Compare between simex and Quassi-likelihood methods in estimation of regression function in the presence of measurement error

مقارنة بين طريقتي السيمكس و شبه الإمكان في تقدير دالة الانحدار شبه المعلمية في ظل وجود خطأ القياس

أ.د. مناف يوسف حمود / كلية الإدارة و الاقتصاد / جامعة بغداد

الباحث / احمد مهدي جابر Gmail / ahmedalwasty01@gmail.com

OPEN ACCESS

P - ISSN 2518 - 5764
E - ISSN 2227 - 703X

Received:16/10/2018
Accepted: 18/11/2018

المستخلص

زاد اهتمام الباحثين في السنوات الاخيرة بنماذج الانحدار شبه المعلمية وذلك اذ يمكن من خلالها دمج نماذج الانحدار اللامعلمي والمعلمي في ان واحد ومن ثم تكوين انموذج انحدار يتميز بامكانية التعامل مع مشكلة تعدد الابعاد الموجودة في النماذج اللامعلمية والتي تحصل عند زيادة عدد المتغيرات التوضيحية الداخلة في التحليل و من ثم تناقص دقة التقدير. فضلا عن امتياز هذا النوع من النماذج بمرونة في المجال التطبيقي مقارنةً مع النماذج المعلمية التي تتقيد بشروط معينة كمعرفة توزيع الاخطاء او قد تكون النماذج المعلمية لا تمثل الظاهرة المدروسة تمثيلاً صحيحاً.

سنستعرض في هذا البحث طرائق شبه معلمية في تقدير دالة الانحدار في ظل وجود أخطاء قياس وهذه الطرائق هي طريقة السيمكس وطريقة شبه الإمكان المعدلة وستتم المقارنة بين هاتين الطريقتين بأستعمال معيار المقارنة معدل متوسط مربعات الخطأ $(MASE)$.

اذ تمت المقارنة من خلال استعمال أسلوب المحاكاة بأستعمال النماذج شبه المعلمية واحجام عينات وتباينات مختلفة للتحقق من أداء الطرائق بأستعمال معيار المقارنة المذكورة آنفاً حيث كانت اهم الاستنتاجات هي ان طريقة شبه الإمكان المعدلة تتفوق على طريقة السيمكس بغض النظر عن حجم العينة المستعملة وقيمة التباين الموضوعة .

المصطلحات الرئيسة للبحث / انموذج شبه معلمي – خطأ القياس – مقدر السيمكس – المتغيرات المساعدة – مقدر شبه الإمكان

*بحث مستل من رسالة ماجستير

1- المقدمة Intrduction

يعد الانحدار من الاساليب المهمة في التحليل الاحصائي والذي يهدف الى معرفة طبيعة العلاقة بين المتغير التابع ومتغير او مجموعة المتغيرات التوضيحية من خلال تكوين صيغة رياضية معينة تدعى بأنموذج الانحدار التي تقوم بتحديد طبيعة واتجاه العلاقة بين المتغيرات والتي تستخدم ايضاً في التنبؤ . ونظراً لأختلاف وتنوع الظواهر في الواقع العملي فقد وجد اكثر من نوع لنماذج الانحدار تبعاً لطبيعة الظاهرة المدروسة .
فالنوع الاول من النماذج يدعى بنماذج الانحدار المعلمي الذي يعتمد في صياغته مجموعة من المعالم المجهولة في النموذج والتي يتم تقديرها بأستعمال عدة طرائق منها المربعات الصغرى والامكان الاعظم وغيرها من الطرائق الا ان هذا النوع من النماذج لا يمكن استعماله دائماً نظراً لكونه يتقيد بشروط معينة لكي يتم استعماله او قد تكون هذه النماذج لا تعبر عن الظاهرة المدروسة نظراً لسلوك بعض المتغيرات الداخلة في عملية التحليل سلوكاً معلمياً والبعض الاخر سلوكاً لامعلمياً .
اما النوع الثاني فيدعى بنماذج الانحدار اللامعلمية و الذي لا يتقيد بشروط صارمة من حيث التوزيع الخاص بمتغير الاستجابة والخطأ ، ولا يتقيد بدالة معينة تفسر العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة و يمتاز هذا النموذج بمرونته العالية في التعامل مع البيانات و كذلك سهولة تفسير نتائجه الا ان هذا النوع من النماذج يعاني من مشكلة تعدد الابعاد والتي تحصل عند زيادة عدد المتغيرات التوضيحية في عملية التحليل مما يؤدي الى تناقص دقة التقدير .
اما النوع الاخير من نماذج الانحدار فيدعى بنماذج الانحدار شبه المعلميه والذي يعد ثمرة التكامل بين انموذجي الانحدار المعلمي واللامعلمي ويمتاز هذا الانموذج بالمرونة العالية في التطبيق بالمقارنة مع النماذج المعلميه وكذلك فأنه يتخلص من مشكلة تعدد الابعاد . بالرغم من ان نموذج الانحدار شبه المعلمي يمتاز بالعديد من مميزات النموذجين المتكون منهما الا انه ايضاً سوف يمتلك بعضاً من سلبيات هذين الانموذجين واحدى هذه السلبيات هي وجود مشكلة خطأ القياس التي تحدث عندما ينتهك الافتراض الاساسي في نموذج الانحدار اذ ان المتغيرات التوضيحية قد تم قياسها بدون وجود اخطاء قياس فبوجود هذه الاخطاء ينتج عنه تقديرات متحيزة وغير متسقة للمعلمة . ومن ثم لا يتم الحصول على انموذج يمثل الظاهرة قيد الدراسة تمثيلاً كاملاً .

2- هدف البحث

يهدف البحث الى الحصول على تقدير لدالة الانحدار شبه المعلمية في حالة وجود اخطاء في القياس في المتغيرات التوضيحية بأستعمال بعض طرائق التقدير والمقارنة فيما بينها للحصول على افضل مقدر .

3- الجانب النظري

3-1 خطأ القياس

يعرف خطأ القياس ببساطه على أنه الفرق بين قيمة متغير المشاهدة (الملاحظة) وقيمة المتغير المقاس بشكل صحيح ويحدث هذا الخطأ نتيجةً لوجود خلل في الجهاز المستعمل في عملية القياس او الغفلة والإهمال من قبل الشخص الذي يقوم بعملية القياس....الخ [6]
يعبر عن خطأ القياس بالصيغة الاتية :

$$U_i = X_i - W_i \qquad ---(1)$$

اذ ان :
: يمثل خطأ القياس . U_i
: تمثل القيمة المشاهدة (المقاسة) .X_i
: تمثل القيمة الحقيقية . W_i
[9] أن وجود خطأ القياس يؤدي الى .
1) تقديرات متحيزة و غير متسقة لمعلمة الانموذج
2) عدم اعطاء وصف صحيح للظاهرة قيد الدراسة

3- إخفاء ملامح البيانات مما يجعل عملية تحليل البيانات صعبة

3-2 الانموذج شبه المعلمي

تعرف النماذج شبه المعلمية (*Semiparametric Regression models*) بأنها تلك النماذج التي تستعمل للإشارة الى فئتين من النماذج ، الفئة الاولى من النماذج تتضمن معلمات مجهولة يتم تقديرها بأحدى طرائق تقدير الانحدار المعلمية على سبيل المثال المربعات الصغرى او الإمكان الأعظم ... الخ . والفئة الثانية من النماذج لا تتضمن المعلمات المجهولة ولكن يوجد دالة مجهولة والتي يمكن تقديرها بأحدى طرائق التقدير اللامعلمية . [3]

يحظى انموذج الانحدار شبه المعلمي (*Semiparametric Regression*) برواج واسع في الآونة الأخيرة ، وذلك لميزته في دمج نماذج الانحدار المعلميه مع نماذج الانحدار اللامعلميه في إن واحد ، تلك الميزة التي جعلت من الانموذج الجديد يتجاوز مشكلة الإبعاد في حالة النماذج اللامعلميه بالكامل ، كما وفر بيئة اكبر للتطبيق من تلك التي تكون في حالة النماذج المعلميه للانحدار كون الأخير قد يتأثر ببعض المتغيرات المستقلة (التوضيحية) والتي ليس لها توزيع معلمي معروف ، ونظراً لكون الانموذج شبه المعلمي يمتلك مميزات كلاً من الانموذجين المعلمي واللامعلمي لذا فأنه يعد انموذج وسيط بين النماذج المعلميه واللامعلمية.[2]

يمكن كتابة الانموذج شبه المعلمي بالصيغة الاتية :

$$Y = X\beta + g(t) + \varepsilon \qquad \dots(2)$$

اذ أن :

Y : يمثل متجه متغير الاستجابة او المتغير المعتمد وهو من الدرجة $n * 1$

X : يمثل مصفوفة المتغيرات التوضيحية من الدرجة $n * p$

β : يمثل متجه المعالم المجهولة من الدرجة $p * 1$

$g(t)$: يمثل دالة تمهيدية غير معرفة من الدرجة $n * 1$

ε : متجه الأخطاء العشوائية من الدرجة $n * 1$ وتكون مستقلة بمتوسط صفر وتباين σ^2

3-3 طرائق التقدير

3-3-1 طريقة السيمكس (*simex method*)

هي اسلوب للتقدير والحد من التحيز الناجم عن خطأ القياس وقد اقترحت هذه الطريقة من قبل كوك وستيفانسكي عام 1994 م لإيجاد تقدير كفوء الى β في ظل وجود اخطاء قياس في المتغيرات التوضيحية لانموذج الانحدار المعلمي وقد اكتسبت منذ ذلك الحين شعبية كبيرة بين الإحصائيين نظراً لبساطتها وعموميتها اذ بأستعمال هذه الطريقة يمكن الحصول تقريباً على مقدرات غير متحيزة وكفوءة للمعلمة بالنسبة للأنواع المختلفة من نماذج الانحدار. [11]

تتكون خوارزمية السيمكس بشكل عام من ثلاث خطوات هي خطوة المحاكاة وخطوة التقدير وخطوة الاستقراء وكما في ادناه : [7] [4] [8]

أفترض مجموعة من القيم $0 = \lambda 1 < \lambda 2 \cdots < \lambda J$

1- في خطوة المحاكاة ، يتم توليد اخطاء القياس بمصفوفة تغاير $\lambda_j \Sigma_u$ وتضاف الى بيانات W الاصلية ، و بذلك تنشأ مجموعة بيانات بتباينات خطأ القياس الاكبر بالتوالي . أن تباين خطأ القياس الكلي لِ jth من مجموعة بيانات هو

$$\Sigma_u + \lambda_j \Sigma_u = (1 + \lambda_j)\Sigma_u.$$

2- في خطوة التقدير يتم الحصول على تقديرات من كل مجموعة بيانات الملوثة المولدة، بأستعمال خوارزمية تقدير معينة في حالة عدم وجود خطأ قياس .

يتم تكرار خطوة المحاكاة وخطوة التقدير ومن ثم يتم حساب القيمة المتوقعة للتقدير لكل مستوى تلوث λj.

3- في خطوة الاستقراء يتم وضع هذه المعدلات مقابل قيم λ ويستعمل اسلوب الانحدار ،على سبيل المثال ، المربعات الصغرى متعددة الحدود لملائمة دالة الاستقراء الى تقديرات المتوسط الملوثة بالخطأ.

3-3-1-1 السيمكس عندما X منمذجة لامعلمياً [5]:

في هذه الحالة لدينا الاستجابة Y ، المتجه Z الذي يمثل المتغير التوضيحي المقاس بدون وجود خطأ قياس والمتجه X الذي يمثل المتغير التوضيحي المقاس بوجود خطأ قياس . وأن دالة الإمكان اللوغاريتمية هي $\mathcal{L}\{Y,Z,\theta(X),\beta\}$

اذ يتم إيجاد مقدر السيمكس كما يأتي

لكل $b=1,2,\dots,B$ حيث B تمثل عدد مرات تكرار خطوة المحاكاة والتقدير المذكورة انفاً، يتم مشاهدة

$$\boldsymbol{W}_{ib}(\boldsymbol{\lambda}) = \boldsymbol{W}_i + \boldsymbol{\lambda}^{1/2}\boldsymbol{U}_{ib}$$

اذ أن :

$$\boldsymbol{W}_i = \boldsymbol{X}_i + \boldsymbol{U}_i$$

$\boldsymbol{U}_{ib}$: هي متغيرات عشوائية يتم توليدها و تتبع التوزيع الطبيعي بمتوسط صفر و تباين Σ_u

ثم نطبق الإمكان لكل b ولكل λ حيث يتم استبدال X_i بالمتغير $W_{ib}(\lambda)$. افترض أن $f_w(x_0,\lambda)$ هي دالة الكثافة ل $\mathcal{L}_\beta W_b(\lambda)$. تعرف $\beta(\lambda)$ و $\theta\{x_0,\beta(\lambda),\lambda\}$ على انها الحلول للمعادلتين الاتيتين :

$$0 = E\big(\mathcal{L}_\beta\big[Y,Z,\theta\{W_b(\lambda),\beta(\lambda),\lambda\},\beta(\lambda)\big]\big); \qquad \dots(3)$$
$$0 = E\big(\mathcal{L}_\theta\big[Y,Z,\theta\{x_0,\beta(\lambda),\lambda\},\beta(\lambda)\big]\big|W_b(\lambda)=x_0\big)$$

في خوارزمية السيمكس ، لأي λ ولـ $b=1,2,\dots,B$ ، افترض أن تقدير المعلمة والدالة يرمز لها بالرمز $\hat{\beta}_b(\lambda)$ و $\hat{\theta}_b\big(\cdot,\hat{\beta}_b(\lambda),\lambda\big)$ على التوالي . فأن متوسط هذين المقدرين يكونان بالشكل الاتي :

$$\hat{\beta}(\lambda) = B^{-1}\sum_{b=1}^{B}\hat{\beta}_b(\lambda)$$

$$\hat{\theta}\big(\cdot,\hat{\beta}(\lambda),\lambda\big) = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b\big(\cdot,\hat{\beta}_b(\lambda),\lambda\big)$$

في خطوة الاستقراء سوف يتم استعمال الاستقراء متعدد الحدود سوف يجعل من مقدرات SIMEX تأخذ الشكل الاتي :

$$\hat{\beta}_{simex} = \sum_{j=1}^{J}s_j\hat{\beta}(\lambda_j);\ \hat{\theta}_{simex}(x_0) = \sum_{j=1}^{J}s_j\,\hat{\theta}\{x_0,\hat{\beta}(\lambda_j),\lambda_j\} \qquad \dots(4)$$

3-3-1-2 السيمكس عندما X منمذجة معلمياً : [5]:

هنا نفترض الحالة التي بها X منمذجة معلمياً والكمية Z منمذجة لامعلمياً . أن دالة الإمكان اللوغارتمية تأخذ الشكل الاتي :

$$\boldsymbol{\mathcal{L}}\{\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\theta}(\boldsymbol{Z}),\boldsymbol{\beta}\}$$

لأي $\boldsymbol{\lambda}$ تعرف $\boldsymbol{\beta}(\boldsymbol{\lambda})$ و $\boldsymbol{\theta}\{\boldsymbol{z}_0,\boldsymbol{\beta}(\boldsymbol{\lambda}),\boldsymbol{\lambda}\}$ على انها حلول المعادلتين الاتيتين :

$$\boldsymbol{0} = \boldsymbol{E}\big(\boldsymbol{\mathcal{L}}_\beta[\boldsymbol{Y},\boldsymbol{W}_b(\boldsymbol{\lambda}),\boldsymbol{\theta}\{\boldsymbol{Z},\boldsymbol{\beta}(\boldsymbol{\lambda}),\boldsymbol{\lambda}\},\boldsymbol{\beta}(\boldsymbol{\lambda})]\big)$$
$$\boldsymbol{0} = \boldsymbol{E}\big(\boldsymbol{\mathcal{L}}_\beta[\boldsymbol{Y},\boldsymbol{W}_b(\boldsymbol{\lambda}),\boldsymbol{\theta}\{\boldsymbol{z}_0,\boldsymbol{\beta}(\boldsymbol{\lambda}),\boldsymbol{\lambda}\},\boldsymbol{\beta}(\boldsymbol{\lambda})]\big|\boldsymbol{Z}=\boldsymbol{z}_0\big)$$

القيم التي تقترن بها تقديرات الامكان التوصيف لـ λ معطاة أيضاً تعرف

$$\theta_\beta\{z_0,\beta(\lambda),\lambda\} = -\frac{E\{\mathcal{L}_{\theta\beta}(\blacksquare)|Z=z_0\}}{E\{\mathcal{L}_{\theta\theta}(\blacksquare)|Z=z_0\}}$$

$$E(\lambda) = E\left[E\{\mathcal{L}_{\beta\beta}(\blacksquare)|Z\} + E\{\mathcal{L}_{\beta\theta}(\blacksquare)|Z\}\theta_\beta^T\{Z,\beta(\lambda),\lambda\}\right]$$

$$\Omega_Z\{z_0,\beta(\lambda),\lambda\} = E\{\mathcal{L}_{\theta\theta}(\blacksquare)|Z=z_0\}$$

حيث هنا $(\blacksquare) = [Y, W_b(\lambda), \theta\{Z,\beta(\lambda),\lambda\}, \beta(\lambda)]$

علاوة على ذلك ، عمل التعريفات

$$\blacksquare_{ib}(\lambda) = [Y_i, W_{ib}(\lambda), \theta\{Z_i,\beta(\lambda),\lambda\}, \beta(\lambda)]$$

$$\xi_{i,B}(\lambda) = -B^{-1}\sum_{b=1}^{B}\left[\mathcal{L}_\beta\{\blacksquare_{ib}(\lambda)\} + \theta_\beta\{Z_i,\beta(\lambda),\lambda\}\mathcal{L}_\theta\{\blacksquare_{ib}(\lambda)\}\right]$$

$$\chi_{i,B}(\lambda) = B^{-1}\sum_{b=1}^{B}\mathcal{L}_\theta\{\blacksquare_{ib}(\lambda)\}$$

بالتعريف ، $\xi_{i,B}(\lambda)$ لها متوسط قيمته صفر بينما $\chi_{i,B}(\lambda)$ له متوسط صفري شرطي على Z_i .

لأي b و λ أفترض المقدرين $\hat{\beta}_b(\lambda)$ و $\hat{\theta}_b\{z_0,\hat{\beta}_b(\lambda),\lambda\}$ على التوالي . متوسط هذه المقدرات عبر $b = 1,2,\dots,B$ يكون بالشكل الاتي :

$$\hat{\beta}(\lambda) = B^{-1}\sum_{b=1}^{B}\hat{\beta}_b(\lambda)$$

$$\hat{\theta}_b\{z_0,\hat{\beta}(\lambda),\lambda\} = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b\{z_0,\hat{\beta}_b(\lambda),\lambda\}$$

وعليه فأن مقدر الـ simex سوف يأخذ الصيغة المذكورة في المعادلة (4) .

2-3-3 مقدر شبه الامكان المعدل للانموذج شبه المعلمي

1-2-3-3 عندما تكون Σ_{uu} معلومة [10]:

يؤدي مقدر شبه الإمكان للباحثين Severini and Staniswalis في الانموذج شبه المعلمي الى تقديرات متحيزة لكل من المعلمة β و الدالة $g(\cdot)$ عندما يتم تجاهل خطأ القياس . أذ تعمل هذه الطريقة تعديل بسيط على مقدرهما الذي هو عبارة عن تصحيح للجزء المعلمي أي يتم تقدير المعلمة المجهولة β والدالة المجهولة $g(\cdot)$ في المعادلة ادناه :

$$Y_i = X_i'\beta + g(T_i) + \epsilon_i \qquad i = 1,2,\dots,n$$

عندما تكون المتغيرات X_i يتم قياسها بوجود خطأ أي بدلاً من مشاهدة X_i يتم مشاهدة

$$W_i = X_i + U_i$$

اذ ان أخطاء القياس U_i موزعة بالتماثل وتكون مستقلة عن المتغيرات (Y_i, X_i, T_i) بمتوسط صفر ومصفوفة تغاير Σ_{uu} .

اذا تم مشاهدة ال $X's$ ، يمكن الحصول على تقدير β بمعدلات التقارب الأعتيادية عن طريق خوارزميه الإمكان الاعتيادية كما يأتي :

لكل ثابت β افترض ان $\hat{g}(T,\beta)$ هو مقدر $g(T)$. على سبيل المثال ، في تطبيق ***Severini*** و ***Staniswalis*** ، $\hat{g}(T,\beta)$ تعظم الإمكان المرجح بأفتراض أن أخطاء الانموذج ϵ_i متجانسة و لها توزيع طبيعي ، مع الترجيحات كونها مرجحات **kernel** بدالة كثافة **kernel** متماثلة $K(\cdot)$ وعرض حزمة h . بعد الحصول على $\hat{g}(T,\beta)$ ، يتم تقدير β بواسطة طريقة المربعات الصغرى بتصغير المقدار

$$\sum_{i=1}^{n}\{Y_i - X_i^T\beta - \hat{g}(T_i,\beta)\}^2$$

في هذه الحالة بالذات ، يمكن تحديد تقدير β بشكل صريح . افترض ان $\hat{g}_{y,h}$ و $\hat{g}_{x,h}$ هما انحداري **kernel** بعرض حزمة h من Y و T من X، على التوالي وعليه فأن

$$\hat{\beta}_n = \left[\sum_{i=1}^{n}\{X_i - \hat{g}_{x,h}(T_i)\}\{X_i - \hat{g}_{x,h}(T_i)\}^T\right]^{-1} * \sum_{i=1}^{n}\{X_i - \hat{g}_{x,h}(T_i)\}\{Y_i - \hat{g}_{y,h}(T_i)\} \qquad ---(5)$$

يمكن استعمال صيغة المربعات الصغرى في المعادلة اعلاه لإظهار أنه إذا تجاهل المرء خطأ القياس واستبدل X ب W، فإن التقدير الناتج سيكون غير متسق بالنسبة إلىβ . على الرغم من أن الانموذج يقترح أكثر من ذلك. فمن المعروف أنه في الانحدار الخطي، يمكن التغلب على عدم الاتساق الناجم عن خطأ القياس بتطبيق ما يسمى " تصحيح للتخفيف " والذي يشير الى تخلص معامل الارتباط من تأثير خطأ القياس أي الاخذ بنظر الاعتبار وجود خطأ قياس عند التقدير. في سياق النماذج شبه المعلمية ، هذا يشير إلى استعمال المقدر

$$\hat{\beta}_n = \left[\sum_{i=1}^{n}\{W_i - \hat{g}_{w,h}(T_i)\}\{W_i - \hat{g}_{w,h}(T_i)\}^T - n\Sigma_{uu}\right]^{-1} * \sum_{i=1}^{n}\{W_i - \hat{g}_{w,h}(T_i)\}\{Y_i - \hat{g}_{y,h}(T_i)\} \qquad ---(6)$$

3-3-2-2عندما تكون Σ_{uu} مجهولة :

على الرغم من أن في بعض الحالات تكون مصفوفة تغاير خطأ القياس Σ_{uu} تم انشاؤها بواسطة تجارب مستقلة ، في تجارب أخرى تكون مجهولة ويجب تقديرها . الأسلوب المعتاد للقيام بذلك هو من خلال التكرار الجزئي، بحيث نشاهد

$$W_{ij} = X_i + U_{ij} \qquad j = 1, \ldots, m_i$$

للسهولة، يتم افتراض الحالة التي فيها تكون $m_i \le 2$ ونفترض أن جزء δ من البيانات لديه مثل هذه المكررات . أفترض أن $\overline{W}_i$ يمثل متوسط عينة من مكررات . فأن التقدير غير المتحيز والمتسق لطريقة العزوم لـ Σ_{uu} هو

$$\hat{\Sigma}_{uu} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(W_{ij} - \overline{W}_i)(W_{ij} - \overline{W}_i)^T}{\sum_{i=1}^{n}(m_i - 1)}$$

يتغير المقدر قليلا فقط لاستيعاب المكررات، ليصبح

$$\hat{\beta}_n = \left[\sum_{i=1}^{n}\{\bar{W}_i - \hat{g}_{w,h}(T_i)\}^{\otimes 2} - \{W_i - \hat{g}_{w,h}(T_i)\}^T - n(1-\delta/2)\hat{\Sigma}_{uu}\right]^{-1} * \sum_{i=1}^{n}\{\bar{W}_i - \hat{g}_{w,h}(T_i)\}\{Y_i - \hat{g}_{y,h}(T_i)\} \quad ---(7)$$

حيث $\hat{g}_{w,h}(\cdot)$ هو *انحدار* **kernel** لـ$\bar{W}_i\,'s$ على T_i . وأن

$\hat{\beta}_n \sim Normal(0, B^{-1}\Gamma_2 B^{-1})$

4- الجانب التجريبي

1-4 مفهوم المحاكاة:

تعد المحاكاة من الأدوات المهمة التي تستعمل للحصول على إجابات مناسبة في دراسة و صياغة وحل النماذج الرياضية والاحصائية نظراً لكون العديد من المسائل و النماذج لا يمكن تمثيلها رياضياً بسبب الصياغة المعقدة للانموذج او بسبب الطبيعة العشوائية للمسألة المدروسة [1] .

ويمكن تعريف المحاكاة بحسب نايلر على أنها تقنية عددية تستعمل للقيام باختبارات على حاسوب عددي، وتتضمن علاقات منطقية ورياضية تتفاعل فيما بينها لتصف سلوك و بنية منظومة معقدة في العالم الحقيقي على امتداد فترة من الزمن. وغالبا ما توصف المحاكاة بأنها عملية خلق روح الواقع دون تحقيق هذا الواقع مطلقاً.

2-4 النماذج المستعملة في المحاكاة

تتنوع النماذج بتنوع الظواهر التي تمثلها، ولذا لا يمكن عرض جميع انماط النماذج في الوقت نفسه، إلا انه تم استيفاء بعض النماذج من بحوث منشورة فعلاً اذ تم استعمال(**PLM**) لصياغة مركبات المركبة اللامعلمية بحيث تكون:

$$Y_i = X_i'\beta + f(t_i) + \varepsilon_i$$

اذ ان:

$$X = (\underline{1}, \underline{X}_1, \underline{X}_2); \beta = (\beta_\circ, \beta_1, \beta_2)'; f(t_i) = (f(t_1), f(t_2), \ldots, f(t_n))'$$

وبذلك كانت المركبة اللامعلمية للنموذج المذكور آنفاً كما يأتي:

الأنموذج الأول

$$f_1(t_{1i}) = 0\cdot 5\ \mathrm{Sin}(2\pi t_i)$$

الأنموذج الثاني

$$f_2(t_{2i}) = 3^{t_i} - 2^{t_i} + exp(-5t_i) + exp\left\{-20\left(t_i - \frac{1}{2}\right)^2\right\}$$

الأنموذج الثالث

$$f_3(t_{3i}) = \frac{16}{3}\, t_i(t_i - 1)^2$$

3-4 تحليل نتائج تجربة المحاكاة:

جدول (4-1) يبين قيم معدل متوسط مربعات الخطأ $MASE$ وبأحجام عينات (50, 150, 250) وتباينات (20,16,10) للأنموذج الاول

	Mean	σ	Simex methods	Quassi likelihood	Best
n=50	40	10	8.65125898	0.01756558	Quassi likelihood
		16	40.7506066	0.047554951	Quassi likelihood
		20	5.89531517	0.133469485	Quassi likelihood
	50	10	13.5046213	0.027364727	Quassi likelihood
		16	63.6336753	0.074166684	Quassi likelihood
		20	9.20700775	0.208308538	Quassi likelihood
	60	10	19.4342101	0.039327065	Quassi likelihood
		16	91.5949276	0.106667729	Quassi likelihood
		20	13.2538474	0.299736391	Quassi likelihood
n=150	40	10	0.01668999	0.000038117	Quassi likelihood
		16	0.07909080	0.000104135	Quassi likelihood
		20	0.01137627	0.000293961	Quassi likelihood
	50	10	0.02612811	0.000059665	Quassi likelihood
		16	0.12372364	0.000162888	Quassi likelihood
		20	0.017787787	0.0004596211	Quassi likelihood
	60	10	0.037672517	0.0000860188	Quassi likelihood
		16	0.17830059	0.0002347277	Quassi likelihood
		20	0.02562627	0.0006621372	Quassi likelihood
n=250	40	10	0.000024316	0.00000005479	Quassi likelihood
		16	0.000114966	0.0000001498	Quassi likelihood
		20	0.000016506	0.00000042306	Quassi likelihood
	50	10	0.000038057	0.00000008579	Quassi likelihood
		16	0.000179813	0.00000023437	Quassi likelihood
		20	0.000028507	0.00000066153	Quassi likelihood
	60	10	0.000054863	0.00000012372	Quassi likelihood
		16	0.000259101	0.00000033777	Quassi likelihood
		20	0.000037177	0.00000095308	Quassi likelihood

جدول (4-2) يبين قيم معدل متوسط مربعات الخطأ $MASE$ وبأحجام عينات (50, 150, 250) وتباينات (20,16,10) للأنموذج الثاني

	Mean	σ	Simex methods	Quassi likelihood	Best
n=50	40	10	57648.9658	0.227280195	Quassi likelihood
		16	157196.3177	0.328287016	Quassi likelihood
		20	16574.2792	0.545576641	Quassi likelihood
	50	10	62915.7351	0.243223962	Quassi likelihood
		16	176107.9454	0.360108188	Quassi likelihood
		20	19056.8354	0.614911364	Quassi likelihood
	60	10	74140.5669	0.276735723	Quassi likelihood
		16	217154.978	0.428169343	Quassi likelihood
		20	24541.6496	0.766035366	Quassi likelihood
n=150	40	10	756.995627	0.000368973	Quassi likelihood
		16	218.238689	0.000563744	Quassi likelihood
		20	242.982955	0.000995510	Quassi likelihood
	50	10	835.861721	0.000399354	Quassi likelihood

		16	247.357062	0.000621161	Quassi likelihood
		20	28.2241471	0.001135577	Quassi likelihood
	60	10	100.531851	0.000463720	Quassi likelihood
		16	311.062552	0.000760676	Quassi likelihood
		20	369.574232	0.001443358	Quassi likelihood
n=250	40	10	0.77067764	0.000000436	Quassi likelihood
		16	2.29709809	0.000000693	Quassi likelihood
		20	0.26380609	0.0000012746	Quassi likelihood
	50	10	0.85699826	0.0000004761	Quassi likelihood
		16	2.62113038	0.0000007765	Quassi likelihood
		20	0.30816119	0.0000014652	Quassi likelihood
	60	10	1.043386536	0.0000005607	Quassi likelihood
		16	3.333317701	0.0000009570	Quassi likelihood
		20	0.407206525	0.0000018861	Quassi likelihood

جدول (4-3) يبين قيم معدل متوسط مربعات الخطأ $MASE$ وبأحجام عينات (50 ,150 ,250) وتباينات (20,16,10) للأنموذج الثالث

	Mean	σ	Simex methods	Quassi likelihood	Best
n=50	40	10	110.071660	0.003069240	Quassi likelihood
		16	1062.71646	0.028247651	Quassi likelihood
		20	207.158251	0.111381461	Quassi likelihood
	50	10	216.013945	0.007795343	Quassi likelihood
		16	1379.52313	0.038104592	Quassi likelihood
		20	262.842462	0.143850791	Quassi likelihood
	60	10	357.332148	0.014684790	Quassi likelihood
		16	2136.93872	0.062235554	Quassi likelihood
		20	394.071784	0.221231147	Quassi likelihood
n=150	40	10	3.22275153	0.000015735	Quassi likelihood
		16	20.0406774	0.000072274	Quassi likelihood
		20	3.76609896	0.000266011	Quassi likelihood
	50	10	4.29642896	0.000022486	Quassi likelihood
		16	25.7691271	0.000095657	Quassi likelihood
		20	4.75564486	0.000340641	Quassi likelihood
	60	10	6.90607827	0.000039593	Quassi likelihood
		16	39.3834864	0.000152239	Quassi likelihood
		20	7.08062439	0.000517550	Quassi likelihood
n=250	40	10	0.00432872	0.0000000270	Quassi likelihood
		16	0.02592370	0.0000001131	Quassi likelihood
		20	0.00477741	0.0000004003	Quassi likelihood
	50	10	0.00571003	0.0000000375	Quassi likelihood
		16	0.03316756	0.0000001481	Quassi likelihood
		20	0.00601706	0.0000005100	Quassi likelihood
	60	10	0.00904556	0.0000000637	Quassi likelihood
		16	0.05032934	0.0000002323	Quassi likelihood
		20	0.00892488	0.0000007692	Quassi likelihood

4- الاستنتاجات و التوصيات

1-5 الاستنتاجات

بعد تنفيذ تجارب المحاكاة و ما تم عرضه من نتائج استنتج الباحث ما يأتي :

1- اشارت نتائج المحاكاة الى ان افضل مقدر شبه الإمكان هو الأفضل بغض النظر عن حجم العينة المستعملة و قيمة التباين الموضوعة .

2- ان اسوء مقدر تم الحصول عليه هو مقدر السيمكس بغض النظر عن حجم العينة المستعملة وقيمة التباين الموضوعة

3- هناك علاقة عكسية بين حجم العينة وقيمة $MASE$ اذ نلاحظ انه كلما زادت قيمة حجم العينة نلاحظ انخفاض في قيمة $MASE$ وهذا ما يطابق النظرية الإحصائية

2-5 التوصيات

على ضوء الاستنتاجات التي توصلنا اليها من خلال البحث يمكن اجمال التوصيات الاتية :

1- استعمال مقدر شبه الإمكان عند التحليل في الانموذج شبه المعلمي في ظل وجود خطأ قياس لما يبديه هذا المقدر من كفاءة و مرونة عند التطبيق

2- أهمية الأخذ بعين الاعتبار كافة الأخطاء المرافقة لقياس المتغيرات المختلفة التي تمثل اي ظاهرة من ظواهر الحياة الاقتصادية او الاجتماعية... الخ وعدم إهمالها او التغاضي عنها لتأثيرها الواضح في نمذجة تلك الظواهر ودقة نتائجها.

3- عمل المزيد من الدراسات حول تقدير دالة الانحدار شبه المعلمية بوجود خطأ القياس بأستعمال نماذج شبه معلمية مختلفة وتطويرها لما تمثله من دقة في تفسير نماذج الانحدار في التطبيقات ذات العلاقة في مجالات العمل المختلفة.

6- المصادر

[1] بتال، احمد حسين، احمد، عصام كامل، خضير، البراء عبد الوهاب (2008) "استخدام المحاكاة في تدريس الانحدار الخطي البسيط" مجلة جامعة الانبار للعلوم الصرفة . العدد الثالث، المجلد الثاني

[2] عيس، اسيل مسلم (2011) " مقارنة بعض المقدرات شبه المعلمية لتقدير دالة استهلاك الطاقة الكهربائية لمدينة بغداد"، رسالة ماجستير في الأحصاء، كلية الأدارة والأقتصاد، جامعة بغداد .

[3] شهاب، طارق عزيز (2016) ، " بعض الطرائق شبه المعلميه في تقدير واختيار المتغير لانموذج المؤشر الواحد" ، اطروحة دكتوراه فلسفة في الاحصاء ، كلية الأدارة والأقتصاد ، جامعة بغداد .

[4] علوان ، اسراء سعدون (2003) " مقارنة بين طريقتي SIMEX و LLSلتقدير دالة الانحدار اللامعلمية بأستخدام المحاكاة "، رسالة ماجستير في الأحصاء، كلية الأدارة والأقتصاد، جامعة بغداد

[5] Apanasovich, T. V., Carroll, R. J., & Maity, A. (2009). SIMEX and standard error estimation in semiparametric measurement error models. Electronic journal of statistics, 3, 318.

[6] B. S. Everitt and A. Skrondal , ''the Cambridge dictionary of statistics'', fourth edition, (2010), United States of America by Cambridge University Press, New York

[7] Buonaccorsi, J. P. (2010). Measurement error: models, methods, and applications. United States of America on acid-free paper

[8] Carroll, R. J., Maca, J. D., & Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. Biometrika, 86(3), 541-554.

[9] Carroll, R.J., Ruppert , D. and Stefanski, L.A., (2006), Measurement Error in Nonlinear Models ,2nd edition , NewYork: Chapman & Hall.

[10] LIANG , H. , Hardle , w , & Carroll, R. J. (1999), '' Estimation in a Semiparametric Partially Linear Errors-in-Variables Model '' The Annals of Statistics, Vol. 27, No. 5 (Oct., 1999), pp. 1519-1535

[11] Shang,yi,2012,'' Measurement error adjustment using the simex method: An application to student growth percentiles',k journal of educational measurement,Vol. 49, No. 4, pp. 446–465.

Compare between Simex and Quasi-likelihood methods in estimation of regression function in the presence of measurement error

Abstract

In recent years, the attention of researchers has increased of semi-parametric regression models, because it is possible to integrate the parametric and non-parametric regression models in one and then form a regression model has the potential to deal with the cruse of dimensionality in non-parametric models that occurs through the increasing of explanatory variables. Involved in the analysis and then decreasing the accuracy of the estimation. As well as the privilege of this type of model with flexibility in the application field compared to the parametric models which comply with certain conditions such as knowledge of the distribution of errors or the parametric models may not represent the phenomenon properly studied.

In this paper, we will show semi-parametric methods in estimation of regression function in the presence of measurement error, and these methods are Simex method and Quasi-likelihood method and will be comparing between this methods by using (MASE) criterion. A simulation had been used to study the empirical behavior for the semi-parametric models, with different sample sizes and variances. The results using represent that the instrument variable is better than Simex method at different sample sizes and variances that been used.

Key word / semi-parametric model – measurement error – Simex estimator – Quasi likelihood estimator