تقدير المرحلتين الحصين وتقريب متعدد الحدود المحلي (kernel) لأنموذج المعاملات المتغيرة زمنياً مع بيانات طولية متزنة[1]

أ. د. ظافر حسين رشيد
الباحث علي سيف الدين عبد الحافظ
كلية الادارة والاقتصاد- جامعة بغداد
قسم الاحصاء

المستخلص

في هذا البحث تم أستعراض تقنية لا معلمية لتقدير دوال المعاملات المتغيرة زمنياً للبيانات الطولية المتزنة والتي تتصف بكون المشاهدات يتم الحصول عليها من n من القطاعات المستقلة كل واحد منها يكرر قياسها خلال مجموعة نقاط زمن محددة m ، وعلى الرغم من أن القياسات هي مستقلة بين القطاعات المختلفة الاّ أنها على الأغلب تكون مرتبطة داخل كل قطاع، والتقنية المستعملة هي تقنية متعدد الحدود الخطي المحلي LLPK ، ولتجاوز مشكلة تعدد الابعاد والحسابات الكثيفة، تم استعمال طريقة المرحلتين لتقدير دوال المعاملات بأستعمال التقنية السابقة، ولكون طريقة المرحلتين تعتمد بالتقدير على طريقة المربعات الصغرى ols التقليدية، وهي حساسة لوجود الشواذ بالبيانات أو تلوث الخطأ، لذا تم اقتراح أستعمال اساليب حصينة مثل LAD و M لتحصين طريقة المرحلتين اتجاه الشواذ او تلوث الخطا، وتمت صياغة تجارب محاكاة في هذا البحث والتحقق من أداء الطرائق التقليدية والحصينة لتقنية LLPK بأستعمال معيارين ولمختلف حجوم العينة ومستويات التباين.

<u>المصطلحات الاساسية للبحث/</u>
المعاملات المتغيرة زمنياً، تقدير المرحلتين، تقديرات M و LAD الحصينة، تمهيد متعدد الحدود المحلي، البيانات الطولية.

---

[1] بحث مستل من رسالة دكتوراه

مجلة العلوم الاقتصادية والإدارية
المجلد 19
العدد 70
الصفحات 297-324

1. المقدمة :

غالباً ما يتم الحصول على المشاهدات في الدراسات الطويلة من n من القطاعات المستقلة كل واحداً منها يكرر قياسها خلال مجموعة نقاط زمن محددة ، وغالباً ما يتركز أهتمام هذه الدراسات على تقييم آثار الزمن (t) وكذلك مجموعة المتغيرات المستقلة $\chi_r(t) \quad r = 1,2,...,d$, على نتيجة المتغير المعتمد $y(t)$ ، نفرض أن (t_j) تمثل الزمن للقياسات للقطاع i^{th} ، وأن y_{ij} و χ_{ij} تمثل مشاهدات القطاع i للمتغير المعتمد والمستقل عند الزمن t_j على التوالي، فأن مجموعة المشاهدات الطولية تعطى كالاتي :-

$$\{(t_j , y_{ij} , \chi_{ij})\}$$

على الرغم من أن القياسات هي مستقلة بين القطاعات المختلفة الاّ أنها على الأغلب تكون مرتبطة داخل كل قطاع.

البيانات الطولية شائعة الأستعمال في الدراسات الطبية، الوبائية، الأقتصادية، المالية ، ... وغيرها ، والتحليل الأحصائي مع هكذا نوع من البيانات مهتم بنمذجة منحنى المتوسط لـ $y(t)$ والتأثيرات للمتغيرات المستقلة في $y(t)$ ، وتطوير التقدير وأجراءات الاستدلال ، وتحت أطار النماذج المعلمية مثل النماذج الخطية ، وقد درست نظريات وطرائق التقدير للمعاملات والاستدلالات بصورة موسعة، ولكن المشكلة في النماذج الخطية أنها غير واقعية في الكثير من التطبيقات،عندها بدأ التفكير بتقنيات أخرى مثل التحليل اللامعلمي.

إن النماذج اللامعلمية للبيانات الطولية لا تقدم افتراضات بشأن تحديد الأنموذج ولكنها قد تخفق في ما يسمى مشكلة البعدية (Curse of Dimensionality) مما يجعل الطرائق اللامعلمية القياسية عاجزة عملياً عندما تكون المتغيرات المستقلة ذات أبعاد عالية ، وتكمن الصعوبة في حدود التفاعل (Interaction) بين الأبعاد المستعملة في الظاهرة، حيث لا يمكن تمثيلها بسهولة، ولحل مشكلة البعدية ظهر نهج بديل يتمثل في تخفيض القيود المفروضة على النماذج المعلمية التقليدية واستكشاف الهيكل الخفي (Hidden Structure) ومنها نماذج المعاملات المتغيرة (Varying Coefficient Models)، وما يميزها فضلا عن تمكنها من حل مشكلة تعدد الابعاد هو أمكانية استعمالها لاستكشاف الملامح (الديناميكية) في البيانات ذات الأبعاد العالية، وقد تم تطبيقها بنجاح في الانحدار اللامعلمي متعدد الأبعاد، النماذج الخطية العامة، نماذج السلاسل الزمنية غير الخطية، البيانات الطولية، بيانات البقاء والبيانات الاقتصادية والمالية،... وغيرها.

وتعد نماذج المعاملات المتغيرة البديل الطبيعي للنماذج المضافة (Additive Models)، اذ إن النماذج المضافة تعتمد على حل مشكلة تعدد الأبعاد بتجاهل حدود التفاعل بين الفضاءات (المتغيرات) والتركيز على المتغيرات المعبرة عن التأثيرات الرئيسية (Main Effects) فقط واعتبارها دوال ممهدة، اما نماذج المعاملات المتغيرة فأنها ستعمد على ايجاد نوع من أنواع التفاعل بين الفضاءات (المتغيرات)، اذ انها تكون خطية في المتغيرات ولكن معاملات هذه المتغيرات يسمح لها بالتغير تمهيدياً وهو ما قد يسمى تعديلات التأثير (Effect Modifiers) ، وأول من أستعمل المعاملات المتغيرة زمنياً كحلٍ لمشكلة تعدد الابعاد للأنموذج اللامعلمي للبيانات الطولية وكذلك استكشاف الديناميكية لها هم ((Hoover et al (1998)[5] ، حيث أعتمدو على أنموذج المعاملات المتغيرة زمنياً التالي :-

$$Y(t) = X'(t)\,\beta(t) + e(t) \qquad \ldots\ldots..(1)$$

اذ ان $e(t)$ عملية عشوائية بمتوسط (0) ، ولجميع قيم t فأن $e(t)$ و $X(t)$ مستقلان، إن الأنموذج المذكور آنفاً يبين أن الزمن (t) يغير المعاملات لـ $X(t)$ و $X(t) = (X_1(t), X_2(t), \ldots, X_d(t))'$ من خلال دوال غير محددة $\beta(t)$ اذ ان $\beta(t) = (\beta_1(t), \beta_2(t), \ldots, \beta_d(t))'$ هو متجة للدوال الممهدة خلال الزمن (t) ، واقترحوا أستعمال تقنيتين لامعلمية لتقدير $\beta(t)$ هما متعدد الحدود الخطي المحلي وشرائح التمهيد ، وأستعملوا خوارزمية (Backfitting) لأيجاد التقديرات، ولكن بأستعمال هذة الخوارزمية في البيانات الطولية ستظهر مشكلة من نوع اخر وهي كثافة العمليات الحسابية والجهد البرمجي ، للتغلب على هذه المشكلة يمكن أجراء تقدير المرحلتين المقترح من قبل ((Fan & Zhang (2000)[3] كبديل لأجل تقدير $\beta(t)$ في (1) ،أن تقديرات دوال المعاملات $\beta(t)$ بأسلوب المرحلتين يتم الحصول علية بأستعمال طريقة المربعات الصغرى ،وكما هو معلوم أن مقدرات المربعات الصغرى تمتلك بعض الخصائص الجيدة ، وخاصة عندما الخطأ العشوائي يتبع التوزيع الطبيعي ، ولكن المقدرات المعتمدة على المربعات الصغرى حساسة جداً الى الشواذ في البيانات أو عند تلوث (contamination) الخطأ، لذلك فأن طرائق التقدير الحصينة مطلوبة أكثر.

في هذا البحث سيتم الأعتماد على البيانات الطولية المتزنة (عندما تكون عدد القياسات للقطاعات متساوية) والمصاغة على وفق المعالم المختلفة زمنياً (TVC) لأيجاد تقديرات متعدد الحدود الخطي المحلي LLPK الحصين لدوال المعالم بطريقة المرحلتين، وبالأعتماد على أسلوبي LAD و M الحصينين، ومقارنتهما مع طرائق التقدير التقليدية عن طريق تجارب محاكاة بنسب تلويث مختلفة وحجوم عينة مختلفة ومستويات تباين مختلفة، ولأغراض الملاءمة والتعميم، تم عرض كل الصيغ والمعادلات بدلالة d من المتغيرات التوضيحية، على الرغم من إستعمال دراسة المحاكاة لحالة ثنائي المتغيرات (متغيرين فقط).

2. طريقة تقدير المرحلتين (Two-Step Estimation Method) (3) ، (4) ، (11)

لنفترض $t_j \;,\;\; j=1,2,\,....,m$ ، هي نقاط زمن محددة، حيث تم جمع البيانات ، أذ ان m تمثل عدد القياسات المكررة لكل قطاع، لأن هناك عدد من المشاهدات التي جمعت في الزمن t_j ، فمن الممكن لهذا الثابت t_j استعمال البيانات المجمعة هناك لمطابقة أنموذج (1) والحصول على المقدرات الخام (raw estimates)

$$b(t_j) \;=\; \big(b_1(t_j)\,,\,....\,,\,b_d\,(t_j)\big)'$$

الى

$$\beta(t_j) \;=\; \big(\beta_1(t_j)\,,\,....\,,\,\beta_d\,(t_j)\big)'$$

هذه هي المرحلة الأولى، عادة المقدرات الخام هي غير ممهدة تحتاج الى تمهيد للحصول على المقدرات الممهدة لدوال المعاملات، لذلك في المرحلة الثانية لكل مركبة معطاة $r=1,2,\,....,d$ نطبق تقنية تمهيد الى البيانات $\{(b_r\,(t_j)\;,t_j\,)\;,\;\; j=1,2,\,....,m\}$.

إن مرحلة تقديرات التمهيد (Smoothing estimates) هذه حاسمة لأنها تعطي مقدرات تمهيدية لدوال معاملات التمهيد الاساسية، وإضافة الى ذلك فان مرحلة التمهيد عادةً ذات بعد واحد (one–dimensional) .

1.2 مرحلة التقديرات الخام (Raw Estimates Step)[(10)]،

في هذه المرحلة سيتم الاعتماد على هيكيلة خاصة لمجموعة البيانات الطولية حيث سنعيد هيكلة أنموذج (1) كالآتي :

نفترض $j=1,2,\ ....,m \quad t_j$ هي نقاط زمن محددة لمجموعة البيانات الطولية لكل نقطة زمن t_j ، لنفترض N_j هو مجموعة المراجع للقطاع الى جميع مشاهدات y_{ij} عند t_j .

نجمع كل χ_{ij} و y_{ij} التي تقابل المراجع للقطاعات في N_j ونشكل مصفوفة التصميم $\tilde{X}_j$ ومتجه الاستجابة $\tilde{Y}_j$ بالتتابع.

الجدير بالذكر ان البحث يعتمد على حالة البيانات المتزنة وبذلك فان N_j مجموعة المراجع للقطاع الى جميع مشاهدات y_{ij} عند t_j ستكون متساوية ولجميع القطاعات. عندئذ فان صيغة أنموذج (1) عندما البيانات تجمع عند الزمن t_j يتبع الأنموذج الخطي التالي :

$$\left.\begin{aligned} \tilde{Y}_i(t_j) = \tilde{X}_i(t_j)\,\beta(t_j) + \tilde{e}_i(t_j) \qquad &, i=1,2,\dots,n \\ &j=1,2,\dots,m \end{aligned}\right\} \quad .....(2)$$

إذ إن :

$$\tilde{Y}_i(t_j) = \left[y_{11}, y_{12}, \dots, y_{1m}\ ,\ y_{21}, \dots, y_{2m}\ , \dots, y_{n1}, \dots, y_{nm}\right]'$$

$$\widetilde{X}_i(t_j) = \begin{bmatrix} X_{1,r} \\ X_{2,r} \\ . \\ . \\ . \\ X_{n,r} \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{1,2} & . & . & . & X_{1,d} \\ X_{2,1} & X_{2,2} & . & . & . & X_{2,d} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ X_{n,1} & X_{n,2} & . & . & . & X_{n,d} \end{bmatrix}$$

$$X_{i,r} = \begin{bmatrix} X'_{i1,r} & & & & & \\ & X'_{i2,r} & & & & \\ & & . & & & \\ & & & . & & \\ & & & & . & \\ & & & & & X'_{im,r} \end{bmatrix} \quad , \ i=1,2,\ldots,n \quad , \ r=1,2,\ldots,d$$

$(m*dm)$

$$\beta(t_j) = \left[\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,m}, \ldots, \beta_{d,1}, \ldots, \beta_{d,2}, \ldots, \beta_{d,m}\right]'$$

وإن $$\widetilde{e}_i(t_j) = \left[e_{11}, e_{12}, \ldots, e_{1m}, e_{21}, \ldots, e_{2m}, \ldots, e_{n1}, \ldots, e_{nm}\right]'$$

المتغيرات $\widetilde{X}_i(t_j)$ مستقلة عن الأخطاء العشوائية $\widetilde{e}_i(t_j)$ ، والأخطاء العشوائية $\widetilde{e}_i(t_j)$ هي عبارة عن عمليات عشوائية مشتركة بمتوسط (0) ودالة تباين كالآتي :

$$\sigma_{js} = \operatorname{cov}\left\{\widetilde{e}_i(j), \widetilde{e}_i(s)\right\}$$

1.1.2 مقدرات المربعات الصغرى العامة المقبولة مع اخطاء AR(1) [(6)]

(Feasible GLS Estimation With AR(1) Errors)

لتقدير معالم الأنموذج (2) وتحت افتراض هيكل الارتباطات للأخطاء يتبع AR(1) كالآتي :

$$\tilde{e}_i(t_j) = \rho\,\tilde{e}_i(t_{j-1}) + u_{ij} \qquad \ldots\ldots\ldots(3)$$

يمكننا تطبيق المربعات الصغرى العامة حيث مقدراتها ستكون كالآتي :

$$b_{GLS}(t_j) = \left(X_i'(t_j)\,\Omega^{-1} X_i(t_j) \right)^{-1} X_i'(t_j)\,\Omega^{-1}\,\tilde{Y}_i(t_j) \quad \ldots.(4)$$

ان المشكلة في مقدرات GLS بأنها تفترض مصفوفة التباين المشترك Ω معلومة، بعبارة آخرى إن ρ معلوم وهذا نادراً ماهو معلوم من الناحية العملية .

ولتجنب افتراض GLS علينا ايجاد تقدير متسق لـ $\hat{\Omega}$ أي $\hat{\rho}$ واستخدامه لايجاد تقدير لمعاملات الأنموذج (2). إن هذا المقدر يدعى مقدرات المربعات الصغرى العامة المقبولة (FGLS) والذي يمكن ايجاده حسب الخطوات التالية :

a. نجد اولاً مقدرات المربعات الصغرى الاعتيادية الى أنموذج (2) والذي سيكون كالآتي :

$$b_{OLS}(t_j) = \left(\tilde{X}_i'(t_j)\,\tilde{X}_i(t_j) \right)^{-1} \tilde{X}_i'(t_j)\,\tilde{Y}_i(t_j) \quad \ldots\ldots(5)$$

ثم نجد الاخطاء باستخدام مقدرات OLS إذ إن :

$$\tilde{e}_i(t_j) = \tilde{Y}_i(t_j) - \tilde{X}_i(t_j)\,b_{OLS}(t_j) \qquad \ldots\ldots\ldots(6)$$

b. وتحت افتراض الاخطاء هي عمليات عشوائية مشتركة فان مقدر ρ المشترك يمكن تقديره كالآتي :

$$\hat{\rho} = \frac{\sum_{i=1}^{n}\sum_{j=2}^{m} \tilde{e}_i(t_j)\,\tilde{e}_i(t_{j-1})}{\sum_{i=1}^{n}\sum_{j=1}^{m} \tilde{e}_i^{\,2}(t_j)} \quad \ldots\ldots\ldots(7)$$

c. إجراء تحويل للبيانات باستخدام (Prais – Winsten) transformation)).

$$\overset{*}{Y}_i(t_j) = \begin{bmatrix} \sqrt{1-\hat{\rho}^2}\,\tilde{y}_i(t_1) \\ \tilde{y}_i(t_2) - \hat{\rho}\, y_i(t_1) \\ \tilde{y}_i(t_3) - \hat{\rho}\, y_i(t_2) \\ . \\ . \\ \tilde{y}_i(t_m) - \hat{\rho}\,\tilde{y}_i(t_{m-1}) \end{bmatrix} \quad , i = 1,2,\ldots,m \quad \ldots\ldots\ldots(8)$$

$$\overset{*}{X}_i(t_j) = \begin{bmatrix} \sqrt{1-\hat{\rho}^2}\,\tilde{\chi}_i(t_1) \\ \tilde{\chi}_i(t_2) - \hat{\rho}\,\tilde{\chi}_i(t_1) \\ \tilde{\chi}_i(t_3) - \hat{\rho}\,\tilde{\chi}_i(t_2) \\ . \\ . \\ \tilde{\chi}_i(t_m) - \hat{\rho}\,\tilde{\chi}_i(t_{m-1}) \end{bmatrix} \quad , i = 1,2,\ldots,m \quad \ldots\ldots\ldots(9)$$

d. وبتطبيق المربعات الصغرى الاعتيادية على البيانات المحولة فاننا نحصل على مقدرات FGLS كالآتي :

$$b_{FGLS}(t_j) = \left(\overset{*\prime}{X}_i(t_j)\,\overset{*}{X}_i(t_j) \right)^{-1} \overset{*\prime}{X}_i(t_j)\,\overset{*}{Y}_i(t_j) \quad \ldots\ldots(10)$$

وان تقديرات كلاً من σ_u^2 و σ_e^2 سيكون كالآتي :

$$\hat{\sigma}^2 = \frac{\overset{*\prime}{e_i}(t_j)\overset{*}{e_i}(t_j)}{nm} = \frac{\left(\overset{*}{Y}_i(t_j) - \overset{*}{X}_i(t_j) b_{FGLS}\right)'\left(\overset{*}{Y}_i(t_j) - \overset{*}{X}_i(t_j) b_{FGLS}\right)}{nm} \quad ..(11)$$

$$\sigma_e^2 = \frac{\hat{\sigma}_u^2}{1-\hat{\rho}^2} \quad ....(12)$$

2.1.2 مقترح التقديرات الخام الحصينة

(Robust Raw Estimate)

اقترح أسلوب M الحصين اولاً من قبل (Huber) [7]،وتستند الفكرة ببساطة على تقليل بعض الدوال للأخطاء بدلاً عن مجموع المربعات لها، والمقدر الحصين يحدد بواسطة الاختيار لدالة وزن، وان اسلوب مقدرات M بحاجة الى بعض التوسيع لتطبيقها على البيانات الطولية المتزنة الموصوفة في أنموذج (2) والتي تحتوي على n من القطاعات و m من القياسات المكررة لكل قطاع.

إذ إن اهم ما يميز هذه البيانات هو ارتباطها ضمن القطاع، اي بمعنى ان الأخطاء مرتبطة، وهذا سينافي الافتراض لاسلوب مقدرات M وهو ان تكون الأخطاء غير مرتبطة، ولتلافي هذه المشكلة سيتم الاعتماد على البيانات المحولة في طريقة (Feasible GLS) .

لنفرض ان $\overset{*}{Y}_{ij} = \overset{*}{Y}_i(t_j)$ و $\overset{*}{X}_{ij} = \overset{*}{X}_i(t_j)$ ومتجه المعالم $\underline{\beta} = \beta(t_j)$ ، ولتوضيح هذا الاسلوب، ان المربعات الصغرى الاعتيادية للبيانات المحولة تخفض مجموع مربعات الخطأ الى اقل ما يمكن كالآتي :

$$\min \sum_{i=1}^{n}\sum_{j=1}^{m} \overset{*}{e}_{ij}^{2} = \min \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\overset{*}{Y}_{ij} - \overset{*}{X}_{ij}\,\beta\right)^2 \quad ....(13)$$

إذ إن :

$\overset{*}{Y}_{ij}$: المشاهدة j للقطاع i للمتغير المعتمد.

$\overset{*}{X}_{ij}$: مصفوفة التصميم $\overset{*}{X}$.

وإن :

$$\overset{*}{X}_{ijk} = \left[\overset{*}{X}_{ij1}, \ldots, \overset{*}{X}_{ijd}\right] \ , k = 1,2,\ldots,d$$

β : متجه معالم ذو بعد dm .

ان تقديرات M المطورة من قبل (Huber) تعتمد على فكرة ابدال مجموع مربعات الأخطاء $\overset{*}{e}_{ij}^{2}$ بدالة اخرى للأخطاء الهدف منها تقليل المقدار الآتي :

$$\sum_{i=1}^{n}\sum_{j=1}^{m} P\left(\overset{*}{Y}_{ij} - \overset{*}{X}_{ij}\,\beta\right)^{2} \qquad \ldots\ldots\ldots(14)$$

وإن P هي دالة محدبة متماثلة (symmetric convex function) ولتقليل المقدار اعلاه تؤخذ المشتقة بالنسبة الى متجه المعالم وجعلها مساوية الى الصفر كالآتي :

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \overset{*}{X}_{ij}\,\Psi\left(\overset{*}{Y}_{ij} - \overset{*}{X}_{ij}\,\beta\right) = 0 \qquad \ldots\ldots\ldots(15)$$

إذ إن :

Ψ : المشتقة الجزئية الى متجه المعالم β للدالة P وبهذا يكون هناك $(d\,m)$ من المعادلات غير الخطية والتي يمكن حلها بعدة طرائق منها طريقة المربعات الصغرى الموزونة التكرارية (IWLS) او طريقة نيوتن رافسون (N−R Method) او طريقة (Huber) وغيرها، ولكي تكون لمقدرات M خاصية (Scale Invariant) يتم إعادة كتابة الصيغة (14) كما يأتي :

$$\sum_{i=1}^{n}\sum_{j=1}^{m} P\left(\frac{\left(\overset{*}{Y}_{ij} - \overset{*}{X}_{ij}\,\beta\right)^{2}}{\sigma_e^{2}}\right) \qquad \ldots\ldots\ldots(16)$$

والصيغة (15) ستكون كما يلي :

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\overset{*}{X}_{ij}\,\Psi\left(\frac{\left(\overset{*}{Y}_{ij}-\overset{*}{X}_{ij}\,\beta\right)}{\sigma_e^2}\right)=0 \qquad ....(17)$$

ولإيجاد مقدرات M بالاعتماد على طريقة IWLS يتطلب حساب دالة الوزن وفيها يتم اعادة كتابة الصيغة (17) كما يلي :

$$\sum_{i=1}^{n}\sum_{j=1}^{m}W_{ij}\,\overset{*}{X}_{ij}\left(\frac{\left(\overset{*}{Y}_{ij}-\overset{*}{X}_{ij}\,\beta\right)}{\sigma_e^2}\right)=0 \qquad ....(18)$$

وبحل المعادلة اعلاه نحصل على تقديرات أسلوب M الحصين b_M باستعمال IWLS .

إذ إن :

$$b_M=\left(\overset{*\prime}{X}W\overset{*}{X}\right)^{-1}\overset{*\prime}{X}WY \qquad ....(19)$$

إذ إن :

b_M : متجه ذو بعد (dm)

W : مصفوفة اوزان قطرية ببعد ($nm*nm$) تحسب عناصرها كما يأتي :

$$W_{ij} = \frac{\Psi\left(\dfrac{\left(\overset{*}{Y}_{ij} - \overset{*}{X}_{ij}\,\beta\right)}{\hat{\sigma}_e}\right)}{\left(\dfrac{\left(\overset{*}{Y}_{ij} - \overset{*}{X}_{ij}\,\beta\right)}{\hat{\sigma}_e}\right)} \qquad \ldots\ldots\ldots(20)$$

إذ إن :

$$\hat{\sigma}_e = 1.483\left[Median \left| \overset{*}{e}_{ij} - Median(\overset{*}{e}_{ij}) \right| \right]$$

وتم استعمال دالة P ومشتقتها Ψ التالية :

دالة (Andrews)

$$P\left(\overset{*}{e}_{ij}\right) = \begin{cases} A^2\left[1-\cos\left(\overset{*}{e}_{ij} \setminus A\right)\right] & \left|\overset{*}{e}_{ij}\right| \le A\pi \\ 2A^2 & \left|\overset{*}{e}_{ij}\right| > A\pi \end{cases}$$

$$\Psi\left(\overset{*}{e}_{ij}\right) = \begin{cases} A\sin\left(\overset{*}{e}_{ij} \setminus A\right) & \left|\overset{*}{e}_{ij}\right| \le A\pi \\ 0 & \left|\overset{*}{e}_{ij}\right| > A\pi \end{cases}$$

$$A = 1.339$$

أن أسلوب M الحصين هو ليس أفضل أسلوب حصانة ، وهذا يجعل مقدراتها ليست بالضرورة هي الأفضل، لذلك نلجأ الى اساليب حصينة أخرى اتجاه الشواذ ومن بين هذه الاساليب هوالانحرافات المطلقة الصغرى ((Least Absolute Deviations (LAD)) المقترحة من قبل (Schlossmacher)[9]، حيث سيتم توسيع هذا الاسلوب لتطبيقه في البيانات الطولية المتزنة المصاغة في أنموذج (2) وبالاعتماد على البيانات المحولة في طريقة FGLS لتفادي مشكلة ارتباط الاخطاء كما مر سابقاً في اسلوب M .

ولتوضيح فكرة هذا الاسلوب لنفرض ان القيم المطلقة للاخطاء يمكن وصفها كالآتي :

$$L = \sum_{i=1}^{n}\sum_{j=1}^{m}\left|\overset{*}{e}_{ij}\right| \qquad \text{.........}(21)$$

إذ إن :

$$\overset{*}{e}_{ij} = \overset{*}{Y}_{ij} - \overset{*}{X}_{ij}\,\beta$$

$\overset{*}{Y}$ ، $\overset{*}{X}$ ، β كما تم تعريفها سابقاً.

ولايجاد مقدرات LAD بالاعتماد على دالة وزن وطريقة IWLS وبتقليل الصيغة التالية :

$$L = \sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij}\, e_{ij}^{*(2)} \qquad \text{.........}(22)$$

فان تقديرات LAD ستكون كالاتي :

$$b_{LAD} = \left(\overset{*}{X}{}'W\overset{*}{X}\right)^{-1}\overset{*}{X}{}'WY \qquad \text{.........}(23)$$

أذ إن :

b_{LAD} : متجه ذو بعد (dm)

W : مصفوفة اوزان قطرية ببعد $(dm*dm)$ تحسب عناصرها كالآتي :

$$W_{ij} = \frac{1}{\left|\overset{*}{e}_{ij}\right|} \qquad \text{.........}(24)$$

2.2 مرحلة تحسين او تمهيد المقدرات الخام [3]

(Refining Or Smoothing The Raw Estimates)

الطريق الطبيعي لتحسين المقدرات الخام هي تمهيدها من خلال الزمن، وببساطة نحن نمهد البيانات الكاذبة التالية $\{(b_r(t_j), t_j),\ j=1,2,\ldots,m,\ r=1,2,\ldots,d\}$ للحصول على تمهيد لدوال المعاملات $\beta_r(t)$ عن طريق احدى تقنيات التمهيد المعروفة، وقبل البدء بهذه المرحلة فان مركبات دوال المعالم المقدرة في المرحلة الاولى نحصل عليها كالآتي :

ولـ $r=1,2,\ldots,d$ نفرض أن :

$b_r(t_j)$ هي المركبة رقم r لتقديرات المرحلة الاولى $b(t_j)$ عندئذ

$$b_r(t) = L_r' \left(\overset{*\prime}{X} \overset{*}{X} \right)^{-1} \overset{*\prime}{X} \overset{*}{Y} \qquad \ldots\ldots\ldots(25)$$

إذ إن :

L_r : يعرف كمتجه وحدة ذو بعد (dm) المدخل رقم r هو (1) والباقي (0).

ومن المعلوم إن

$$E\left[b_r(t)\right] = \beta_r(t) \qquad \ldots\ldots\ldots(26)$$

وكذلك

$$\text{var}-\text{cov}\left[b_r(t)\right] = \sigma^2_{\overset{*}{e}} L_r' \left(\overset{*\prime}{X} \overset{*}{X} \right)^{-1} L_r \qquad \ldots\ldots\ldots(27)$$

وبذلك سيصبح التمهيد عند كل مركبة (r) ، حيث سنستعمل تقنية تمهيد متعدد الحدود الخطي المحلي LLPK، ولكن للبيانات التالية :

$$\left(b_r(t_j),\ t_j\right),\ j=1,2,\ldots,m$$

إذ إن :

t_j : هي نقاط زمن التصميم.

$b_r(t_j)$: هي الاستجابة عند نقاط زمن التصميم، مع التأكيد بانها اي تقدير مستخرج من المرحلة الاولى.

ان أنموذج الانحدار اللامعلمي البسيط للبيانات السابقة سيكون كالآتي :

$$b_r(t_j) = f\left(t_j\right) + \varepsilon_j \quad , \; j = 1,2,\ldots,m \quad \ldots\ldots\ldots(28)$$

ونحن نريد تقدير الدالة الممهدة $f(t_j)$.

إذ إن :

ε_j : هي اخطاء القياسات التي لايمكن شرحها بواسطة دالة الانحدار $f(t_j)$.

رياضياً $f(t)$ هي التوقع الشرطي لـ $b_r(t_j)\backslash t_j = t$ أي

$$f(t) = E\left(b_r(t_j)\backslash t_j = t\right)$$

1.2.2 ممهد Kernel متعدد الحدود الخطي الداخلي[2]

(Local Linear Polynomial Kernel Smoother (LLPK))

لايجاد ممهد (Kernel) متعدد الحدود الخطي المحلي للأنموذج (28) نفرض t_0 هي نقطة زمن ثابتة محددة ونريد تقدير $f(t_j)$ ، ونفرض ان $f(t)$ تمتلك $(P+1)$ من المشتقات المستمرة لبعض القيم الصحيحة $(P > 0)$ ، وهنا $(P = 1)$ وهي درجة متعدد الحدود الداخلي، وباستعمال متسلسلة تايلر، التقريب الخطي المحلي سيكون كالآتي :

$$f(t) \approx f(t_0) + (t - t_0)\, f^{(1)}(t_0)$$

وان

$$\tilde{\beta}_V = f^{(V)}(t_0) \big/ V! \qquad , V = 0,1$$

نفرض ان $\hat{\tilde{\beta}}_V \; , V = 0,1$ يقلل صيغة المربعات الصغرى الموزونة WLS التالية :

$$\sum_{j=1}^{m}\left\{b_r(t_j) - \left[\tilde{\beta}_0 + (t_j - t_0)\tilde{\beta}_1\right]\right\}^2 K_h(t_j - t_0) \qquad \ldots\ldots\ldots(29)$$

إذ إن :

$K_h = {K(./h)}\big/{h}$: يتم الحصول عليها عن طريق دالة (Kernel) . $K($ مع ثابت $(h > 0)$.

h : تسمى معلمة التمهيد او عرض الحزمة.

ان عرض الحزمة h يستعمل أساساً لتحديد أقرب تجاور محلي إذ إن :

$$I_h(t_0) = [t_0 - h \; , \; t_0 + h] \ldots\ldots\ldots(30)$$

ودالة (Kernel) تحدد كيف المشاهدات داخل $I_n(t_0)$ تشارك في المطابقة عند (t_0) ، واحدى اشهر دوال (Kernel) شيوعاً هي دالة الكثافة الاحتمالية الطبيعية المتماثلة الآتية:

$$K\left(\frac{t_i - t_0}{h}\right) = {\exp\left\{-\frac{\left(\frac{t_i - t_0}{h}\right)}{2}\right\}}\Big/{\sqrt{2\pi}} \qquad \ldots\ldots\ldots(31)$$

وان

$$\hat{f}_h^{(V)}(t_0) = V!\,\hat{\tilde{\beta}}_V \quad , V = 0,1 \qquad \ldots\ldots\ldots(32)$$

وللتعبير عن $\hat{f}_h^{(V)}(t_0)$ من المفيد وضعها بصيغة المصفوفات كالآتي :

$$b_r = \left[b_{r,1}, b_{r,2}, \ldots, b_{r,m}\right]'$$

متجه ذو بعد (m) .

$$X = \begin{bmatrix} 1 & (t_1 - t_0) \\ 1 & (t_2 - t_0) \\ . & . \\ . & . \\ . & . \\ 1 & (t_m - t_0) \end{bmatrix} \qquad \ldots\ldots\ldots(33)$$

مصفوفة ببعد $(m*2)$.

$$\tilde{\beta} = \left[\tilde{\beta}_0 \ , \ \tilde{\beta}_1\right]'$$

متجه ببعد (2).

$$W = diag\left(K_h(t_1 - t_0) \ , \ \ldots \ , \ K_h(t_m - t_0)\right)$$

مصفوفة أوزان قطرية ببعد $(m*m)$.

ولايجاد مطابقة LLPK عند (t_0) وباستخدام WLS فان

$$(b_r - X\,\tilde{\beta})'W(b_r - X\,\tilde{\beta})$$

وان قيمة المطابقة هي

$$\hat{f}_h^{(V)}(t_0) = V!\, e'_{V+1}\ S_m^{-1}\ T_m\ b_r \qquad , V = 0,1$$

e_{V+1} : يعرف كمتجه وحدة ببعد $(m*m)$ ولـ $(2*1)$ من المدخلات (1) والبقية (0).

وان

$$S_m = X'W\,X$$

$$T_m = X'W$$

وبذلك يصبح

$$\hat{f}_h(t_0) = e'_2\ S_m^{-1}\ T_m\ b_r \qquad \ldots\ldots\ldots(34)$$

عندما $\overset{*}{b}_r = \hat{f}_h(t_j)$ هو قيمة المطابقة لـ $f(t_j)$ بواسطة الصيغة (34) نحصل على

$$\overset{*}{b}_r = \hat{f}_h(t_j) = a'\left(t_j\right) b_r \qquad \ldots\ldots\ldots(35)$$

إذ إن :

$a(t_j)$: هي $e_2' \, S_m^{-1} \, T_m$ بعد ابدال (t_0) بـ (t_j) .

افترض ان

$$\overset{*}{b}_{r,h} = \left[\overset{*}{b}_{r,1}, \overset{*}{b}_{r,2}, \ldots, \overset{*}{b}_{r,m} \right]'$$

ويعرف بانه قيم المطابقة عند جميع نقاط الزمن المحددة عندئذ

$$\overset{*}{b}_{r,h} = A_h \, b_r \qquad \ldots\ldots\ldots(36)$$

إذ إن :

A_h : هي مصفوفة قطرية تعرف كالآتي :

$$A_h = diag\left(a(t_1) \, , \, a(t_2) \, , \ldots , \, a(t_m) \right)$$

2.2.2 الممهد الخطي وأختيار معلمة التمهيد(1)

(Linear Smoother and Smoothing Parameter Selection)

الممهدات الخطية تعبر عن المطابقة لمتجه الاستجابة $\overset{*}{b}_r$ عند نقاط الزمن المحددة كتراكيب خطية (Linear Combinations) الى متجه الاستجابة b_r ، وبشكل أكثر تحديداً الممهد $\hat{f}_P(t)$ يسمى ممهد خطي إذا أستطعنا أن نجعل المطابقة لمتجه الاستجابة $\overset{*}{b}_r$ ان يكون مرتبط مع متجه الاستجابة b_r عن طريق الصيغة البسيطة التالية :

$$\overset{*}{b}_r = A_P \, b_r \qquad \ldots\ldots\ldots(37)$$

إذ إن :

$$\overset{*}{b}_r = \left[\overset{*}{b}_{r,1}, \overset{*}{b}_{r,2}, \ldots, \overset{*}{b}_{r,m} \right]'$$

مع

$$\overset{*}{b}_{r,j} = \hat{f}_P(t_j)$$

وإن :

A_P : هي مصفوفة التمهيد بدرجة $(m*m)$ ،وتحدد بواسطة الممهد مع معلمة تمهيد P .

إن مصفوفة التمهيد A_P غالباً تعتمد على نقاط زمن التصميم ولكن يجب أن تكون غير معتمدة على متجه الاستجابة b_r ، وهذا يعني ان :

$$E\left(\overset{*}{b}_r \backslash t_1, t_2, \ldots, t_m \right) = A_P\, E\left(b_r \backslash t_1, t_2, \ldots, t_m \right) \quad \ldots\ldots(38)$$

$$\operatorname{cov}\left(\overset{*}{b}_r \backslash t_1, t_2, \ldots, t_m \right) = A_P \operatorname{cov}\left(b_r \backslash t_1, t_2, \ldots, t_m \right) A_p' \quad \ldots\ldots(39)$$

وباستعمال الممهدات الخطية تم تطوير العديد من طرائق اختيار معلمة التمهيد وبشكل موحدة الى الممهدات الخطية $\hat{f}_P(t)$ المعرفة في (37) الاختيار الجيد يحول عادةً لتحديد معلمة تمهيد جيدة (P) .

إن احد أفضل طرائق اختيار معلمة التمهيد وأكثرها انتشاراً هو (GCV) – (Generalized Cross Validation) ، ولاختيار معلمة التمهيد (P) للممهد الخطي $\hat{f}_P(t)$ في (37) وللأنموذج اللامعلمي في الصيغة (28) ، عن طريق تصغير المعيار الآتي :

$$GCV(P)=\frac{m^{-1}\sum_{j=1}^{m}\left[b_{r,j}-\overset{*}{b}_{r,j}\right]^2}{\left\{1-{t_r(A_P)}\big/{m}\right\}^2}=\frac{m^{-1}\,SSE_P}{\left(1-{df}\big/{m}\right)^2} \qquad \ldots\ldots\ldots(40)$$

ويتم اختيار معلمة التمهيد (P) التي تقابل اقل GCV .

3.2.2 مقترح التقديرات الممهدة الحصينة [8]

(Robust Smoothing Estimates)

ان طريقة تقدير الأنموذج اللامعلمي في (28) وهي طريقة متعدد الحدود الخطي المحلي حساسة إتجاه وجود قيم شاذة وجعلها أكثر صرامة بوجود الشواذ يمكن اجراء الاساليب الحصينة في القسم (2.1.2)، إذ إن الاخطاء للأنموذج اللامعلمي في (28) ستكون كالآتي :

$$e_j=b_r\left(t_j\right)-\overset{*}{b}_{r,j} \qquad , \; j=1,2,\ldots,m$$

لذلك سيكون من الواجب تحصين معيار GCV(P) كالآتي :

$$GCV_{Rob}(P)=\frac{m^{-1}\sum W_j\left(b_{r,j}-\overset{*}{b}_{r,j}\right)^2}{\left\{1-{t_r(A_P)}\big/{m}\right\}^2} \qquad \ldots\ldots\ldots(41)$$

إذ إن :

W_j : هي دالة وزن تحتوي على (m) من العناصر ويمكن حسابها دون الحاجة الى التوسيع لاسلوب M أو LAD في القسم (2.1.2) وفيما يلي وصف لدالة الاوزان.

في اسلوب M الحصين فان دالة الوزن تكون كالآتي :

$$W_j = \frac{\Psi\left(\frac{b_{r,j} - \overset{*}{b}_{r,j}}{\hat{\sigma}}\right)}{\left(\frac{b_{r,j} - \overset{*}{b}_{r,j}}{\hat{\sigma}}\right)} = \frac{\Psi(e_j)}{e_j} \qquad \text{.........(42)}$$

وبالاعتماد على دالة Andrews يمكننا ايجاد دالة الوزن W_j

وفي اسلوب LAD فان دالة الوزن W_j تكون كالآتي :

$$W_j = \frac{1}{|e_j|} = \frac{1}{\left|b_{r,j} - \overset{*}{b}_{r,j}\right|} \qquad \text{.........(43)}$$

3. المحاكاة Simulation

تم تنفيذ تجارب المحاكاة بإستعمال (n=10) ويمثل عدد القطاعـات مـع (m=5 , m=10 , m=15) وتمثل القياسات المتكررة لكل قطاع. وبذلك سيكون لدينا ثلاث حجوم للعينات (nm=50) و(nm=100) وأخيراً(*nm=150*) ، وللأنموذج التالي:

$$Y_{ij} = X_{1,i}(t_j)\,\beta_1(t_j) + X_{2,i}(t_j)\,\beta_2(t_j) + e_i(t_j) \quad , i=1,2,\dots,n \,;\, j=1,2,\dots,m$$

إذ إن :

$\beta_r(t_j)\ , r = 1\ , 2$: هي دوال معاملات ممهدة.

١

لمتغيران التوضيحيان $X_{1,i}\,(t_j)$ و $X_{2,i}\,(t_j)$ يتبعان التوزيع الطبيعي بمتوسط μ وتباين σ^2 ويتم توليدهما باستعمال طريقة (Box – Muller) وبصورة مستقلة لكل واحداً منهما،أما الاخطاء العشوائية فيتم توليدها كالاتي:

1. متجه الأخطاء $e_i\,(t_j)$ يتبع التوزيع الطبيعي بمتوسط (0) وتباين σ^2 يتم عن طريق استعمال طريقة (Box – Muller) ، وقد تم تناول ثلاث مستويات للتباين:

(High Noise) * تباين عالي

$$\sigma = \left(\frac{1}{2}\right) * Function\ Range$$

(Medium Noise)* تباين متوسط

$$\sigma = \left(\frac{1}{4}\right) * Function\ Range$$

(Low Noise)* تباين واطئ

$$\sigma = \left(\frac{1}{8}\right) * Function\ Range$$

إذ إن :

σ : هو الانحراف المعياري للخطأ e.

2. أما التوزيع الآخر للخطأ العشوائي $e_i\,(t_j)$ فهو التوزيع الملوث ويستعمل في حالة تلوث البيانات بقيم شاذه وبنسب 10% و 20% إذ تم توليد بيانات تتبع توزيع طبيعي بمتوسط (0) وتباين (σ^2=36).

أما دوال المعاملات فهي كلأتي :

$$\beta_1(t) = \sin(4\,\pi\,t)$$
$$\beta_2(t) = \cos(0.5\,\pi\,t)$$

أما المتغير المعتمد فيتم توليدا مباشراً من خلال أستعمال الأنموذج في دراسة المحاكاة.

ولتقييم إداء طرائق التقدير لمرحلة التقدير الخام ومرحلة التقدير الخام الحصينة وكذلك مرحلة التمهيد ومرحلة التمهيد الحصينة تم استعمال المعايير التالية:

1. متوسط الانحرافات المطلقة للأخطاء (Mean Absolute Deviation Error):

$$MADE = (dm)^{-1} \sum_{j=1}^{m} \sum_{r=1}^{2} \frac{|\beta_r(t_j) - \hat{\beta}_r(t_j)|}{range\ (\beta_r)}$$

2. متوسط مربعات الخطأ الموزون (Weighted Average Squared Error)

$$WASE = (dm)^{-1} \sum_{j=1}^{m} \sum_{r=1}^{2} \frac{\{\beta_r(t_j) - \hat{\beta}_r(t_j)\}^2}{range^2\ (\beta_r)}$$

إذ إن :

$range(\beta_r)$ هو المدى الى دالة $\beta_r(t_j)$.

وتم تكرار جميع تجارب المحاكاة $(Replicates = 200)$ مرة لكل تجربة وتم وضع جميع النتائج في الجداول من رقم (1) الى (5).

جدول (1) معايير تقدير Two Step لحالة عدم استعمال أساليب الحصانة في المرحلة الاولى والثانية، ولجميع حجوم العينة ولجميع مستويات التباين

ϵ	method	n	m	WASE			MADE		
				$\sigma = \frac{1}{2}$	$\sigma = \frac{1}{4}$	$\sigma = \frac{1}{8}$	$\sigma = \frac{1}{2}$	$\sigma = \frac{1}{4}$	$\sigma = \frac{1}{8}$
10 %	LLPK	10	5	1612.4	1203.59	2833.78	35.37	30.04	35.02
		10	10	269.28	193.95	277.09	14.35	12.72	13.02
		10	15	77.34	50.76	44.02	13.73	11.30	10.35
20 %	LLPK	10	5	994.58	1165.55	1265.65	23.45	26.10	28.36
		10	10	212.81	168.28	195.33	13.34	10.54	12.04
		10	15	65.48	47.08	46.83	12.59	10.68	10.48

جدول (2) معايير تقدير Two Step لحالة استعمال أساليب الحصانة LAD و M في المرحلة الاولى فقط ، ولجميع حجوم العينة ولجميع مستويات التباين

ϵ	method	n	m	Rob	WASE			MADE		
					$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$	$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$
10 %	LLPK	10	5	LAD	1609.46	1162.14	1472.46	34.82	28.79	27.50
				M	1495.20	1769.44	2100.70	34.49	32.91	32.78
		10	10	LAD	133.77	186.61	219.76	8.86	12.19	12.84
				M	169.76	226.50	208.90	11.74	12.90	11.77
		10	15	LAD	66.23	45.43	41.17	8.28	6.55	9.34
				M	63.41	49.54	40.17	6.75	7.83	8.43
20 %	LLPK	10	5	LAD	931.50	1174.78	1037.48	20.93	26.49	27.55
				M	2887.55	1660.75	1341.42	46.92	35.51	29.92
		10	10	LAD	208.46	105.07	145.33	12.34	9.02	10.57
				M	354.71	211.79	453.88	15.31	12.31	14.72
		10	15	LAD	47.23	37.52	39.89	6.19	5.74	7.91
				M	48.98	43.21	45.72	6.52	6.18	10.39

جدول (3) معايير تقدير Two Step لحالة استعمال أساليب الحصانة LAD و M في المرحلة الثانية فقط،ولجميع حجوم العينة ولجميع مستويات التباين

ϵ	method	n	m	Rob	WASE			MADE		
					$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$	$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$
10 %	LLPK	10	5	LAD	421.74	1248.64	1842.85	13.95	30.80	32.14
				M	2285.43	1010.16	2839.37	39.34	27.98	38.94
		10	10	LAD	165.17	158.66	179.74	11.61	11.23	12.54
				M	206.04	217.72	72.06	13.02	13.17	7.36
		10	15	LAD	175.89	28.74	39.14	13.83	8.21	9.40
				M	42.27	220.07	83.48	10.44	13.20	11.51
20 %	LLPK	10	5	LAD	720.15	1227.45	1158.75	22.67	29.61	24.82
				M	816.09	1338.69	1358.42	23.03	31.87	30.16
		10	10	LAD	47.35	214.74	194.81	9.59	13.03	12.00
				M	179.20	127.70	187.64	11.68	9.64	11.80
		10	15	LAD	51.30	36.47	142.22	11.70	9.65	12.48
				M	50.93	32.13	56.98	10.36	8.49	11.58

جدول (4) معايير تقدير Two Step لحالة استعمال أسلوب LAD في المرحلة الاولى، واستعمال اسلوبي LAD و M في المرحلة الثانية، ولجميع حجوم العينة ولجميع مستويات التباين

ϵ	method	n	m	Rob	WASE			MADE		
					$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$	$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$
10 %	LLPK	10	5	LAD	1278.31	1111.17	1704.12	35.09	27.80	33.92
				M	1855.61	1177.40	1834.85	38.71	28.31	33.50
		10	10	LAD	191.68	103.24	121.51	11.93	11.90	8.74
				M	72.20	97.82	178.08	6.78	8.03	11.47
		10	15	LAD	62.95	45.56	40.22	7.13	7.27	7.52
				M	72.24	47.44	39.85	7.07	7.38	6.81
20 %	LLPK	10	5	LAD	862.38	2875.13	1150.40	20.06	37.66	26.73
				M	927	2019.07	960.98	22.62	35.90	25.71
		10	10	LAD	178.91	166.86	242.98	11.34	10.39	13.28
				M	184.31	141.46	260.54	11.19	10.08	13.91
		10	15	LAD	62.53	43.92	44.85	7.34	7.68	8.45
				M	64.08	38.91	40.71	7.96	5.17	6.09

جدول (5) معايير تقدير Two Step لحالة استعمال أسلوب M في المرحلة الاولى، واستعمال اسلوبي LAD و M في المرحلة الثانية ، ولجميع حجوم العينة ولجميع مستويات التباين

ϵ	method	n	m	Rob	WASE			MADE		
					$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$	$\sigma = 1/2$	$\sigma = 1/4$	$\sigma = 1/8$
10 %	LLPK	10	5	LAD	1053.92	585.94	1007.65	23.93	21.22	25.72
				M	1606.01	1324.61	2795.92	29.04	32.29	34.01
		10	10	LAD	186.16	111.64	328.08	11.94	9.57	15.95
				M	268.55	161.71	170.57	14.11	11.31	10.70
		10	15	LAD	67.55	44.29	32.03	10.93	9.33	8.24
				M	59.77	48.32	23.02	9.39	10.05	7.86
20 %	LLPK	10	5	LAD	865.17	1159.86	1177.07	22.74	25.93	24.48
				M	2200.42	1167.20	3537.22	38.69	27.12	44.86
		10	10	LAD	210.69	108.68	150.45	12.43	9.69	11.15
				M	208.44	120.33	40.50	12.12	10.38	6.91
		10	15	LAD	60.61	46.91	39.98	9.46	7.16	6.61
				M	58.16	46.54	42.91	8.93	7.05	7.49

4. تحليل النتائج

لحالة عدم استعمال أسلوب الحصانة في المرحلة الاولى والثانية،ومن نتائج جدول (1) ، قيم معياري (MADE , WASE) سجلا أنخفاضاً ترافق مع زيادة حجم العينة (الزمن) ولكلا حالتي التلوث عدا حالة واحدة عند تلوث 20% وحجم عينة (n=10,m=15) و مستوى التباين σ=(1/4).

ولحالة استعمال أساليب الحصانة LAD و M في المرحلة الاولى فقط، ومن نتائج جدول (2) ، قيم معياري (MADE , WASE) سجلا أنخفاضاً ترافق مع زيادة حجم العينة (الزمن) ولكلا حالتي التلوث ، وبأخذ أفضل النتيجتين ومقارنتها مع نتائج جدول (1) نلاحظ أنخفاض قيم المعيارين بنسبة كبيرة عند تلوث 10% وبنسبة اقل عند تلوث 20% ما عدا اخفاق حالة واحدة عند تلوث 20% وحجم عينة (n=10,m=5) ومستوى تباين σ=(1/4) ،وأن أسلوب LAD تفوق قليلاً على أسلوب M عند تلوث 10% وتفوق بشكل كلي عند تلوث 20%.

ولحالة استعمال أساليب الحصانة LAD و M في المرحلة الثانية فقط، ومن نتائج جدول (3) ، قيم معياري (MADE , WASE) سجلا أنخفاضاً ترافق مع زيادة حجم العينة (الزمن) ولكلا حالتي التلوث الا اربع حالات اثنان عند تلوث 10% واثنان عند تلوث 20% وجميعهم عند حجم عينة (n=10,m=15) ، وبأخذ أفضل النتيجتين ومقارنتها مع نتائج جدول (1) نلاحظ أنخفاض قيم المعيارين بنسبة كبيرة عند تلوث 10% وبنسبة اقل عند تلوث 20% ما عدا اخفاق حالة واحدة عند تلوث 20% وحجم عينة (n=10,m=5) ومستوى تباين σ=(1/4) ، وأن اسلوب LAD تفوق على أسلوب M عند تلوث 10% ، وأسلوب M تفوق قليلاً على أسلوب LAD عند تلوث 20% .

ولحالة استعمال أسلوب الحصانة LAD في المرحلة الاولى وأستعمال أساليب الحصانة LAD و M في المرحلة الثانية ، ومن نتائج جدول (4) ، قيم معياري (MADE , WASE) سجلا أنخفاضاً ترافق مع زيادة حجم العينة (الزمن) ولكلا حالتي التلوث عدا حالة واحدة عند تلوث 10% وحجم عينة (n=10,m=15) ، وبأخذ أفضل النتيجتين ومقارنتها مع نتائج جدول (1) نلاحظ أنخفاض قيم المعيارين بنسة كبيرة عند تلوث 10% و 20% ما عدا أخفاق حالتين عند تلوث 20% الاولى عند حجم عينة (n=10,m=5) ومستوى تباين σ=(1/4) والثانية عند حجم عينة (n=10,m=10) ومستوى تباين σ=(1/8) ، وأن اسلوب M تفوق قليلاً على أسلوب LAD عند تلوث 10% وبشكل اكبر عند تلوث 20% .

ولحالة استعمال أسلوب الحصانة M في المرحلة الاولى وأستعمال أساليب الحصانة LAD و M في المرحلة الثانية، ومن نتائج جدول (5) ، قيم معياري (MADE , WASE) سجلا أنخفاضاً ترافق مع زيادة حجم العينة (الزمن) ولكلا حالتي التلوث عدا حالة واحدة عند تلوث 20% وحجم عينة (n=10,m=15) ومستوى تباين σ=(1/8) ، وبأخذ أفضل النتيجتين ومقارنتها مع نتائج جدول (1) نلاحظ أنخفاض قيم المعيارين بنسة كبيرة عند تلوث 10% و 20% ،وأن اسلوب LAD تفوق على أسلوب M عند تلوث 10% وبشكل اقل عند تلوث 20% .

ومن خلال متابعة الجداول من (1) الى (5) فأن أفضل النتائج لطريقة تقدير LLPK عند تلوث 10% هي أستعمال أسلوب الحصانة LAD في المرحلة الاولى مع أستعمال أسلوب الحصانة M في المرحلة الثانية، وعند تلوث 20% أستعمال أسلوب الحصانة LAD في المرحلة الاولى فقط، ولوحظ أنخفاض قيم المعيارين لحالات أستعمال اساليب الحصانة LAD و M مقارنة مع الطريقة التقليدية ولكلا حالتي التلوث ولمختلف حجوم العينة ومستويات التباين، وكذلك أنخفاض قيم المعيارين على الاغلب لمستوى التباين المتوسط $\sigma=(1/4)$ مقارنة مع مستوى التباين العالي $\sigma=(1/2)$ والواطئ $\sigma=(1/8)$ ، كما أظهر المعيارين (MADE و WASE) تطابق في تفضيل ومقارنة طرائق التقدير.

المصادر

1- Buja, A., Hastie, T. and Tibshirani, R., (1989), "Linear Smoothers and Additive Models", Annals of statistics, vol. 17, pp. 453-510.

2- Fan, J. and Gijbels, I., (1996), "Local Polynomial Modelling and its Applications", Chapman and Hall, London.

3- Fan, J. and Zhang, J., (2000), "Two – Step Estimation of Functional Linear Models with Applications to Longitudinal Data", Journal of Royal statistical society, vol. 62, no. 2, pp. 303-322.

4- Fan, J. and Zhang, W., (2008), "Statistical Methods with Varying Coefficient Models" Statistics and Interface, vol. 1, pp. 179-195.

5- Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L., (1998), " Non Parametric Smoothing Estimates of time - Varying Coefficient Models with Longitudinal Data", Biometrika, vol. 85, no. 4, pp. 809-822.

6- Hsiao, C., (2003), "Analysis of Panel Data", second edition, Cambridge University Press.

7- Huber, P. J., (1981), "Robust Statistics", John Wiley & Sons, New York.

8- Kovac, A., (2002), "Robust Nonparametric Regression and Modality", http://Maths.bris.ac.uk/~Maxak/

9-Schlossmacher, E. J., (1973), "An Iterative Technique for Absolute deviations curve fitting", JASA, vol. 68, no. 344, pp.857-859.

10- Senturk, D. and Muller, H. G. (2008), "Generalized varying coefficient models for longitudinal data", Biometrika, vol. 95, Iss.3, pp.653-666.

11- Wu, C. O., Tian, X. and Yu, J., (2010), "Non parametric estimation for time-varying transformation models with longitudinal data", Journal of non parametric statistics, vol. 22, no. 2, pp. 133-147.

Robust Two-Step Estimation and Approximation Local Polynomial Kernel For Time-Varying Coefficient Model With Balance Longitudinal Data

Abstract

In this research, the nonparametric technique has been presented to estimate the time-varying coefficients functions for the longitudinal balanced data that characterized by observations obtained through (n) from the independent subjects, each one of them is measured repeatedly by group of specific time points (m). Although the measurements are independent among the different subjects; they are mostly connected within each subject and the applied techniques is the *Local Linear kernel* LLPK technique. To avoid the problems of dimensionality, and thick computation, the two-steps method has been used to estimate the coefficients functions by using the two former technique. Since, the two-steps method depends, in estimation, on (OLS) method, which is sensitive for the existence of abnormality in data or contamination of error; robust methods have been proposed such as LAD & M to strengthen the two-steps method towards the abnormality and contamination of error. In this research imitating experiments have been performed, with verifying the performance of the traditional and robust methods for *Local Linear kernel* LLPK technique by using two criteria, for different sample sizes and disparity levels.

Keywords: Time varying coefficient;Two step estimation;Robust M&LAD estimation;Local polynomial smoothing;Longitudinal data.