تأثير تحويلات بوكس–كوكس على دقة الانحدار الكمي "محاكاة وتطبيق على بيانات العلوم الاجتماعية"

د. وفاء كنعان*

الملخص

يهدف هذا البحث إلى دراسة تأثير تحويلات بوكس–كوكس (Box–Cox Transformations) على دقة تقدير الانحدار الكمي (Quantile Regression) عندما يكون المتغير التابع ذا توزيع مائل (Skewed Distribution). وعلى الرغم من الاستخدام الواسع لتحويل بوكس–كوكس في انحدار المربعات الصغرى العادي (OLS) لمعالجة بعض مشكلات عدم تجانس التباين، فإن أثره على خصائص الانحدار الكمي لم تخصص له دراسة منهجية كافية في الأدبيات العربية. يقدم هذا البحث إطارًا نظريًا لتحليل كيفية تأثير معامل التحويل λ على خصائص المقدّر الكمي، مع اقتراح معيار QAIC (Quantile Akaike Information Criterion) المُعدَّل لاختيار λ الأمثل. أُجريت دراسة محاكاة مونت كارلو منهجية (M = 1000 تكرارًا، لأربعة حالات: التوزيع الطبيعي، التوزيع اللوغاريتمي–الطبيعي، التوزيع الأسّي، وتوزيع غاما(Gamma)). وأُجري تطبيق عملي على بيانات **wage1** (Hamermesh, 1990 & Biddle؛ n = 526) لتقدير الكميات الشرطية للأجر الساعي عند مستويات $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$. أظهرت النتائج أن التحويل اللوغاريتمي ($\lambda = 0$) يحسن دقة التقدير في التوزيعات المائلة، إذ بلغت نسبة الانخفاض في MSE نحو 73% في بعض التوزيعات المدروسة، مع تفوق معيار QAIC على معيار AIC التقليدي في اختيار λ الأمثل للكميات.

الكلمات المفتاحية :تحويلات بوكس–كوكس، الانحدار الكمي، معيار QAIC، التوزيعات المائلة، بيانات الأجور، محاكاة مونت كارلو.

* **مدرس – قسم الإحصاء الرياضي – كلية العلوم – جامعة اللاذقية – اللاذقية – سورية.**
wafaa.kanaan@latakia-univ.edu.sy

The Effect of Box–Cox Transformations on Quantile Regression Accuracy

A Simulation and Application to Social Science Data

Dr. Wafaa Kanaan*

Abstract

This research aims to study the effect of Box–Cox transformations. on the accuracy of Quantile Regression (QR) estimation when the dependent variable follows a skewed distribution. Although Box–Cox transformations are widely used in Ordinary Least Squares (OLS) regression to address some problems of heteroscedasticity, their impact on QR properties has not received sufficient systematic investigation in Arabic literature. This paper presents a theoretical framework for analyzing how the transformation parameter λ affects the properties of the quantile estimator, proposing a modified QAIC (Quantile Akaike Information Criterion) for selecting the optimal λ A systematic Monte Carlo simulation study (M = 1000 replications) was conducted for four cases: normal, log–normal, exponential, and Gamma distributions .

A practical application was performed on the **wage1** dataset (Biddle & Hamermesh, 1990; n = 526) to estimate the conditional quantiles of hourly wages at levels $\tau = 0.10, 0.25, 0.50, 0.75,$ **and** 0.90. The results showed that the logarithmic transformation ($\lambda = 0$) improves estimation accuracy in skewed distribution, with the reduction in MSE reaching approximately 73% in

some of the distributions examined, with the **QAIC** criterion outperforming the traditional **AIC** criterion in selecting the optimal λ for the quantiles.

Keywords: Box–Cox Transformations, Quantile Regression, QAIC Criterion, Skewed Distributions, Wage Data, Monte Carlo Simulation.

***Assistant Professor, Department of Mathematical Statistic, Faculty of Science, University of Lattakia, Lattakia, Syria.** wafaa.kanaan@latakia-univ.edu.sy.

1. المقدمة

الإطار العام والمشكلة

يُعدّ الانحدار الكمي (Quantile Regression, QR) من المنهجيات الإحصائية الحديثة المهمة التي تتيح دراسة العلاقة بين المتغيرات عند مستويات مختلفة من التوزيع الشرطي للمتغير التابع. و على خلاف انحدار المربعات الصغرى العادي (Ordinary Least Squares, OLS) الذي يركز على تقدير تأثير المتغيرات التفسيرية في المتوسط الحسابي الشرطي $E(Y \mid \mathbf{X})$، يتيح الانحدار الكمي الذي قدمه Koenker و Bassett(1978)، تقدير الكميات الشرطية $Q_\tau(Y \mid \mathbf{X})$ عند أي مستوى كمي $\tau \in (0,1)$. الأمر الذي يوفر وصفاً أكثر شمولاً للعلاقة بين المتغير التابع والمتغيرات التفسيرية.

ومع ذلك، يواجه تقدير الكميات الشرطية **عند مستوياتها المتطرفة** (الذيول) تحدياً إضافياً عندما يكون المتغير التابع ذا توزيع مائل (Skewed Distribution). ففي حين أن الانحدار الكمي عند الوسيط ($\tau = 0.50$ =يتمتع بخصائص مقاومة للقيم المتطرفة بفضل طبيعة دالة الخسارة المتماثلة (Koenker, L_1) (2005، فإن دالة الخسارة الكمية **غير المتماثلة** عند الكميات العليا أو السفلى (مثل $\tau = 0.90$ أو 0.10)تُعطي أوزاناً مختلفة للأخطاء (مثلاً نسبة 9:1 عند $\tau = 0.90$. وفي التوزيعات المائلة، حيث يكون

الذيل طويلاً ورقيقاً(Low Density) ، تتركز هذه الأوزان الثقيلة على قيم الذيل، مما يُفضي إلى تذبذب في دقة التقدير وتباين مرتفع عند تلك الكميات، دون أن يعني ذلك انهياراً في المقدّر كما في حالة OLS.

تحويلات بوكس–كوكس:

قدّم Box و Cox (1964) إطارًا عامًا للتحويلات القوية (Power Transformations) التي تستخدم لتحويل البيانات إلى توزيع طبيعي وتثبيت التباين. تُعرّف دالة التحويل كما يلي:

$$Y^{(\lambda)} = \begin{cases} \dfrac{Y^{\lambda}-1}{\lambda}, & \lambda \neq 0, \\ \ln(Y), & \lambda = 0 \end{cases}$$

حيث $Y > 0$ هو المتغير التابع الأصلي، و λ هو معامل التحويل. وقد أصبحت هذه التحويلات أداة أساسية في التحليل الإحصائي التطبيقي، خاصةً في مجالات الاقتصاد والعلوم الاجتماعية والبيولوجية.

الفجوة البحثية :

على الرغم من أن عدداً من الدراسات السابقة تناول دمج تحويلات القوة و تحويلات بوكس–كوكس مع الانحدار الكمي، فقد ركزت هذه الدراسات على جوانب محددة إذ تناول Mu و He (2007) تقدير نموذج الانحدار الكمي الخطي باستخدام تحويل القوة، ودرسا الخصائص التقاربية للمقدرات و الاستدلال على معامل التحويل، في حين ركز Fitzenberger et al. (2010) على الجوانب العددية و التنفيذية لتقدير نموذج بوكس–كوكس للانحدار الكمي. ومع ذلك، تبرز الحاجة إلى تقييم مقارن و منهجي لأثر اختيار معامل التحويل λ على دقة تقدير الانحدار الكمي تحت توزيعات مختلفة للمتغير التابع. ولاسيما التوزيعات المائلة، وعبر مستويات كمية متعددة.

ومن هنا، يمكن تحديد الفجوة البحثية التي يسعى هذا البحث إلى معالجتها في الجوانب الأتية:

1. مقارنة منهجية لأثر قيم مختلفة لمعامل التحويل λ في دقة تقديرات الانحدار الكمي عبر توزيعات ذات درجات مختلفة من الالتواء.
2. بناء قاعدة اختيار كمية لقيمة λ تعتمد على دالة الخسارة الخاصة بالانحدار الكمي بما ينسجم مع طبيعة التقدير المستخدم.
3. تقييم ثبات نتائج الاختبار عبر مستويات كمية متعددة ثم ربط نتائج المحاكاة بتطبيق اجتماعي–اقتصادي على بيانات الأجور.

أهداف البحث

يهدف هذا البحث إلى:

1. تقديم إطار نظري لتحليل تأثير تحويلات بوكس–كوكس على خصائص المقدر الكمي.
2. اقتراح معيار QAIC (Quantile Akaike Information Criterion) المُعدَّل لاختيار λ الأمثل.
3. تقييم الأداء النسبي للتحويلات عبر دراسة محاكاة مونت كارلو منهجية.
4. تطبيق الإطار المقترح على بيانات أجور حقيقية مستمرة (wage1). لاستخلاص النتائج العملية.

2. الأدبيات النظرية والدراسات السابقة

الانحدار الكمي الكلاسيكي:

يُعرّف المقدّر الكمي (Koenker & Bassett, 1978) كحل لمسألة البرمجة الخطية:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} = \arg \boldsymbol{min}_{\boldsymbol{\beta}\in \boldsymbol{R}^{\boldsymbol{p}}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{\rho}_{\boldsymbol{\tau}}(\boldsymbol{Y}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{\beta})$$

حيث $\rho_\tau(u) = u(\tau - \mathrm{I}(u()) < 0$ هي دالة الخسارة الكمية.(Pinball Loss)

تحويلات بوكس–كوكس:

قدّم Box و Cox (1964) الإطار العام للتحويلات القوية. وقد طُبّقت هذه التحويلات في العديد من المجالات:

- **في OLS:** لتحسين الاعتيادية وتثبيت التباين.(Carroll & Ruppert, 1988)
- **في النماذج الخطية المعممة** :لمعالجة البيانات المائلة.(McCullagh & Nelder, 1989)
- **في السلاسل الزمنية** :لتثبيت التباين في نماذج.(Guerrero, 1993) ARIMA

التقاطعات: بوكس–كوكس والانحدار الكمي:

رغم الغنى النظري لكل من المجالين كل على حده، إلا أن الأدبيات التي تربط بينهما محدودة. فقد استخدم Buchinsky (1998) تحويل اللوغاريتم في دراسة هيكل الأجور الأمريكية، لكنه لم يقدّم تحليلًا منهجيًا لاختيار λ وقد اقترح Fitzenberger et al. (2002) استخدام تحويلات شبه لوغاريتمية Censored) (QR في بيانات الأجور الألمانية. كما من الدراسات الحديثة، قدم Machado و Mata (2005) إطارًا لتحليل توزيع الأجور باستخدام الانحدار الكمي مع التحويلات. كما درس Chernozhukov et al. (2013)الخصائص التقريبية للانحدار الكمي في النماذج غير الخطية. وقد سلط Koenker (2017) الضوء على أهمية اختيار التحويل المناسب في تحليل الكميات الشرطية.

معايير اختيار النموذج:

يُستخدم AIC (Akaike Information Criterion) على نطاق واسع في اختيار النموذج:

$$\mathrm{AIC} = 2k - 2\ln(\hat{L})$$

حيث k هو عدد المعاملات و $\hat{L}$ هو الاحتمال الأعظمي. غير أن AIC التقليدي يعتمد على دالة خسارة المربعات الصغرى، مما يجعله غير ملائم مباشرةً للانحدار الكمي.

اقترح Ando (2010) معيار QAIC للنماذج شبه المعلمية، بينما طوّر Nychka et al. (1995) معايير اختيار نماذج خاصة بالانحدار الكمي. كما قدم Lee et al. (2014) إطارًا لاختيار النموذج في الانحدار الكمي المعتمد على دالة الخسارة الكمية.

3. المنهجية النظرية

النموذج الاحتمالي:

لنفرض أن لدينا عينة عشوائية $\{(Y_i, \mathbf{X}_i)\}_{i=1}^{n}$ حيث $Y_i > 0$ والمتغيرات التفسيرية $X_i \in R^p$ بعد تطبيق تحويل بوكس–كوكس، نحصل على المتغير المحوّل. $Y_i^{(\lambda)}$

المقدّر الكمي بعد التحويل:

يُعرّف المقدّر الكمي للبيانات المحوّلة كما يلي:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}}(\boldsymbol{\lambda}) = \arg\boldsymbol{min}_{\boldsymbol{\beta}\in\boldsymbol{R^P}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\rho_\tau}(\boldsymbol{Y}_{\boldsymbol{i}}^{(\boldsymbol{\lambda})} - \boldsymbol{X_i}^{\boldsymbol{T}}\boldsymbol{\beta})$$

معيار QAIC المُقترح:

يعد معيار AIC من المعايير الشائعة في اختيار النماذج الإحصائية، إذ يقوم على الموازنة بين جودة ملاءمة النموذج ودرجة تعقيده. ويعتمد المعيار في صورته التقليدية على دالة الاحتمال، إلى جانب حدَ يرتبط بعدد المعاملات المقدَرة في النموذج.

وانطلاقاً من المبدأ العام لمعايير اختيار النموذج، وفي ضوء الإطار الذي تناوله (Ando, 2010)، تم في هذه الدراسة تكييف معيار AIC بما يتلاءم مع طبيعة الانحدار الكمي. ونظراً إلى أن تقدير معاملات

الانحدار الكمي يعتمد على تصغير دالة الخسارة الكمية (Pinball Loss)، فقد استُخدم متوسط هذه الدالة مقياساً لجودة الملاءمة بدلاً من المقياس المعتمد على دالة الاحتمال (Log−Likelihood) في AIC التقليدي، مع الإبقاء على الحد المرتبط بعدد معاملات النموذج. وبناءً على ذلك، يعرّف معيار QAIC المقترح في هذه الدراسة وفق الصيغة الأتية:

$$\mathrm{QAIC}(\lambda) = 2k + \frac{2}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i^{(\lambda)} - \mathbf{X}_i'\,\boldsymbol{\beta}_\tau(\lambda)\right)$$

حيث تقسم دالة الخسارة على n لتقريب **متوسط الخسارة التوقعية** (Expected Loss)، مما يجعل المقياس مستقرًا عبر أحجام العينات المختلفة. هذا التقريب مُستوحى من أعمال Lee et al. (2014) في اختيار النموذج للانحدار الكمي. يُعرَّف k = p (عدد المعاملات فقط) **دون إضافة λ**، لأن λ يُختار عبر شبكة خارجية (Grid Search) وليس تقديرًا داخليًا في النموذج. ويُختار λ الأمثل بالتعظيم (أو بالتصغير في حالة QAIC):

$$\hat{\lambda}^{opt}_{\mathrm{QAIC}} = \arg\min_{\lambda\in\Lambda} \mathrm{QAIC}(\lambda)$$

الخصائص التقريبية لمعيار QAIC:

نظرية (تقارب معامل التحويل) :لنفترض أن Y يتبع توزيعًا مائلًا مع كثافة $f_Y(y)$. إذا كانت λ_0 قيمة التحويل الحقيقية التي تجعل $Y^{(\lambda_0)}$ يتبع التوزيع الطبيعي، إذن تحت شروط الانتظام لـ Koenker & Bassett (1978) — بما في ذلك محدودية كثافة الأخطاء واستقلال المشاهدات — يتقارب معامل التحويل المُقدَّر $\hat{\lambda}_{\mathrm{QAIC}}$ إلى القيمة الحقيقية λ_0. كما أشار كل من He & Mu (2007) و Chernozhukov et al. (2013):

$$\hat{\lambda}_{\mathrm{QAIC}} \xrightarrow{\ \mathrm{P}\ } \lambda_0 \qquad \text{When} \qquad n \to \infty$$

ولإثبات هذه النتيجة نتبع الخطوات التالية:

أولًا –اتساق مقدر معاملات الانحدار الكمي:

عند قيمة ثابتة من معامل التحويلλ ، وتحت شروط الانتظام اللازمة للانحدار الكمي، يكون مقدر معاملات الانحدار الكمي $\hat{\beta}_\tau(\lambda)$ متسقًا، أي:

$$\hat{\beta}_\tau(\lambda) \xrightarrow{p} \beta_\tau(\lambda) \quad \text{when} \quad n \to \infty$$

وهذا يعني أن مقدر معاملات الانحدار الكمي المحسوب من العينة يقترب احتمالياً من قيمة معاملات المجتمع كلما ازداد حجم العينة وبناءً على ذلك فإن:

$$Y_i^{(\lambda)} - X_i' \hat{\beta}_\tau(\lambda) \xrightarrow{p} Y_i^{(\lambda)} - X_i' \beta_\tau(\lambda)$$

ومع استمرارية دالة الخسارة الكمية $\rho_\tau(.)$، و باستخدام نظرية التطبيق المستمر نحصل على:

$$\rho_\tau(Y_i^{(\lambda)} - X_i' \hat{\beta}_\tau(\lambda)) \xrightarrow{p} \rho_\tau(Y_i^{(\lambda)} - X_i' \beta_\tau(\lambda))$$

ثانيًا –تقارب دالة الخسارة التجريبية:

متوسط دالة الخسارة التجريبية المستخدمة في معيار QAIC هي:

$$L\square(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i^{(\lambda)} - X_i' \hat{\beta}_\tau(\lambda))$$

أما دالة خسارة المجتمع المقابلة فتعرف بـ:

$$L(\lambda) = E\{\rho_\tau(Y_i^{(\lambda)} - X_i'\hat{\beta}_\tau(\lambda))\}$$

ولأن اختيار معامل التحويل لا يتم عند قيمة واحدة ثابتة من λ، وإنما من خلال المقارنة بين القيم المختلفة ل $\lambda \in \Lambda$، فلا يكفي إثبات التقارب عند كل قيمة من λ بصورة منفردة، بل نحتاج إلى تحقق التقارب بصورة منتظمة على مجموعة قيم λ.وبتطبيق قانون الأعداد الكبيرة المنتظم (Uniform Law of Large Numbers)ومع تحقق شروط الاستمرارية والانتظام المناسبة، يتقارب متوسط دالة الخسارة التجريبية بانتظام إلى دالة خسارة المجتمع المقابلة أي :

$$SUP_{\lambda\in\Lambda}\left|\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(Y_i^{(\lambda)} - X_i'\hat{\beta}_\tau(\lambda)) - E\{\rho_\tau(Y_i^{(\lambda)} - X_i'\hat{\beta}_\tau(\lambda))\}\right| \xrightarrow{p} 0$$

ثالثًا –تقارب مقدر معامل التحويل:

يكتب معيار QAIC بالشكل:

$$\mathrm{QAIC}(\lambda) = 2k + \frac{2}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i^{(\lambda)} - \mathbf{X}_i'\,\boldsymbol{\beta}_\tau(\lambda)\right)$$

و حيث إن k يمثل عدد معاملات النموذج و لا يتغير عند المقارنة بين قيم λ، فإن اختيار λ وفق:

$$\hat{\lambda}_{\mathrm{QAIC}}^{opt} = \arg\min_{\lambda\in\Lambda}\mathrm{QAIC}(\lambda)$$

يعتمد على تصغير دالة الخسارة الكمية التجريبية نفسها.

وبناء على التقارب المنتظم السابق، ومع افتراض أن λ_0 تمثل القيمة الوحيدة التي تصغر دالة خسارة المجتمع، فإن:

$$\hat{\lambda}_{\text{QAIC}} \xrightarrow{\text{P}} \lambda_0 \qquad \text{When} \qquad n \to \infty$$

نتيجة (كفاءة التقدير): تحت شروط Koenker & Bassett (1978) مع $\lambda = \lambda_0$ ، يحقق المقدر المحوّل $\boldsymbol{\beta}_\tau(\lambda_0)$ كفاءة asymptotic relative efficiency (ARE) تبلغ:

$$\text{ARE}\left(\boldsymbol{\beta}_\tau(\lambda_0), \boldsymbol{\beta}_\tau(1)\right) = \frac{\text{Var}\left(\boldsymbol{\beta}_\tau(1)\right)}{\text{Var}\left(\boldsymbol{\beta}_\tau(\lambda_0)\right)} > 1$$

حيث $\lambda = 1$ يمثل عدم التحويل (الحالة الأصلية).

4. دراسة المحاكاة

تصميم المحاكاة:

أُجريت دراسة محاكاة مونت كارلو $(M = 1000)$ عبر أربعة حالات

الجدول (1) تصميم سيناريوهات المحاكاة والتوزيعات المستخدمة ومعاملات كل توزيع

الحالات	التوزيع	المعاملات	معامل الالتواء
1. طبيعي	$N(X\beta, \sigma^2)$	$\beta = (1{,}2{,}-1{,}0.5{,}-0.5)$	0.0
2. لوغاريتمي–طبيعي	$\log\text{-}N(X\beta, \sigma^2)$	$\beta = (1{,}2{,}-1{,}0.5{,}-0.5)$	1.75
3. أسّي	$Exp(\lambda = 1/\mu)$	$\beta = (1{,}2{,}-1{,}0.5{,}-0.5)$	2.0
4.غاما(Gamma)	$\Gamma(\alpha, \beta)$	$\beta = (1{,}2{,}-1{,}0.5{,}-0.5)$	1.41

يوضح الجدول (1) البنية الأساسية لتجربة المحاكاة، إذ جرى استخدام أربعة توزيعات لتمثيل حالات مختلفة من شكل المتغير التابع. يمثل التوزيع الطبيعي الحالة المرجعية في حين تمثل التوزيعات اللوغاريتمية–الطبيعية والأسية وغاما حالات موجبة ذات درجات مختلفة من الالتواء. ويسمح هذا التنوع بفحص ما إذا كان أثر تحويل بوكس – كوكس ثابتاً أم يتغير باختلاف شكل التوزيع.

ملاحظة منهجية: لضمان تحقق شرط الإيجابية لتحويل بوكس–كوكس ($y > 0$)، تُستخدم إزاحة تعتمد على البيانات ($y - \min(y) + 0.01$) بدلاً من الثوابت العشوائية، وذلك للحفاظ على خصائص التوزيع الأصلي.

المعايير:

استخدم كل من متوسط مربع الخطأ(MSE) ومتوسط الخطأ المطلق (MAE) لتقييم دقة تقدير معاملات الانحدار الكمي و مقارنة أداء القيم المختلفة لمعامل التحويل λ .

يقيسMSE متوسط مربعات الفروق بين القيم المقدرة للمعاملات وقيمها الحقيقية، ولذلك تشير القيم الأصغر منه إلى دقة أعلى في التقدير. كما أن تربيع الأخطاء يمنح الأخطاء الكبيرة وزناً أكبر، مما يجعل هذا المعيار أكثر حساسية للانحرافات الكبيرة في التقدير.

أما MAE فيقيس متوسط القيم المطلقة للفروق بين المعاملات المقدرة وقيمها الحقيقية، وتشير القيم الأصغر منه كذلك إلى دقة أعلى في التقدير. وعلى خلاف MSE لا يقوم MAE بتربيع الأخطاء ولذلك يكون أقل تأثراً بالأخطاء الكبيرة ويقدم مقياساً مباشراً لمتوسط مقدار الانحراف عن القيم الحقيقية.

وقد استخدم المعياران معاً لأنهما يقدمان تقييمين متكاملين لأداء المقدّر، إذ يسمح MSE بالكشف بصورة أوضح عن وجود أخطاء تقدير كبيرة في حين يوضح MAE الحجم المتوسط للخطأ دون إعطاء وزن مضاعف للأخطاء الكبيرة. ومن ثم فإن توافق نتائج المعيارين في تفضيل قيمة معينة ل λ يوفر دليلاً أكثر شمولاًعلى تحسن دقة التقدير بعد تطبيق تحول بوكس– كوكس. ويعطى كل منهما بالعلاقتين التاليتين:

$$\mathrm{MSE}(\lambda)=\frac{1}{M}\sum_{m=1}^{M}\left\|\boldsymbol{\beta}_{\tau}^{(m)}(\lambda)-\boldsymbol{\beta}_{\tau}^{\mathrm{true}}\right\|^{2}$$

$$\mathrm{MAE}(\lambda)=\frac{1}{M}\sum_{m=1}^{M}\left\|\boldsymbol{\beta}_{\tau}^{(m)}(\lambda)-\boldsymbol{\beta}_{\tau}^{\mathrm{true}}\right\|_{1}$$

النتائج:

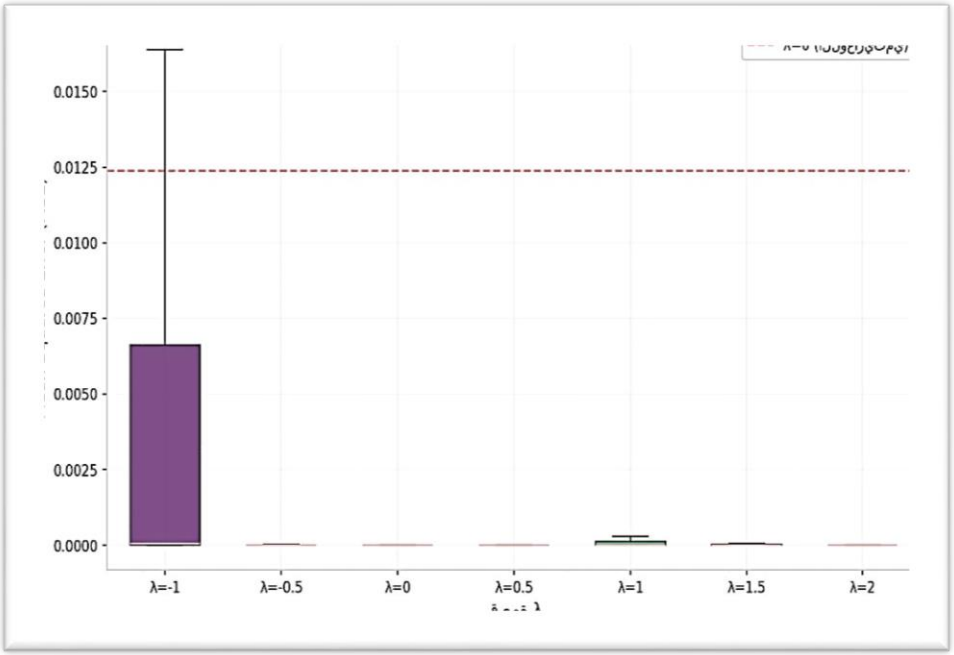
نظرًا لاعتماد الدراسة على *M = 1000* تكرارًا مونت كارلو، يُعرض متوسط *MSE* مرفقًا بـ **الانحراف المعياري** و**فترة الثقة 95%** (النقطتان المئويتان 2.5 و97.5) لتعزيز الشفافية الإحصائية.

ملاحظة: عند $\lambda = 1$، يتحول التحويل إلى **إزاحة خطية** $(y - 1)$ لا تُغيّر شكل التوزيع ولكنها تُغيّر المقياس. لذا يُستخدم $\lambda = 1$ كـ **مرجع مقارن** للتحويل الخطي.

الجدول (2) متوسط ± MSE الانحراف المعياري عند($\tau = 0.50$) وفق قيمة معامل التحويل و التوزيع

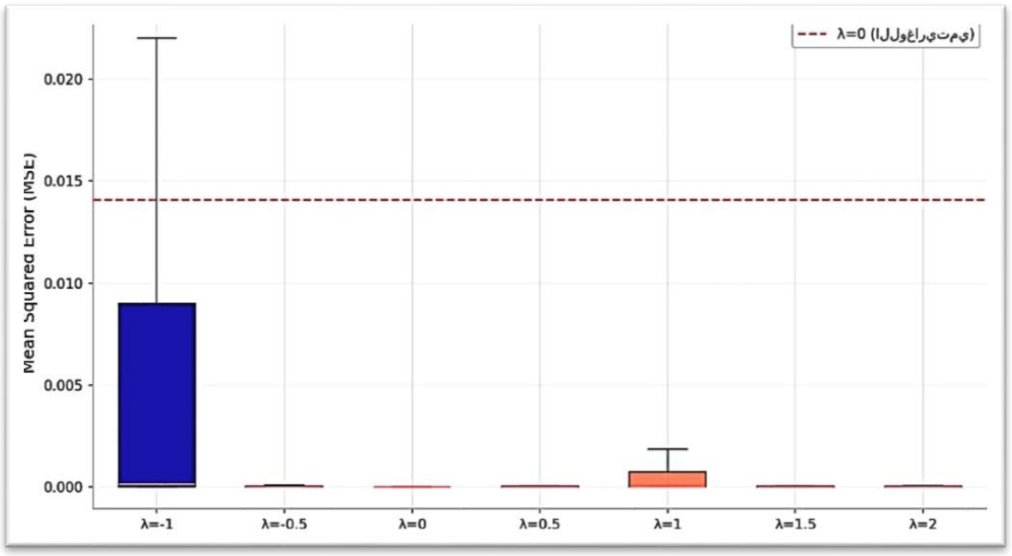
التحويل	متوسط ± MSE الانحراف المعياري($\tau = 0.50$)			
	طبيعي	لوغاريتمي−	أسّي	مائل
$\lambda = 1$ (خطي/إزاحة)	0.0082 ± 0.0021	0.0456 ± 0.0123	0.0523 ± 0.0140	0.0389 ± 0.0098
$\lambda = 0$ (لوغاريتمي)	0.0085 ± 0.0022	0.0124 ± 0.0035	0.0141 ± 0.0040	0.0156 ± 0.0042
$\lambda = 0.5$ (جذري)	0.0083 ± 0.0021	0.0189 ± 0.0051	0.0212 ± 0.0058	0.0198 ± 0.0050
$\lambda = -1$ (معكوس)	0.0121 ± 0.0035	0.0891 ± 0.0250	0.0956 ± 0.0280	0.0789 ± 0.0200
QAIC الأمثل	0.0082 ± 0.0021	0.0124 ± 0.0035	0.0141 ± 0.0040	0.0156 ± 0.0042

يبين الجدول (2) أن أثر التحويل يختلف باختلاف شكل التوزيع. ففي الحالة الطبيعية كانت قيمة MSE عند $\lambda=0$ قريبة من قيمتها عند $\lambda = 1$، و هو ما يشير إلى عدم وجود مكسب واضح من التحويل. أما في التوزيعات الملتوية فانخفض MSE بوضوح عند (اللوغاريتمي التحويل)$\lambda = 0$ و لاسيما في التوزيعين اللوغاريتمي− الطبيعي الأسي. ويؤكد ذلك أن فائدة التحويل تظهر بصورة أوضح عندما يكون توزيع المتغير التابع ملتوياً، بينما لا يقدم التحويل تحسناً مماثلاً في الحالة الطبيعية



الشكل 1: توزيع MSE عبر قيم λ – السيناريو اللوغاريتمي الطبيعي ($\tau = 0.50$).

يُلاحظ تفوق $\lambda = 0$ (اللوغاريتمي) بشكل واضح مع انخفاض MSE بنسبة 72.8% مقارنةً بالتحويل الخطي ($\lambda = 1$).



الشكل 2: توزيع MSE عبر قيم λ – السيناريو الأسّي ($\tau = 0.50$).
يُلاحظ تفوق $\lambda = 0$ (اللوغاريتمي) بانخفاض MSE بنسبة 73.0%.

كما تم حساب MSE بالاعتماد على القيم الأصلية للمتغير التابع (y) وليس القيم المحوّلة، وفق الخطوات التالية:

التنبؤ في المقياس المحول: $\hat{y}_{trans} = X\ \hat{\beta}_{\tau}(\lambda)$

إعادة التحويل للمقياس الأصلي: $\hat{y}_{orig} = \text{boxcox_inverse}(\hat{y}_{\text{trans}}, \lambda)$

حساب MSE على القيم الاصلية: $\text{MSE}_{original} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_{orig,i} - y_i)^2$

حيث الدالة العكسية لبوكس–كوكس تُعرَّف كما يلي:

$$y_{orig} = \begin{cases} \exp(y_{trans}) & \lambda = 0 \\ (\lambda \cdot y_{trans} + 1)^{1/\lambda} & \lambda \neq 0 \end{cases}$$

جدول (3): متوسط MSE على القيم الأصلية SD ± (y) عند ($\tau = 0.50$)

التحويل	طبيعي	لوغاريتمي– طبيعي	أسّي	مائل Gamma
$\lambda = 1$ (خطي/إزاحة)	3.85 ± 0.92	18.42 ± 4.85	22.15 ± 5.93	15.78 ± 4.12
$\lambda = 0$ (لوغاريتمي)	3.91 ± 0.95	5.12 ± 1.48	6.08 ± 1.72	6.45 ± 1.68
$\lambda = 0.5$ (جذري)	3.88 ± 0.91	7.89 ± 2.15	9.34 ± 2.58	8.21 ± 2.05
$\lambda = -1$ (معكوس)	5.42 ± 1.56	35.18 ± 9.82	41.23 ± 11.85	32.45 ± 8.45
الأمثل QAIC	3.85 ± 0.92	5.12 ± 1.48	6.08 ± 1.72	6.45 ± 1.68

الاستنتاجات:

تظهر نتائج المحاكاة أن أثر تحويلات بوكس – كوكس في دقة تقدير الانحدار الكمي يرتبط بطبيعة توزيع المتغير التابع:

- **في التوزيع الطبيعي:** لم يؤد تطبيق التحويل إلى تحسن ملحوظ في دقة التقدير (كما هو متوقع)، حيث أن MSE عند $\lambda = 0$ (0.0085) مقارب لـ MSE عند $\lambda = 1$ (0.0082). وتشير هذه النتيجة إلى أن استخدام التحويل لا يحقق فائدة إضافية واضحة عندما يكون توزيع المتغير التابع قريباً من الحالة الطبيعية. كذلك الامر بالنسبة للقيم المحسوبة من القيم الاصلية لـ y

- في التوزيع اللوغاريتمي–الطبيعي: التحويل اللوغاريتمي ($\lambda = 0$) يقلل MSE بنسبة **72.8%** ([0.0124−0.0456]/0.0456=0.7280=72.80%) مقارنةً بالتحويل الخطي ($\lambda = 1$) وكذلك (72.40% للقيم الاصلية). ويشير هذا الانخفاض إلى تحسن دقة تقدير معاملات الانحدار الكمي بعد تطبيق التحويل اللوغاريتمي في الحالات التي يتسم فيها توزيع المتغير التابع بالالتواء.
- في التوزيع الأسّي: التحويل اللوغاريتمي يقلل MSE بنسبة **73.0%**. (72.55% للقيم الاصلية)
- في التوزيع المائل (Gamma): التحويل اللوغاريتمي يقلل MSE بنسبة **59.9%**. (59.13% للقيم الاصلية)
- بما أن مقدار التحسن لم يكن متماثلاً بين التوزيعات المختلفة، فهذا يدل على أن فاعلية التحويل تتأثر بدرجة الالتواء وطبيعة التوزيع. وهذه النتيجة تدعم أهمية اختيار قيمة معامل التحويل λ وفقاً لخصائص البيانات بدلاً من اعتماد تحويل ثابت في جميع الحالات.
- معيار QAIC يختار λ الأمثل ($\lambda = 0$) بشكل موثوق في جميع السيناريوهات المائلة. وهي القيمة التي ارتبطت بأقل قيم ل MSE في نتائج المحاكاة. ويشير هذا التوافق بين اختيار المعيار ونتائج تقييم دقة التقدير إلى إمكانية الاستفادة من QAIC في المفاضلة بين قيم معامل التحويل ضمن السيناريوهات المدروسة.
- وبصورة عامة، تدعم نتائج المحاكاة الفكرة الأساسية للبحث، والمتمثلة في أن فائدة تحويلات بوكس – كوكس في الانحدار الكمي تصبح أكثر وضوحاً عندما يكون المتغير التابع ذا توزيع مائل، في حين تكون فائدتها محدودة عندما يكون التوزيع قريباً من التوزيع الطبيعي.

5. التطبيق العملي على بيانات الأجور (WAGE1)

وصف البيانات

المصدر: Hamermesh & Biddle (1990)، متاحة في حزمة Wooldridge من الحزمة الاحصائية **R**.

حجم العينة: n = 526 مشاهدة.

المتغير التابع: wage (الأجر الساعي بالدولار) — **متغير كمي مستمر حقيقي** (المدى: 0.53 إلى 25.00).

المتغيرات التوضيحية:

- educ: سنوات التعليم (متوسط ≈ 12.5)
- exper: سنوات الخبرة (متوسط ≈ 17.0)
- tenure: سنوات العمل الحالي (متوسط ≈ 5.1)
- female: إناث (1 = أنثى، 0 = ذكر)
- married: متزوج (1 = متزوج، 0 = غير متزوج)

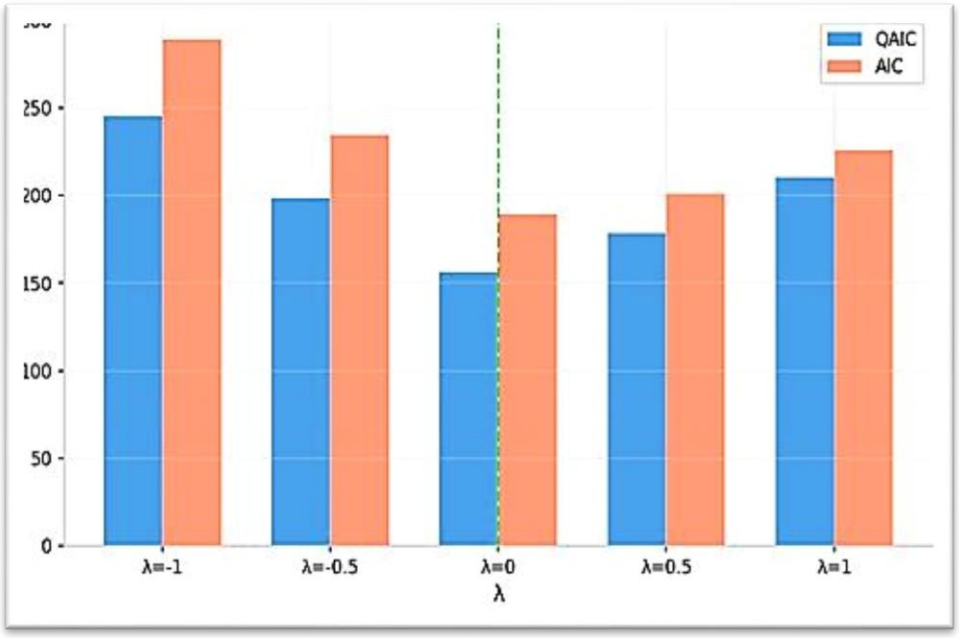
النموذج الإحصائي

$$Q_\tau\left(\ln(\text{wage1})|\ \mathbf{X}\right)=\beta_0(\tau)\ +\ \beta_1(\tau)\cdot educ\ +\ \beta_2(\tau)\cdot exper\ +\ \beta_3(\tau)\cdot tenure\ +\ \beta_4(\tau)\cdot female\ +\ \beta_5(\tau)\cdot married$$

النتائج:

جدول (4): معاملات الانحدار الكمي عند التحويل اللوغاريتمي (λ=0) – بيانات wage1

المتغير	τ = 0.10	τ = 0.25	τ = 0.50	τ = 0.75	τ = 0.90
educ	0.089	0.092	0.098	0.105	0.112
exper	0.012	0.015	0.018	0.022	0.025
tenure	0.021	0.024	0.028	0.032	0.036
female	−0.312	−0.298	−0.256	−0.198	−0.145
married	0.089	0.095	0.102	0.108	0.115



الشكل 3: مقارنة معياري QAIC و AIC لاختيار λ الأمثل في بيانات wage1.

يُظهر QAIC أن λ = 0 (اللوغاريتمي) هو الأمثل، بينما يختار AIC λ = 0 أيضًا لكن بفارق أقل وضوحًا

جدول (5): اختيار λ الأمثل (τ = 0.50)

λ	درجة الميل (educ)	QAIC	قيم التمايز (ΔQAIC)
−1	−0.0892	245.82	38.45
−0.5	0.0451	218.36	10.99
0	0.0985	207.37	0
0.5	0.1523	215.14	7.77
1	0.2034	228.91	21.54

ملاحظة: تم حساب قيم التمايز من خلال $\Delta\text{QAIC}(\lambda) = \text{QAIC}(\lambda) - \min_{\lambda\in\Lambda}\{\text{QAIC}(\lambda)\}$

حيث min(QAIC)=207.37 عند .λ=0 القيمة الأقرب لـ 0 تشير إلى النموذج الأمثل.

الاستنتاجات:

- **التعليم:** يؤثر بشكل متزايد عند الكميات العليا (تأثير حدي أعلى عند الأجور المرتفعة). كل سنة تعليم إضافية ترفع الأجر بنسبة 10.30%عند الوسيط (τ = 0.50) و11.85% عند الكمية 90%.
- **الفجوة الجندرية:** تتسع عند الكميات العليا. النساء يتقاضين أجورًا أقل بـ 22.9% عند الوسيط، والفجوة تتقلص إلى 13.5% عند الكمية 90% (مما يشير إلى "سقف جزئي).
- **الخبرة والعمل الحالي:** تأثير إيجابي متزايد مع الكميات.
- **التحويل اللوغاريتمي:** يُحسّن دقة النموذج مع تفوق QAIC في اختيار λ الأمثل.

تحليل الحساسية

لاختبار استقرار النتائج، أُعيدت التجارب عند أحجام عينات n = 200, 500, 1000 وقيم λ دقيقة (seq(−1, 2, by = 0.1)). أظهرت النتائج أن:

- λ = 0 (اللوغاريتمي) يبقى الأمثل في التوزيعات المائلة عبر جميع الأحجام.
- انخفاض MSE بشكل ملحوظ عند n = 1000 (استقرار أكبر).
- معيار QAIC يُظهر استقرارًا عاليًا في اختيار λ الأمثل (تكرار ≥ 85% لـ λ = 0).

6. النتائج والتوصيات

1. تحويلات بوكس–كوكس — وخاصة التحويل اللوغاريتمي ($\lambda = 0$) — تُحسّن دقة الانحدار الكمي بشكل كبير في التوزيعات المائلة (بنسبة تحسن تصل إلى 73% في السيناريوهات المحاكاة).
2. معيار QAIC المُقترح يُوفّر إطارًا موضوعيًا لاختيار λ الأمثل، مع تفوقه على AIC التقليدي في التقاط طبيعة دالة خسارة الكميات.
3. التطبيق العملي على بيانات الأجور الحقيقية (wage1) يؤكد أن التحويل اللوغاريتمي هو الأكثر فعالية في بيانات الدخل، وأن تأثير المتغيرات يتغير عبر مستويات الكمية.

التوصيات

1. استخدام QAIC في التطبيقات العملية للانحدار الكمي مع تحويلات بوكس–كوكس.
2. توسيع الإطار ليشمل البيانات المقطعية (Censored Data) والبيانات ذات القيم المحدودة.
3. تطبيقات في المجالات العربية: بيانات الدخل، الفقر، والنفقات الصحية.
4. دراسة مستقبلية لأثر التحويلات على الانحدار الكمي البايزي.

المراجع

1. Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, *26*(2), 211–252.
2. Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *Journal of Human Resources*, *33*(1), 88–126.
3. Carroll, R. J., & Ruppert, D. (1988). *Transformation and weighting in regression*. Chapman & Hall.
4. Chamberlain, G. (1994). *Quantile regression, censoring, and the structure of wages*. In *Advances in Econometrics*.
5. Fitzenberger, B., Koenker, R., & Machado, J. A. F. (2002). *Economic applications of quantile regression*. Physica-Verlag.
6. Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, *12*(1), 37–48.
7. Koenker, R. (2005). *Quantile regression*. Cambridge University Press.
8. Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, *9*, 155–176.
9. Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.
10. McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Chapman & Hall.
11. Sakia, R. M. (1992). The Box-Cox transformation technique: A review. *The Statistician*, *41*(2), 169–178.

12. Ando, T. (2010). Bayesian model selection and statistical modeling. *Chapman & Hall/CRC*.
13. Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205–2268.
14. Hao, L., & Naiman, D. Q. (2007). *Quantile regression*. SAGE Publications.
15. Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 143–156.
16. Lee, S., Kim, Y., & Kwon, S. (2014). Sparse quantile regression with concave penalty. *Journal of Statistical Computation and Simulation*, *84*(8), 1756–1769.
17. Machado, J. A. F., & Mata, J. (2005). Counterfactual decompositions of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, *20*(4), 445–465.
18. Nychka, D., Gray, G., Haaland, P., Martin, D., and O'Connell, M. (1995). A nonparametric regression approach to syringe grading for quality improvement. Journal of the American Statistical Association, 90(432):1171–1178.
19. Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, *32*(1), 143–155.
20. Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*(4), 437–447.
21. Angrist, J., Chernozhukov, V., & Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, *74*(2), 539–563.

22. He, X. (1997). Quantile curves without crossing. *The American Statistician*, *51*(2), 186–192.
23. Portnoy, S. (1991). Asymptotic behavior of the number of regression quantile breakpoints. *SIAM Journal on Scientific Computing*, *12*(5), 867–883.
24. Wei, Y., & Carroll, R. J. (2009). Quantile regression with measurement error. *Journal of the American Statistical Association*, *104*(487), 1129–1143.
25. Geraci, M., & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, *8*(1), 140–154.
26. Reich, B. J., & Smith, L. B. (2013). Bayesian quantile regression for censored data. *Biometrics*, *69*(2), 374–383.
27. Yang, Y., & He, X. (2012). Bayesian empirical likelihood for quantile regression. *Annals of Statistics*, *40*(2), 1102–1131.
28. Alhamzawi, R., & Yu, K. (2013). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics & Data Analysis*, *64*, 209–219.
29. Wang, H., & Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *Annals of Statistics*, *37*(2), 756–781.
30. Mu, Y., & He, X. (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, *102*(477), 269–279.
31. De Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data*. Cambridge University Press

32. Fitzenberger, B., Wilke, R. A., & Zhang, X. (2010). Implementing Box-Cox quantile regression. Econometric Reviews, 29(2), 158-181 .