Models))

ان صيغة الأنموذجات التجمعية المعممة (GAM) Generalized Additive Models)) والمبينة بالصيغة الاتية التي تفترض ان الاستجابة عبارة تركيب خطي لمجموعة الدوال غير المعلومة [6].

$$\mathrm{E}(y) = a_1(x_1) + a_2(x_2) + \cdots + a_L(x_L) \qquad \cdots (1)$$

اذ ان $\mathrm{E}(y)$ يمثل توقع متغير الاستجابة y وان x_k ، $k = 1,2,\cdots,L$ هي المتغيرات التوضيحية وان a_k تمثل دوال غير محددة بدلالة المتغيرات التوضيحية (التي نريد تقديرها) غير ان هنالك حالات تكون فيها معاملات المتغيرات التوضيحية غير ثابتة وانما هي دالة تعتمد على متغير آخر مما يجعلها ذات قيم متغيرة والأنموذج الرياضي الانسب لوصف مثل هذه العلاقات يعرف بأنموذج المعاملات المتغيرة (Varying Coefficient Model (VCM)) وصيغته الرياضية تكتب كالاتي [2] [3] :

$$y_t = \sum_{k=1}^{L} a_k(u_t)\, x_{tk} + \varepsilon_t \qquad , t = 1,2,\cdots,n \qquad \cdots (2)$$

اذ ان y_t ، $t = 1,2,\cdots,n$ يمثل متغير قيم الاستجابة وان x_{tk} ، $k = 1,2,\cdots,L$ هي المتغيرات التوضيحية المراد تقدير تأثيرها في متغير الاستجابة y اما $a_k(u_t)$ فتمثل معاملات المتغيرات التوضيحية في معادلة الانحدار والتي هي عبارة عن دالة في المتغير u وليست ثابتة و ان ε_t تشير الى الخطأ عند الزمن t وهو متغير عشوائي بمتوسط صفر وتباين ثابت σ^2. من الأنموذج (٢) نلاحظ ان المتغيرات u تؤثر في المعاملات $a_k(u_t)$ الامر الذي يؤدي الى جعلها ذات قيما متغيرة وليست ثابتة ومن خلال اعتمادية تلك المعاملات على المتغيرات u نجد انها تؤدي الى حالة خاصة من التفاعل بين المتغير u والمتغير التوضيحي x .

٢-٢ طريقة النواة لمتعدد الحدود الموضعي ((LPK) Local Polynomial Kernel) [2][3]

تعتمد طريقة النواة لمتعدد الحدود الموضعي على فكرة تقريب معاملات المتغيرات التوضيحية لأنموذج المعاملات المتغيرة، والمبين في الصيغة (٢) اذ ان تلك المعاملات $a_k(u)$ هي عبارة عن دوال في المتغير u. هذا التقريب يعتمد بالدرجة الاساس على متعدد الحدود من الدرجة p وكذلك النقطة الثابتة u_0 ، بأفتراض ان الدالة $a_k(u)$ التي نود تقديرها تمتلك مشتقات مستمرة لغاية المشتقة $p+1$ وبالاعتماد على تقريب تايلر (**Taylor Expansion**) يمكن كتابة الدالة $a_k(u)$ كالاتي:

$$\begin{aligned} a_k(u) &= \sum_{j=0}^{p} \frac{1}{j!}\, a_k^{(j)}(u_0)(u_i - u_0)^j \\ &= \sum_{j=0}^{p} \beta_{jk}\,(u_i - u_0)^j \qquad \cdots (3) \end{aligned}$$

اذ ان $\beta_{jk} = \frac{1}{j!}\, a_k^{(j)}(u_0)$ تمثل المعاملات التي سنقوم بتقديرها وان $a_k^{(j)}(u_0)$ تمثل مشتقة الدالة $a_k(u_0)$ للرتبة j عند النقطة u_0. ان متعدد الحدود (٣) يمكننا من التقدير الموضعي لمعاملات متعدد الحدود β_{jk} ، $j = 0,1,\cdots,p$ وذلك بتصغير معيار المربعات الصغرى الموزون (**Weighted Least Squares (WLS)**) والذي يكتب بهيئة المصفوفات بالشكل الاتي:

$$\min_{\beta} (Y - X\beta)^T W (Y - X\beta) \qquad \cdots (4)$$

اذ ان $Y=[y_1, y_2, \cdots, y_n]^T$ يمثل متجه لقيم الاستجابة ذو درجة $n\times1$ وان X تمثل مصفوفة التصميم ذات درجة $n\times(p+1)L$ والتي تكتب كالاتي:

$$X=\begin{bmatrix} x_1^T & (u_1-u_0)x_1^T & \cdots & (u_1-u_0)^p x_1^T \\ x_2^T & (u_2-u_0)x_2^T & \cdots & (u_2-u_0)^p x_2^T \\ \vdots & \vdots & \cdots & \vdots \\ x_n^T & (u_n-u_0)x_n^T & \cdots & (u_n-u_0)^p x_n^T \end{bmatrix}$$

علما ان x_t^T ، $t=1,2,\cdots,n$ تحدد تبعاً لعدد المتغيرات التوضيحية وتكتب كالاتي:

$$x_t=[x_{t1}\ ,\ x_{t2}\ ,\ \cdots\ ,\ x_{tL}]^T \qquad , \qquad t=1,2,\cdots,n$$

اما المصفوفة W فتمثل مصفوفة قطرية للاوزان من الدرجة $n\times n$ وتكتب بالشكل الاتي:

$$W=diag\ \{K_h(u_t-u_0)\} \qquad , \qquad t=1,2,\cdots,n$$

اذ ان K_h يتم تحديدها بالاعتماد على دالة النواة K اذ تم اعتماد دالة النواة المتمثلة بالصيغة الرياضية الاتية والتي تم اقتراحها من قبل الباحثين:

$$K(u)=\frac{3315}{4096}(1-u^4)^4\ ,\ I_{\{|u|\le1\}}$$

وباعتماد الصيغة الاتية K_h اذ ان عرض الحزمة (Bandwidth) الذي يرمز له بـ h

$$K_h(u_t-u_0)=\frac{1}{h}K(\frac{1}{h}(u_t-u_0)) \qquad \cdots (5)$$

ان دالة $K_h(\cdot)$ نحصل عليها بأعادة احتساب دالة النواة $K(\cdot)$ وتحديد عرض الحزمة المناسب الذي يكون ذو قيماً موجبة $(h>0)$ ويستخدم لتحديد البيانات ضمن مدى الجوار الموضعي (Local Neighborhood)، والمعرف بالشكل $I_h(u_0)=[u_0-h, u_0+h]$. ان دالة النواة تمكننا من تحديد المشاهدات ضمن المدى $I_h(u_0)$ التي تسهم في توفيق منحنى دالة الانحدار عند نقطة (u_0).

ان متجه المعاملات المراد تقديره هو $\beta = \left[\beta_0^T, \beta_1^T, \cdots, \beta_p^T\right]$ وتكون ذات سعة $(p+1)L \times 1$، اذ ان عناصر هذا المتجه هي β_j^T، $j = 0,1,\cdots,p$ وتحدد بشكل عام تبعاً لعدد المتغيرات التوضيحية وتمثل بالمتجه $\beta_j = \left[\beta_{j1}\,,\ \beta_{j2}\ ,\cdots,\ \beta_{jL}\right]^T$ ، $j = 0,1,\cdots,p$ والصيغة التقديرية لمتجه معاملات متعدد الحدود الناتج من تصغير معيار المربعات الصغرى الموزون والمبين بالصيغة (٤) تكتب كالآتي:

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \qquad \cdots (6)$$

بايجاد تقدير معاملات دوال متعدد الحدود نتمكن من ايجاد مقدر الدالة $a_k^{(j)}(u_o)$ والذي يكتب على وفق الصيغة الأتية:

$$\hat{a}_k^{(j)}(u_o) = j!\ e_{(j,k)}^T\ (X^T W X)^{-1} X^T W Y \qquad \cdots (7)$$

اذ ان $e_{(j,k)}$ تمثل متجه احادي (**Unit Vector**) ذو درجة $((p+1)L \times 1)$ والذي يحتوي على واحد عند الموقع (j,k) اما المواقع المتبقية فتكون اصفار. ومن الجدير بالملاحظة ان الاهتمام ينصب على ايجاد مقدر الدالة $a_k^{(j)}(u_o)$ عندما $(j=0)$ اي عند المشتقة الصفرية لتلك الدالة والممثلة بالاتي:

$$\hat{a}_k(u_0) = e_k^T (X^T W X)^{-1} X^T W Y \qquad \cdots (8)$$

ان تعويض تقدير المعاملات المتغيرة (٨) في معادلة الانحدار (٢) يمكننا من ايجاد القيم التنبؤيه لمتغير الاستجابة y_t , $t = 1,2,\cdots,n$

$$\hat{y}_t = \sum_{k=1}^{L} \hat{a}_k(u_t)\, x_{tk} \qquad ,t = 1,2,\cdots,n \qquad \cdots (9)$$

٢-٣ طريقة الشريحة الجزائية ((PSM) Penalized Spline Method)[8]

ذكرنا سابقا ان ايجاد مواقع العقد يتطلب تحديد عددها مسبقاً، تمهيد الشرائح الجزائية تتغلب على مشكلة التحديد المسبق والامثل لعدد العقد وذلك بأخذ جميع نقاط المتغير التوضيحي المختلفة كعقد. ان اعتماد عدد كبير من العقد بشكل قريب او مساوي لحجم العينة يؤدي الى الحصول على ما يسمى بمنحنى ممهد دون التمهيد (**Under Smoothing**) والمنحنى الناتج يكون ذو نتوءات كثيرة تعرف بالجزء غير الممهد (**Roughness**) من المنحنى، ولمعالجة هذه المشكلة يتم الاعتماد على معيار المربعات الصغرى الجزائية بعد اضافة حد الجزاء لتقدير معاملات الأنموذج المتغيرة $a_k(u)$. ولاجل تقدير أنموذج المعاملات المتغيرة سيعتمد معيار المربعات الصغرى الجزائية (**Penalized Least Squares**) والذي يكتب كالاتي:

$$\min_{\beta}(Y-Z\beta)^T\ (Y-Z\beta)+\lambda\beta^T D\beta \qquad \cdots(10)$$

وبتصغير المربعات الصغرى الجزائية (Penalized Least Squares) نحصل على تقدير المربعات الصغرى لمتجه معاملات دوال اساس قوة القطع على وفق الصيغة الاتية:

$$\hat{\beta}=(Z^T Z+\lambda D)^{-1} Z^T Y \qquad \cdots(11)$$

اذ ان $Y=[y_1,y_2,\cdots,y_n]^T$ يمثل متجه لقيم الاستجابة ذو درجة $n\times 1$.
اذ ان Z هي مصفوفة التصميم ذات سعة $n\times(p+K+1)L$ وتكتب كالاتي:

$$Z=\begin{bmatrix} x_1^T & u_1x_1^T & \cdots & u_1^p x_1^T & (u_1-\kappa_1)_+^p x_1^T & \cdots & (u_1-\kappa_K)_+^p x_1^T \\ x_2^T & u_2x_2^T & \cdots & u_2^p x_2^T & (u_2-\kappa_1)_+^p x_2^p & \cdots & (u_2-\kappa_K)_+^p x_2^T \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^T & u_nx_n^T & \cdots & u_n^p x_n^p & (u_n-\kappa_1)_+^p x_n^T & \cdots & (u_n-\kappa_K)_+^p x_n^T \end{bmatrix}$$

وان x_t ، $t=1,\cdots,n$ تكتب تبعاً لعدد المتغيرات التوضيحيه وكالاتي:

$$x_t=[x_{t1},\ x_{t2}\ ,\cdots,x_{tL}]^T$$

وان β هو متجه معاملات دوال اساس القطع ذو سعة $(P+K+1)L\times 1$ ويكتب كالاتي:

$$\beta=[\beta_o^T,\cdots,\beta_p^T,\beta_{p+1}^T,\cdots,\beta_{p+K}^T]^T$$

وان β_q ، $q=0,1,\cdots,m$ ، $m=p+K+1$ تكتب تبعاً لعدد المتغيرات التوضيحيه وكالاتي:

$$\beta_q=[\beta_{q1},\beta_{q2}\ ,\cdots,\beta_{qL}]^T$$

وان D تمثل مصفوفة التمهيد ذات سعة $(p+K+1)L\times(p+K+1)L$ وتكتب كالاتي:

$$D=\begin{bmatrix} \underset{(p+1)\mathrm{L}\times(p+1)\mathrm{L}}{0} & \underset{(p+1)\mathrm{L}\times K\mathrm{L}}{0} \\ \underset{K\mathrm{L}\times(P+1)\mathrm{L}}{0} & \underset{K\mathrm{L}\times K\mathrm{L}}{I} \end{bmatrix}$$

وبذلك نحصل على تقدير دالة معاملات الأنموذج المتغيرة بهيئة المصفوفات وكالاتي:

$$\hat{a}_k(u) = Z\hat{\beta} = Z(Z^T Z + \lambda D)^{-1} Z^T Y \qquad \cdots (12)$$

اما متجه الاستجابة المقدر تبعاً الى نقاط الزمن $t = 1,2,\cdots,n$ يكتب كالاتي:

$$\hat{y}_t = \sum_{k=1}^{L} \hat{a}_k(u_t)\ x_t \qquad \cdots (13)$$

٢-٤ الممهدات الخطية (Linear Smoothers) [10]

الممهد الخطي يعرف بانه التعبير عن متجه قيم متغير الاستجابة المقدر $\hat{y}$ ولكل قيمة من قيم المتغير التوضيحي بهيئة تركيبة خطية في متجه قيم متغير الاستجابة الحقيقية y ويعتبر الممهد خطياً اذا استطعنا ان نكتب متجه قيم متغير الاستجابة المقدر $\hat{y}$ بدلالة متجه قيم متغير الاستجابة الحقيقية y وفق الصيغة المبسطة الاتية:

$$\hat{y} = A_\rho y \qquad \cdots (14)$$

اذ ان $\hat{y} = [\hat{y}_1,\cdots,\hat{y}_n]^T$ يمثل متجة قيم متغير الاستجابة المقدرة وان A_ρ هي مصفوفة من الدرجة $n\times n$ تعرف بمصفوفة التمهيد (**Smoothing Matrix**) والتي تأخذ صيغا مختلفة تبعاً لطريقة التقدير المعتمدة والتي تعتمد بدورها على معلمة التمهيد (**Smoothing Parameter**) ρ .

٢-٤-١ درجة حرية الأنموذج (degree of freedom (df))

درجة حرية التمهيد تستخدم كمقياس لتحديد عدد المعلمات ذات التأثير والمتضمنة في الأنموذج فمثلا في انحدار الشريحة درجة حرية التمهيد تساوي $m = p + K + 1$ والتي تمثل عدد دوال الاساس متعددة الحدود من الدرجة p مضافاً اليها دوال قوة القطع K. ان درجة حرية التمهيد يعبر عنها بأثر (**Trace**) مصفوفة التمهيد A_ρ اي ان

$$df = tr(A_\rho) = \sum_{t=1}^{n} a_{tt} \qquad \cdots (15)$$

اذ ان a_{tt} هي العناصر القطرية للمصفوفة A_ρ **[10]**.

٢-٤-٢ معيار العبور الشرعي ((CV) Cross Validation)

معيار العبور الشرعي CV_ρ يستعمل بشكل واسع لاختيار انسب قيمة لمعلمة التمهيد ρ اذ ان صيغته تكتب كالاتي:

$$CV_\rho = n^{-1}\sum_{t=1}^{n}\left\{\frac{y_t-\hat{y}_t}{1-a_{tt}}\right\}^2 \qquad \cdots(16)$$

اذ ان a_{tt} تمثل العنصر t من عناصر القطر في مصفوفة التمهيد A_ρ وبذلك سيكون بامكاننا حساب قيمة العبور الشرعي لكل قيمة من قيم معلمة التمهيد ρ. نلاحظ من الصيغة (١٦) ان هذا المعيار يعتمد على الانحرافات المشاهدة $y_t-\hat{y}_t$ وعناصر القطر الرئيس لمصفوفة التمهيد فقط مما يسهل عملية الحساب [10].

٣-٤-٢ معيار العبور الشرعي العام ((GCV) Generalized Cross Validation)

يعد معيار العبور الشرعي العام **GCV** تقريب لمعيار العبور الشرعي **CV** اذ يستعمل لاختيار انسب قيمة لمعلمة التمهيد ρ ويمكن الحصول على معيار العبور الشرعي العام **GCV** من صيغة معيار العبور الشرعي **CV** المبين بالصيغه (١٦) وذلك باستبدال عناصر القطر الرئيسي a_{tt} لمصفوفة التمهيد A_ρ بمعدلها اي ان $n^{-1}\sum_{t=1}^{n}a_{tt}=n^{-1}tr(A_\rho)=df/n$ وبهذا فأن صيغة معيار العبور الشرعي العام **GCV** تكون كالاتي [10]:

$$GCV_\rho = \frac{n^{-1}\sum_{i=1}^{n}\left[y_i-\hat{y}_i\right]^2}{\left\{1-tr(A_\rho)/n\right\}^2} = n^{-1}SSE_\rho/(1-df/n)^2 \qquad \cdots(17)$$

٣. الجانب التجريبي

لغرض تطبيق طرائق التقدير اللامعلمية المعتمدة التي تم ذكرها في انفا تم اعتماد اسلوب المحاكاة (**Simulation**) والذي يحاكي عدداً كبيراً من الحالات الافتراضية المماثلة لنظيراتها في الواقع العملي بما يضمن شمولية اكبر للنتائج المتوقعة وذلك لأعتمادها في عملية التقدير والمقارنة بين تلك الطرائق المعتمدة في تقدير أنموذج المعاملات المتغيرة. تم تنفيذ تجارب المحاكاة بأعتماد ثلاثة حجوم للعينات هي ($n_3=150$, $n_2=100$, $n_1=50$) وذلك لتوليد البيانات الخاصة بالمتغيرات العشوائية المتضمنة في أنموذج المعاملات المتغيرة (**VCM**) اذ تم تكرار كل تجربة 1000 مرة وذلك لاجل الحصول على نتائج متسقة. لقد تم الاعتماد على عدة دوال رياضية مختلفة لتمثيل المعاملات المتغيرة للمتغير التوضيحي حيث تم مراعاة الصيغة الخطية واللاخطية في تلك الدوال وفيما يلي ندرج الدوال المعتمدة:

١- الدالة الخطية وصيغتها[11] $a(u)=2u$

٢- الدالة الخطية بوجود حد ثابت[9] [1] $a(u)=2u+1$

٣- الدالة من الدرجة الثانية[11] $a(u)=8(u-0.5)^2$

٤- الدالة من الدرجة الرابعة[7] [1] $a(u) = 1 - 48u + 218\ u^2 - 315\ u^3 + 145\ u^4$

٥- الدالة الغير خطية[4] [1] $a(u) = u + 2\ \exp\ (-16u^2)$

وبالاعتماد على دوال المعاملات المتغيرة المذكورة انفا يمكن كتابة صيغة أنموذج المعاملات المتغيرة والتي ستجري محاكاتهما وكالاتي:

١- الأنموذج الخطي $y = (2u)x + \in$

٢- الأنموذج الخطي $y = (2u + 1)x + \in$

٣- الأنموذج من الدرجة الثانية $y = 8(u - 0.5)^2 x + \in$

٤- الأنموذج من الدرجة الرابعة $y = (1 - 48u + 218\ u^2 - 315\ u^3 + 145\ u^4)x + \in$

٥- الأنموذج الغير خطي $y = (u + 2\ \exp(-16u^2))x + \in$

تضمن تنفيذ تجارب المحاكاة حول أنموذجات المعاملات المتغيرة كتابة عدة دوال بلغة MATLAB الاصدار (2011) لكل أنموذج من الأنموذجات الخمسة ذات المعاملات المتغيرة اذ تم تنفيذ مايأتي:

١- توليد المتغير التوضيحي x_1 بحيث يتوزع توزيعا طبيعياً بمتوسط يساوي صفر وتباين يساوي 0.02.

٢- توليد المتغير u الذي تعتمد عليه المعاملات المتغيرة بحيث يتوزع طبيعياً بمتوسط يساوي صفر وتباين يساوي 0.03.

٣- توليد المتغير العشوائي $\in$ الذي يمثل حد الخطأ العشوائي في أنموذج المعاملات المتغيرة قيد البحث بحيث يتبع التوزيع الطبيعي بمتوسط يساوي صفر وثلاث مستويات للانحراف المعياري هي $(\sigma_3 = \frac{1}{8}\ ,\ \sigma_2 = \frac{1}{4}\ ,\ \sigma_1 = \frac{1}{2})$.

٤- تعويض المتغيرات التي تم توليدها في أنموذجات المعاملات المتغيرة لأجل الحصول على متغير الاستجابة y ولكل مستوى من مستويات الانحراف المعياري في (٣) وكذلك لكل أنموذج من أنموذجات المعاملات المتغيرة الخمسة قيد البحث.

وفيما يلي نتائج الطريقتين اللامعلميتين التي تمت المقارنه بينها:

١. طريقة النواة لمتعدد الحدود الموضعي (LPK).

٢. طريقة الشريحة الجزائية (PS).

بعد الحصول على نتائج تجارب المحاكاة التي تم تكرارها 1000 مرة وبأعتماد خمسة دوال تمثل معامل المتغير التوضيحي والتي اختيرت لمحاكاة أنموذج المعاملات المتغيرة والمبينة بالجدوال من (١) الى (٥) ولجميع الأنموذجات الخمسة ولجميع حجوم العينات ومستويات الانحراف المعياري لحد خطأ الأنموذج اذ لوحظ ان قيم معدل متوسط مربعات الخطأ (MSE) تشير الى افضلية طريقة النواة لمتعدد الحدود الموضعي (LPK) وان قيم الانحراف المعياري لمتوسط مربعات الخطأ (MSE) تشير الى تجانس قيم متوسط مربعات الخطأ (MSE) ولجميع التكرارات.

جدول (١) يوضح معدل قيم معيار (MSE) ولجميع حجوم العينات ومستويات الانحراف المعياري للطرائق اللامعلمية المستخدمة في تقدير أنموذج المعاملات المتغيرة الاول

		مستويات الانحراف المعياري لحد الخطأ		
حجوم العينات	الطرائق اللامعلميه	σ=1/8	σ=1/4	σ=1/2
n_1=50	LPK	0.000098914540261 (0.000021083960484)	0.003540283147170 (0.000720458974552)	0.059426387108747 (0.011485497828554)
	PS	0.000103495243650 (0.000021990513319)	0.003648108344709 (0.000734510240536)	0.061901764259542 (0.011739756705088)
n_2=100	LPK	0.000098395054915 (0.000016115197025)	0.003520111722584 (0.000493242050924)	0.062103312813904 (0.009111342621385)
	PS	0.000101720489575 (0.000016217666858)	0.003579010388012 (0.000489337753151)	0.063052868601794 (0.009114959802796)
n_3=150	LPK	0.000098505444721 (0.000011959660920)	0.003547376481152 (0.000354393273961)	0.063117685396637 (0.007011833628771)
	PS	0.000101036678783 (0.000012378018882)	0.003584647200335 (0.000356533880629)	0.063857857256820 (0.007118649148747)

جدول (٢) يوضح معدل قيم معيار (MSE) ولجميع حجوم العينات ومستويات الانحراف المعياري للطرائق اللامعلمية المستخدمة في تقدير أنموذج المعاملات المتغيرة الثاني

		مستويات الانحراف المعياري لحد الخطأ		
حجوم العينات	الطرائق اللامعلميه	σ=1/8	σ=1/4	σ=1/2
n_1=50	LPK	0.000097760005020 (0.000020011694892)	0.003531070990053 (0.000751890214986)	0.060173596028792 (0.012797842563552)
	PS	0.000495283617050 (0.000101825361918)	0.004061956673870 (0.000812357631246)	0.062786739714586 (0.012936112531024)
n_2=100	LPK	0.000097100132212 (0.000013316786534)	0.003472382724420 (0.000521521099063)	0.060752328848187 (0.009020842424664)
	PS	0.000504157429182 (0.000067600882578)	0.003945193245062 (0.000612337559553)	0.062350698383555 (0.009032988545470)
n_3=150	LPK	0.000097253839503 (0.000010960992400)	0.003570544685812 (0.000430160580780)	0.063137862452745 (0.007707729024013)
	PS	0.000495115599425 (0.000058700789107)	0.004025622759663 (0.000459594742309)	0.064329922257749 (0.007912915898293)

جدول (٣) يوضح معدل قيم معيار (MSE) ولجميع حجوم العينات ومستويات الانحراف المعياري للطرائق اللامعلمية المستخدمة في تقدير أنموذج المعاملات المتغيرة الثالث

حجوم العينات	الطرائق اللامعلميه	مستويات الانحراف المعياري لحد الخطأ σ=1/2	σ=1/4	σ=1/8
n_1=50	LPK	0.000099160379671 (0.000017827402944)	0.003407244317579 (0.000631520784691)	0.000099716648962 (0.000019781534729)
	PS	0.001767435127701 (0.000397091944304)	0.005150774876904 (0.000965290992402)	0.001724552080232 (0.000312727323040)
n_2=100	LPK	0.000101150691607 (0.000014355486515)	0.003512686532970 (0.000601706264224)	0.000102936081512 (0.000014996785179)
	PS	0.001729489040109 (0.000245901233005)	0.005278099801777 (0.000732558513415)	0.001731175827524 (0.000258888856933)
n_3=150	LPK	0.000101125411601 (0.000011488215779)	0.003578629814083 (0.000455903249400)	0.000101984947886 (0.000011150497774)
	PS	0.001723091464703 (0.000199987083771)	0.005296871226536 (0.000543986071243)	0.001746197421893 (0.000197020174690)

جدول (٤) يوضح معدل قيم معيار (MSE) ولجميع حجوم العينات ومستويات الانحراف المعياري للطرائق اللامعلمية المستخدمة في تقدير أنموذج المعاملات المتغيرة الرابع

حجوم العينات	الطرائق اللامعلميه	مستويات الانحراف المعياري لحد الخطأ σ=1/2	σ=1/4	σ=1/8
n_1=50	LPK	0.061334996125405 (0.012840559155659)	0.003456598463195 (0.000712757650223)	0.000231708027587 (0.000075621891892)
	PS	0.064936441721075 (0.013316373175210)	0.004691386685937 (0.001109896065897)	0.001567472233727 (0.000647470628588)
n_2=100	LPK	0.060879625970763 (0.008146681734796)	0.003609688033503 (0.000535546820757)	0.000251262794899 (0.000072742546539)
	PS	0.063468397854826 (0.008759500834556)	0.005002162347441 (0.000868800388927)	0.001549836084585 (0.000483775943162)
n_3=150	LPK	0.061542042562676 (0.007857340633162)	0.003703242040865 (0.000496162046133)	0.000246362645900 (0.000050365665839)
	PS	0.063539045238946 (0.007871095534113)	0.005009803968250 (0.000778462437048)	0.001470687475140 (0.000365939122069)

جدول (٥) يوضح معدل قيم معيار (MSE) ولجميع حجوم العينات ومستويات الانحراف المعياري للطرائق اللامعلمية المستخدمة في تقدير أنموذج المعاملات المتغيرة الخامس

حجوم العينات	الطرائق اللامعلميه	مستويات الانحراف المعياري لحد الخطأ σ=1/2	σ=1/4	σ=1/8
n_1=50	LPK	0.058866914405853 (0.012700183512511)	0.003517485859272 (0.000784395723644)	0.000098429346371 (0.000016179081966)
	PS	0.062213701802166 (0.013462222335561)	0.005192456126782 (0.001039781515305)	0.001671430195260 (0.000310482993395)
n_2=100	LPK	0.060454289406132 (0.008496145301877)	0.003625519855759 (0.000511697947306)	0.000097763872490 (0.000013750173817)
	PS	0.063165121157419 (0.008937579105352)	0.005225459854829 (0.000671703977670)	0.001641504123915 (0.000225883672491)
n_3=150	LPK	0.061492130799033 (0.007276389872775)	0.003548350550404 (0.000417107774795)	0.000100487286035 (0.000011764707673)
	PS	0.063658020662998 (0.007466035078427)	0.005170486004449 (0.000673592308483)	0.001641689205991 (0.000184039063260)

٤. الاستنتاجات والتوصيات

١. من خلال مقارنة طرائق التقدير اللامعلمية المعتمدة في تقدير أنموذج المعاملات المتغيرة وجد ان افضل طريقة كانت طريقة النواة لمتعدد الحدود الموضعي (LPK) اذ حققت أقل قيمة لمعيار متوسط مربعات الخطأ (MSE) ولجميع حجوم العينات الثلاثة ومستويات الانحراف المعياري وكذلك لجميع الأنموذجات الخمسة المعتمدة في تجارب المحاكاة.

٢. لوحظ ان قيم الانحراف المعياري لمتوسط مربعات الخطأ لكافة تكرارات تجارب المحاكاة لكل حالة من حجوم العينات الثلاثة ومستويات الانحراف المعياري الثلاثة وكذلك للأنموذجات الخمسة المعتمدة في المحاكاة، كانت صغيرة ومتقاربة مما يشير الى تجانس قيم متوسط مربعات الخطأ لكل حالة محاكاة.

٣. وجد ان كلما زاد حجم العينة ادى الى زيادة قيم معدل متوسط مربعات الخطأ وتقليل قيم الانحراف المعياري لمتوسط مربعات الخطأ مما يشير الى تحسين نتائج تقدير أنموذج المعاملات المتغيرة.

٤. تبين انهُ كلما قلت قيمة الانحراف المعياري المستخدم في توليد قيم حدود الخطأ العشوائي لأنموذج المعاملات المتغيرة أدى الى تقليل قيمة متوسط مربعات الخطأ لتلك الأنموذجات.

٥. نوصي بأستخدام طريقة النواة لمتعدد الحدود الموضعي (LPK) لتقدير انموذج المعاملات المتغيرة (VCM)

٦. نوصي بمقارنة طرائق لامعلمية اخرى مثل طريقة الشريحة الجزائية التكعيبية وطريقة المويجة لتقدير أنموذج المعاملات المتغيرة (VCM).

٥. المصادر

١. خمو ، خلود يوسف (٢٠٠٤) " مقارنة أساليب بيز مع طرائق أخرى لتقدير منحنى الانحدار اللامعلمي " أطروحة دكتوراه في الإحصاء ، كلية الإدارة والاقتصاد ، جامعة بغداد.

2. Chan, S. C., Zhang, Z. G.(2009) "Local polynomial modeling and bandwidth selection for time-varying linear models", IEEE

3. Chan, S. C., Zhang, Z. G.(2011) "Local polynomial modeling and variable bandwidth selection for time-varying linear system", IEEE Vol 60, No. 3

4. Denison, D.G.T., Mallick, B.K. & Smith, A.F.M., (1998),"Automatic Bayesian Curve Fitting", J.R. Statist. Soc. B, Vol. 60, pp.333-350.

5. Hastie, T. J. and Tibshirani, R. J. (1993). "Varying-coefficient models". J. Roy. Statist. Soc. B. 55 pp.757–796.

6. McCullagh,P.and Nelder,J.A.;(2001);"Generalized linear models"; Chapman and Hall, London.

7. Ruppert, D., Sheather, S.J. & Wand, M.P., (1995), "An effective Bandwidth Selector for Local Least Squares Regression", J.Amer. Statist. Assoc.,Vol. 90, pp.1257-1270.

8. Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). "Semiparametric Regression." Cambridge University Press.

9. Smith, M. & Kohn, R., (1996), "Nonparametric Regression using Bayesian variable Selection", Journal of Econometrics, Vol. 75, pp.317-343.

10. Wu, H. and Zhang, J., (2006), "Nonparametric regression methods for longitudinal data analysis: Mixed-Effects modeling approaches", John Wiley & Sons, New Jersey.

11. Zhang, Chongshan, Yin, Xiangrong, (2012) "Trending time-varying coefficient market models" Quantitative Finance Volume 12, Issue 10.

Comparison between the Local Polynomial Kernel and Penalized Spline to Estimating Varying Coefficient Model

Abstract

Analysis the economic and financial phenomena and other requires to build the appropriate model, which represents the causal relations between factors. The operation building of the model depends on Imaging conditions and factors surrounding an in mathematical formula and the Researchers target to build that formula appropriately. Classical linear regression models are an important statistical tool, but used in a limited way, where is assumed that the relationship between the variables illustrations and response variables identifiable. To expand the representation of relationships between variables that represent the phenomenon under discussion we used Varying Coefficient Models (VCM) as it assumes the effects of variables illustrations be variable adoption of another explanatory variable. These structural avoided what is known as Curse of Dimensionality, which appears when we used nonparametric methods in estimation. We estimate the varying coefficients by using nonparametric methods which is the Local Polynomial Kernel (LPK) and Penalized Spline (PS), and by using simulation technique for comparison we found that the LPK method is the best.

Keywords: Varying Coefficient Model, Local Polynomial Kernel, Penalized Spline, Generalized Cross Validation.