تقدير دالة الانحدار شبه المعلميه في ظل وجود خطأ القياس•

ا.د. مناف يوسف حمود
munaf_yousif@yahoo.com
جامعة بغداد - كلية الإدارة والاقتصاد - قسم الإحصاء
احمد مهدي جابر
ahmedalwasty01@gmail.com
جامعة بغداد - كلية الإدارة والاقتصاد - قسم الإحصاء

المستخلص

زاد اهتمام الباحثين في السنوات الاخيرة بنماذج الانحدار شبه المعلميه و ذلك لأنه يمكن من خلالها دمج نماذج الانحدار اللامعلمي و المعلمي في ان واحد و من ثم تكوين انموذج انحدار يتميز بامكانية التعامل مع مشكلة تعدد الابعاد الموجودة في النماذج اللامعلميه و التي 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 المعلميه التي تتقيد بشروط معينة كمعرفة توزيع الاخطاء او قد تكون النماذج المعلميه لا تمثل الظاهرة المدروسة تمثيلاً صحيحاً.

سنستعرض في هذا البحث طرائق شبه معلمية في تقدير دالة الانحدار في ظل وجود أخطاء قياس و هذه الطرائق هي طريقة السيمكس و طريقة المتغيرات المساعدة بالإضافة الى طريقة شبه الإمكان المعدلة وسيتم المقارنة بين هذه الطرائق بأستعمال معيار المقارنة معدل متوسط مربعات الخطأ ($MASE$). اذ تمت المقارنة من خلال استعمال أسلوب المحاكاة بأستعمال النماذج شبه المعلمية واحجام عينات و تباينات مختلفة للتحقق من أداء الطرائق بأستعمال معيار المقارنة أعلاه حيث كانت اهم الاستنتاجات هي ان طريقة المتغيرات المساعدة هي الأفضل بأختلاف حجم العينة المستعملة وقيمة التباين الموضوعة.

الكلمات الرئيسية: انموذج شبه معلمي، خطأ القياس، مقدر السيمكس، المتغيرات المساعدة، مقدر شبه الإمكان.

• بحث مستل من رسالة ماجستير

1. المقدمة Introduction

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 و التي يتم تقديرها بأستعمال عدة طرائق منها المربعات الصغرى و الامكان الاعظم و غيرها 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 و البعض الاخر سلوكاً لامعلمياً .

اما النوع الثاني فيدعى بنماذج الانحدار اللامعلميه والذي لا يتقيد بشروط صارمة من حيث التوزيع الخاص بمتغير الاستجابة والخطأ ، ولا يتقيد بدالة معينة تفسر العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة ويمتاز هذا النموذج بمرونته العالية في التعامل مع البيانات وكذلك 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 وإحدى هذه السلبيات هي وجود مشكلة خطأ القياس التي تحدث عندما ينتهك الافتراض الاساسي في نموذج الانحدار الذي هو ان المتغيرات التوضيحية قد تم قياسها بدون وجود اخطاء قياس والذي بوجود 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 أفضل مقدر.

3. الجانب النظري

3.1 خطأ القياس

يعرف خطأ القياس ببساطه على أنه الفرق بين قيمة متغير المشاهدة (الملاحظة) و قيمة 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 :

$$U_i = X_i - W_i \tag{1}$$

اذ ان :

U_i : يمثل خطأ القياس .

X_i : تمثل القيمة المشاهدة (المقاسة) .

W_i: تمثل القيمة الحقيقية .

أن وجود خطأ القياس يؤدي الى [9]:

1) تقديرات متحيزة وغير 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 المعلميه 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 و اللامعلمي لذا فأنه يعد انموذج وسيط بين النماذج المعلميه واللامعلميه.[2]

3-3 طرائق التقدير

3.3.1 طريقة السيمكس simex method

هي اسلوب للتقدير و الحد من التحيز الناجم عن خطأ القياس وقد اقترحت هذه الطريقة من قبل كوك و ستيفانسكي عام 1994 م لإيجاد تقدير كفوء الى β في ظل وجود اخطاء قياس في المتغيرات التوضيحية لانموذج الانحدار المعلمي و قد اكتسبت منذ ذلك الحين شعبية كبيرة بين الإحصائيين نظراً لبساطتها و عموميتها اذ بأستعمال هذه الطريقة يمكن الحصول تقريباً على مقدرات غير متحيزة و كفوءة للمعلمة بالنسبة للأنواع المختلفة من نماذج الانحدار. [12]

تتكون خوارزمية السيمكس بشكل عام من ثلاث خطوات هي خطوة المحاكاة و خطوة التقدير و خطوة 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 $\sum_{\mathrm{u}} \lambda_j$ و تضاف الى بيانات W الاصلية، وبذلك تنشأ مجموعة بيانات بتباينات خطأ القياس الاكبر بالتوالي . أن تباين خطأ القياس الكلي لـ 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 و من 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 في ادناه:

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

$$W_{ib}(\lambda) = W_i + \lambda^{1/2} U_{ib}$$

اذ أن:

$$W_i = X_i + U_i$$

U_{ib}: هي متغيرات عشوائية يتم توليدها وتتبع التوزيع الطبيعي بمتوسط صفر وتباين 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 $W_b(\lambda)$. تعرف $\beta(\lambda)$ و $\theta\{x_0, \beta(\lambda), \lambda\}$ على انها الحلول للمعادلتين الاتيتين:

$$0 = E\left(\mathcal{L}_\beta\left[Y, Z, \theta\{W_b(\lambda), \beta(\lambda), \lambda\}, \beta(\lambda)\right]\right); \qquad \dots (2)$$

$$0 = E\left(\mathcal{L}_\theta\left[Y, Z, \theta\{x_0, \beta(\lambda), \lambda\}, \beta(\lambda)\right] \middle| W_b(\lambda) = x_0\right)$$

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

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

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

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

$$\hat{\beta}_{simex} = \sum_{j=1}^{J} s_j \hat{\beta}(\lambda_j); \hat{\theta}_{simex}(x_0) = \sum_{j=1}^{J} s_j \hat{\theta}\{x_0, \hat{\beta}(\lambda_j), \lambda_j\} \qquad \dots (3)$$

3.3.1.2 السيمكس عندما X منمذجة معلمياً : [5]

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

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

لأي λ تعرف $\beta(\lambda)$ و $\theta\{z_0, \beta(\lambda), \lambda\}$ على انها حلول المعادلتين الاتيتين :

$$0 = E\left(\mathcal{L}_\beta[Y, W_b(\lambda), \theta\{Z, \beta(\lambda), \lambda\}, \beta(\lambda)]\right)$$

$$0 = E\left(\mathcal{L}_\beta[Y, W_b(\lambda), \theta\{z_0, \beta(\lambda), \lambda\}, \beta(\lambda)]\middle|Z = z_0\right)$$

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

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

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

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

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

علاوة على ذلك، عمل التعريفات

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

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

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

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 $b = 1,2, \ldots, B$ يكون بالشكل الاتي :

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

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

و عليه فأن مقدر ال simex سوف يأخذ الصيغة المذكورة في المعادلة (3) .

3.3.2 مقدر شبه الامكان المعدل للانموذج شبه المعلمي

3.3.2.1 عندما تكون Σ_{uu} معلومة: [11]

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 و الدالة المجهولة $g(\cdot)$ في المعادلة ادناه :

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

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 و تكون مستقله عن المتغيرات (Y_i, X_i, T_i) بمتوسط صفر و مصفوفة 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 $K(\cdot)$ و عرض حزمة h . بعد الحصول على $\hat{g}(T,\beta)$ ، يتم تقدير β بواسطة طريقة المربعات الصغرى بتصغير المقدار

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

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 و عليه فأن

$$\hat{\beta}_n = \left[\sum_{i=1}^{n}\{X_i - \hat{g}_{x,h}(T_i)\}\{X_i - \hat{g}_{x,h}(T_i)\}^T\right]^{-1} * \sum_{i=1}^{n}\{X_i - \hat{g}_{x,h}(T_i)\}\{Y_i - \hat{g}_{y,h}(T_i)\} \qquad (4)$$

يمكن استعمال صيغة المربعات الصغرى في المعادلة اعلاه لإظهار أنه إذا تجاهل المرء خطأ القياس واستبدل X بـ W، فإن التقدير الناتج سيكون غير متسق بالنسبة إلى β . 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 " و الذي 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

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

3.3.2.2: عندما 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 و يجب تقديرها . الأسلوب المعتاد للقيام بذلك هو من خلال التكرار الجزئي، بحيث نشاهد

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

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

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

يتغير المقدر قليلا فقط لاستيعاب المكررات، ليصبح

$$\hat{\beta}_n = \left[\sum_{i=1}^{n}\{\overline{W}_i - \hat{g}_{w,h}(T_i)\}^{\otimes 2} - \{W_i - \hat{g}_{w,h}(T_i)\}^{T} - n(1-\delta/2)\hat{\Sigma}_{uu}\right]^{-1} * \sum_{i=1}^{n}\{\overline{w}_j - \hat{g}_{w,h}(T_i)\}\{Y_i - \hat{g}_{y,h}(T_i)\} \quad (6)$$

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

3.3.3 طريقة المتغيرات المساعدة

هي طريقة تعتمد على المتغيرات الموجودة خارج المعادلة التفسيرية في عملية التقدير اذ استعمل مصطلح المتغيرات المساعدة لأول مرة من قبل Philip G. Wright في منشوره عام 1928م بعنوان التعريفة الجمركية على الزيوت الحيوانية والنباتية، ومنذ ذلك الحين ظهرت لهذه الطريقة تطبيقات واسعة في الإحصاء والاقتصاد. وهنالك عدة شروط لتطبيق هذه الطريقة و تختلف هذه الشروط بأختلاف نوعية الانموذج المستعمل الا أن هنالك شرطين رئيسيين يجب توافرهما لأستعمال هذه الطريقة و هما: [10]

1. يجب ان يكون المتغير المساعد المستخدم له ارتباط قوي مع المتغير التوضيحي غير الملوث بأخطاء القياس لكي يكون مُقدر المتغير المساعد جيد ويمكن الاعتماد عليه.
2. عدم وجود ارتباط بين المتغير المساعد المستعمل وحد الخطأ وكذلك الحال بالنسبة لأخطاء القياس.

اذ ان عند انتهاك الشرطين المذكورين انفاً أي افتراض وجود متغير مساعد خاطئ فأن مقدرات المتغير المساعد المستخدم يمكن ان تكون متحيزة و غير متسقة حتى في حالة كون حجم العينة كبير وخطأ القياس صغير و عليه لا تستخدم تقديرات المتغير المساعد الا عندما تكون هناك أدله مقنعه على ان افتراضات المتغير المساعد متحققة. [9]

سنتطرق الان الى إيجاد مقدر المتغير المساعد في انموذج شبه معلمي كالاتي : [13] افترض أن النموذج المستعمل في عملية التقدير هو انموذج الانحدار الجزئي شبه المعلمي الاتـي:

$$Y_t = g(Z_{1t}, X_t, \epsilon_t) = m(Z_{1t}) + X_t\beta + \epsilon_t \quad t = 1, \dots, n \quad (7)$$

وافترض وجود متغيرات مساعدة $Z_t = (Z_{1t}, Z_{2t})$ بحيث أن

$$E(\epsilon_t|Z_t) = E(\epsilon_t|Z_{1t}) = 0 \qquad \text{for all } t$$

افترض أن التوقع الشرطي الحقيقي معرف كالاتي:

$$E(Y_t|Z_{1t}) = m(Z_{1t}) + E(X_t|Z_{1t})\beta$$

بطرح $E(Y_t|Z_{1t})$ من المعادلة (7) نحصل على

$$Y_t - E(Y_t|Z_{1t}) = \left(X_t - E(X_t|Z_{1t})\right)\beta + \epsilon_t \qquad (8)$$

بصورة عامة التوقعات الشرطية مجهولة و لكن يمكن استبدالها بمقدرات التوقع الشرطية اللامعلمية $\widehat{E}(Y_t|Z_{1t})$ و $\widehat{E}(X_t|Z_{1t})$. ومع ذلك ، بسبب الارتباط بين X و ϵ لا يمكن تطبيق مقدر روبنسون عن طريق انحدار $Y_t - \widehat{E}(Y_t|Z_{1t})$ على $X_t - \widehat{E}(X_t|Z_{1t})$. يمكن تقدير β كالاتي :

$$\tilde{\beta} = \left(Q'\breve{X}\right)^{-1}Q'\breve{Y} \qquad (9)$$

باستعمال المتغيرات المساعدة $Q = \{Q'_1, Q'_2, \dots, Q'_n\}$ اذا كان $Q'_t \in \Re^K$ وQ_t يحتوي Z_{1t} و Z_{2t} . هنا

$$\widetilde{X} = \begin{bmatrix} X_1 - \widehat{E}(X_1|Z_{11}) \\ \vdots \\ X_n - \widehat{E}(X_n|Z_{1n}) \end{bmatrix} \quad , \quad \widetilde{Y} = \begin{bmatrix} Y_1 - \widehat{E}(Y_1|Z_{11}) \\ \vdots \\ Y_n - \widehat{E}(Y_n|Z_{1n}) \end{bmatrix}$$

سوف يتم استعمال مجموعات مقدر لي وستينجوس (1996)

$$Q_t = Z_{2t} - \widehat{E}(Z_{2t}|Z_{1t})$$

في التعامل مع التداخل في نماذج الخطية جزئياً . في هذه الحالة $Q_t \in \Re^{K+q}$ لعدد صحيح موجب q نفترض ان المقدر يكون بالصيغة الاتية :

$$\breve{\beta} = \left(\widetilde{X}\,Q(Q'Q)^{-1}Q'\breve{X}\right)^{-1}\widetilde{X}\,Q(Q'Q)^{-1}Q'\breve{Y} \qquad (10)$$

في المعادلة المذكورة انفاً، أذا كان التوقع الشرطي لـ $\tilde{X}$ بالنسبة الى Z أو النموذج المختزل ليس خطياً في Z ، و بما أن التوقع الشرطي هو التنبؤ الأمثل ل $\left(\tilde{X}\middle|Z\right)$ بمعنى مربع المتوسط، حيث يتوقع تحقيق كفاءة عن طريق استبدال الإسقاط $Z(Z'Z)^{-1}Z'\,\tilde{X}$ مع تقدير لامعلمي . كما هو مبين في روبنسون *(1988, p. 945)* المتغير المساعد لـ $X_t - E(X_t|Z_{1t})$ هو متجه دالة Z_t التي تتضمن Z_{1t} و مستقل عن ϵ_t بحيث مصفوفة التباين في التوزيع المحدود لمقدر متسق $n^{\frac{1}{2}}$ ل β الموجودة. المرشح للمساعد

$\hat{E}(X_t|Z_t) - \hat{E}(X_t|Z_{1t})$ حيث $\hat{E}(X_t|Z_t)$ هو مقدر لامعلمي ل $E(X_t|Z_t)$. يُلاحظ أنه يمكن تقدير $E(X_t - E(X_t|Z_{1t})|Z_t)$ بأستمرار حتى لو أن $E(X_t - E(X_t|Z_{1t})|Z_t)$ غير خطي في Z_t . التوقع الشرطي بصورة عامة صيغة مجهولة إذا كانت المعادلة الهيكلية ل X_t Y_t و يحتوي على اللاخطية في المتغيرات الداخلية (و / أو) الخارجية للصيغة المجهولة ، أو حتى لو كانت صيغة من أشكال اللاخطية معروفة ولكن المعلومات عن التوزيع الشرطي ل $X_t's$ بالنسبة الى Z_t غير كافية لتحديد $E(X_t|Z_t)$. تعرف دالة كثافة Z_{1t} في z_{10} على أنها $f_1(z_{10})$ و دالة كثافة Z_t في z_0 على انها $f(z_0)$. ونحن نقدر هاتين الدالتين مع مقدر كثافة $Rosenblat$ مع كل من المتغيرات المستمرة والمتقطعة، ونحن نستخدم مقدر نادرايا-واتسون ل$E(A|z_{10})$ و$E(A|z_0)$. وهي على وجه التحديد:

$$\hat{f}_1(z_{10}) = \frac{1}{nh_1^{l_{1c}}} \sum_{t=1}^{n} K_1\left(\frac{Z_{1t}^c - z_{10}^c}{h_1}\right) I\left(Z_{1t}^d = z_{10}^d\right)$$

$$\hat{E}(A|z_{10}) = \frac{\frac{1}{nh_1^{l_{1c}}} \sum_{t=1}^{n} K_1\left(\frac{Z_{1t}^c - z_{10}^c}{h_1}\right) I\left(Z_{1t}^d = z_{10}^d\right) A_t}{\hat{f}_1(z_{10})}$$

$$\hat{f}(z_0) = \frac{1}{nh_2^{l_{1c}+l_{2c}}} \sum_{t=1}^{n} K_2\left(\frac{Z_{1t}^c - z_{10}^c}{h_2}, \frac{Z_{2t}^c - z_{20}^c}{h_2}\right) I\left(Z_{1t}^d = z_{10}^d, Z_{2t}^d = z_{20}^d\right)$$

$$\hat{E}(A|z_0) = \frac{\frac{1}{nh_2^{l_{1c}+l_{2c}}} \sum_{t=1}^{n} K_2\left(\frac{Z_{1t}^c - z_{10}^c}{h_2}, \frac{Z_{2t}^c - z_{20}^c}{h_2}\right) I\left(Z_{1t}^d = z_{10}^d, Z_{2t}^d = z_{20}^d\right) A_t}{\hat{f}(z_0)}$$

اذ :

h_1 و h_2 : هما عرض الحزمة و اللذان يقتربان من الصفر عندما $n \to \infty$

$K_1(\cdot), K_2(\cdot), l(\cdot)$: تمثل دوال و مؤشرات kernel .

افترض أن:

$$\widehat{W}=\begin{bmatrix}\widehat{W}_1\\ \widehat{W}_2\\ \vdots\\ \widehat{W}_n\end{bmatrix}=\begin{bmatrix}\widehat{W}_{1,1} & \widehat{W}_{1,2} & \cdots & \widehat{W}_{1,k}\\ \widehat{W}_{2,1} & \widehat{W}_{2,2} & \cdots & \widehat{W}_{2,k}\\ \vdots & \vdots & \ddots & \vdots\\ \widehat{W}_{n,1} & \widehat{W}_{n,2} & \cdots & \widehat{W}_{n,k}\end{bmatrix},\widehat{W}_{t,k}$$
$$=\widehat{E}\left(X_{t,k}\middle|Z_t\right)-\widehat{E}\left(X_{t,k}\middle|Z_{1t}\right)$$

حيث:

$X_{t,k}$ هي العنصر K_{th} للمتجه العشوائي X_t . يعرف

$$Y=(Y_1,\dots,Y_n)'$$

$$\vec{Z}=(Z_1,\dots,Z_n)'$$

$$\vec{Z}_1=(Z_{11},\dots,Z_{1n})'$$

$$\widehat{E}\left(Y\middle|\vec{Z}_1\right)=\left(\widehat{E}\left(Y_1\middle|\vec{Z}_{11}\right),\widehat{E}\left(Y_2\middle|\vec{Z}_{12}\right),\dots,\widehat{E}\left(Y_n\middle|\vec{Z}_{1n}\right)\right)'$$

يعرف مقدر المتغير المساعد كالاتي :

$$\widehat{\beta}=\left(\widehat{W}'\,\widehat{W}\right)^{-1}\,\widehat{W}'\,\left(\widehat{E}\left(Y\middle|\vec{Z}\right)-\widehat{E}\left(Y\middle|\vec{Z}_1\right)\right)\qquad ---(11)$$

4. الجانب التجريبي

4.1 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 حسب نايلر على أنها تقنية عددية تستعمل للقيام باختبارات على حاسوب عددي، و تتضمن علاقات منطقية و رياضية تتفاعل فيما بينها لتصف سلوك و بنية منظومة معقدة في العالم الحقيقي على امتداد فترة من الزمن. و غالبا ما توصف المحاكاة بأنها عملية خلق روح الواقع دون تحقيق هذا الواقع مطلقاً.

4.2 النماذج المستعملة في المحاكاة

تتنــوع النمــاذج بتنــوع الظــواهر التــي تمثلهــا، ولــذا لايمكــن عــرض جميــع انمــاط النمــاذج فــي الوقــت نفســه، إلا انــه تــم اســتيفاء بعــض النمــاذج مــن بحــوث منشــورة فعــلاً اذ تــم اســتعمال(PLM) لصــياغة مركبــات 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 اعلاه كما يأتي:

الأنموذج الأول:

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

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:

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

4.3 تحليل نتائج تجربة المحاكاة:

جدول (4.1): قيم معدل متوسط مربعات الخطأ MASE و بأحجام عينات (50, 150, 250) و تباينات (20,16,10) للأنموذج الاول

	Mean	σ	Semex methods	Instrumental Variable methods	Quassi likelihood	Best
n=50	40	10	8.65125898	4.226351289	0.01756558	Quassi likelihood
		16	40.7506066	1.632598745	0.047554951	Quassi likelihood
		20	5.89531517	0.3269874522	0.133469485	Quassi likelihood
	50	10	13.5046213	0.0045478823	0.027364727	InstrumentalVariable
		16	63.6336753	0.0094523297	0.074166684	Instrumental Variable
		20	9.20700775	0.0516216685	0.208308538	Instrumental Variable
	60	10	19.4342101	0.0109268736	0.039327065	Instrumental Variable
		16	91.5949276	0.0293354396	0.106667729	Instrumental Variable
		20	13.2538474	0.0972263551	0.299736391	Instrumental Variable
n=150	40	10	0.01668999	1.3265922452	0.000038117	Quassi likelihood
		16	0.07909080	0.3265980021	0.000104135	Quassi likelihood
		20	0.01137627	0.0589742315	0.000293961	Quassi likelihood
	50	10	0.02612811	0.0000175998	0.000059665	Instrumental Variable
		16	0.12372364	0.0000386598	0.000162888	Instrumental Variable
		20	0.017787787	0.0001149042	0.0004596211	Instrumental Variable
	60	10	0.037672517	0.00001192159	0.0000860188	Instrumental Variable
		16	0.17830059	0.00006652317	0.0002347277	Instrumental Variable
		20	0.02562627	0.00020611786	0.0006621372	Instrumental Variable
n=250	40	10	0.000024316	0.25698743256	0.00000005479	Quassi likelihood
		16	0.000114966	0.02156325899	0.0000001498	Quassi likelihood
		20	0.000016506	0.00012365854	0.00000042306	Quassi likelihood
	50	10	0.000038057	0.00000002159	0.00000008579	Instrumental Variable
		16	0.000179813	0.000000113468	0.00000023437	Instrumental Variable
		20	0.000028507	0.000000165787	0.00000066153	Instrumental Variable
	60	10	0.000054863	0.000000023286	0.00000012372	Instrumental Variable
		16	0.000259101	0.000000084737	0.00000033777	Instrumental Variable
		20	0.000037177	0.000000238757	0.00000095308	Instrumental Variable

جدول (4.2): قيم معدل متوسط مربعات الخطأ MASE و بأحجام عينات (50, 150, 250) و تباينات (20,16,10) للأنموذج الثاني

	Mean	σ	Semex methods	Instrumental Variable methods	Quassi likelihood	Best
n=50	40	10	57648.9658	0.7325698417	0.227280195	Quassi likelihood
		16	157196.3177	1.3256895551	0.328287016	Quassi likelihood
		20	16574.2792	0.6258964002	0.545576641	Quassi likelihood
	50	10	62915.7351	0.0063275758	0.243223962	InstrumentalVariable
		16	176107.9454	0.00601801125	0.360108188	Instrumental Variable
		20	19056.8354	0.05294830575	0.614911364	Instrumental Variable
	60	10	74140.5669	0.00293072083	0.276735723	Instrumental Variable
		16	217154.978	0.04425422360	0.428169343	Instrumental Variable
		20	24541.6496	0.06806000208	0.766035366	Instrumental Variable

n=150	40	10	756.995627	0.00001300335	0.000368973	Instrumental Variable
		16	218.238689	0.00001586620	0.000563744	Instrumental Variable
		20	242.982955	0.00012141945	0.000995510	Instrumental Variable
	50	10	835.861721	0.00002376440	0.000399354	Instrumental Variable
		16	247.357062	0.00004734118	0.000621161	Instrumental Variable
		20	28.2241471	0.00023963885	0.001135577	Instrumental Variable
	60	10	100.531851	0.00003480145	0.000463720	Instrumental Variable
		16	311.062552	0.00007101149	0.000760676	Instrumental Variable
		20	369.574232	0.00003547929	0.001443358	Instrumental Variable
n=250	40	10	0.77067764	0.000000015317	0.000000436	Instrumental Variable
		16	2.29709809	0.000000008949	0.000000693	Instrumental Variable
		20	0.26380609	0.000000187393	0.0000012746	Instrumental Variable
	50	10	0.85699826	0.000000025101	0.0000004761	Instrumental Variable
		16	2.62113038	0.000000064531	0.0000007765	Instrumental Variable
		20	0.30816119	0.000000175257	0.0000014652	Instrumental Variable
	60	10	1.043386536	0.000000057309	0.0000005607	Instrumental Variable
		16	3.333317701	0.000000091328	0.0000009570	Instrumental Variable
		20	0.407206525	0.000000254843	0.0000018861	Instrumental Variable

جدول (4.3): قيم معدل متوسط مربعات الخطأ MASE و بأحجام عينات (50 ,150 ,250) و تباينات (20,16,10) للأنموذج الثالث

	Mean	σ	Semex methods	Instrumental Variable methods	Quassi likelihood	Best
n=50	40	10	110.071660	0.3569871251	0.003069240	Quassi likelihood
		16	1062.71646	0.1237465580	0.028247651	Quassi likelihood
		20	207.158251	0.3698741582	0.111381461	Quassi likelihood
	50	10	216.013945	0.0054872658	0.007795343	InstrumentalVariable
		16	1379.52313	0.0100182726	0.038104592	Instrumental Variable
		20	262.842462	0.0529154174	0.143850791	Instrumental Variable
	60	10	357.332148	0.0103195897	0.014684790	Instrumental Variable
		16	2136.93872	0.0395689899	0.062235554	Instrumental Variable
		20	394.071784	0.0395210370	0.221231147	Instrumental Variable
n=150	40	10	3.22275153	0.0000098275	0.000015735	Instrumental Variable
		16	20.0406774	0.0000340696	0.000072274	Instrumental Variable
		20	3.76609896	0.0000760952	0.000266011	Instrumental Variable
	50	10	4.29642896	0.0000177958	0.000022486	Instrumental Variable
		16	25.7691271	0.0000363496	0.000095657	Instrumental Variable
		20	4.75564486	0.0001239011	0.000340641	Instrumental Variable
	60	10	6.90607827	0.0000218188	0.000039593	Instrumental Variable
		16	39.3834864	0.0000485574	0.000152239	Instrumental Variable
		20	7.08062439	0.0001926271	0.000517550	Instrumental Variable
n=250	40	10	0.00432872	0.00000001857	0.0000000270	Instrumental Variable
		16	0.02592370	0.00000004606	0.0000001131	Instrumental Variable
		20	0.00477741	0.00000014177	0.0000004003	Instrumental Variable
	50	10	0.00571003	0.00000002160	0.0000000375	Instrumental Variable
		16	0.03316756	0.00000005883	0.0000001481	Instrumental Variable
		20	0.00601706	0.00000016578	0.0000005100	Instrumental Variable
	60	10	0.00904556	0.00000003545	0.0000000637	Instrumental Variable
		16	0.05032934	0.00000008524	0.0000002323	Instrumental Variable
		20	0.00892488	0.00000023554	0.0000007692	Instrumental Variable

5. الاستنتاجات والتوصيات

5.1 الاستنتاجات

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

1. اشارت اغلب نتائج المحاكاة الى ان طريقة المتغيرات المساعدة هي الأفضل بأختلاف حجم العينة المستعملة و قيمة التباين الموضوعة.
2. ان اسوء مقدر تم الحصول عليه هو مقدر السيمكس بغض النظر عن حجم العينة المستعملة و قيمة التباين الموضوعة.
3. هنالك علاقة عكسية بين حجم العينة و قيمة MASE اذ نلاحظ انه كلما زادت قيمة حجم العينة نلاحظ انخفاض في قيمة MASE وهذا ما يطابق النظرية الإحصائية.

5.2 التوصيات

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

1) استعمال مقدر المتغيرات المساعدة عند التحليل في الانموذج شبه المعلمي في ظل وجود خطأ قياس لما يبديه هذا المقدر من كفاءة ومرونة 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.

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] عيسى، 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 وLLS لتقدير 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, New York: Chapman & Hall.

[10] Leigh, J.P. and Schembri, M., (2004), “Instrumental variables technique: cigarette price provided better estimate of effects of smoking”, Journal of Clinical Epidemiology, 57(3), 284-293.

[11] Liang , H. , Hardle , W. , & Carroll, R. J. (1999), "Estimation in a Semiparametric Partially Linear Errors-in-Variables Model, The Annuals of Statistics, Vol. 27, No. 5 (Oct., 1999), pp. 1519-1535

[12] Shang, Yi., (2012), " Measurement error adjustment using the simex method: An application to student growth percentiles", journal of educational measurement, Vol. 49, No. 4, pp. 446–465.

[13] Yao, F., & Zhang, J. (2010). Efficient semiparametric instrumental variable estimation (No. 10-11). Morgantown, W Va. West Virginia Univ., Dep. of Economics.

Estimation of Semiparametric Regression Function in Presence of Measurement Error

Prof. Dr. Munaf Yousif Hmood
munaf_yousif@yahoo.com
University of Baghdad - College of Administration and Economy
Ahmed Mahdi Jaber
ahmedalwasty01@gmail.com
University of Baghdad - College of Administration and Economy

Abstract: *In recent years, the attention of researchers has increased on 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 that has the potential to deal with the case of dimensionality in non-parametric models. This 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 this methods are Simex method and instrument variable method and Quassi-likelihood method and will compare between these 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 showed that the instrument variable is better than simex and Quassi-likelihood methods at different sample sizes and variances that been used.

Keywords: Semi-parametric model, Measurement error, Simex estimator, Instrument variable, Quassi likelihood estimator