معالجة آثار الازدواج الخطي في نموذج الانحدار المتعدد

محمود عبد الحليم جاد محمد هبة فهمي محمد حسين
المعمل المركزي لبحوث التصميم والتحليل الإحصائي– مركز البحوث الزراعية

مقدمة:

يعد تحليل الانحدار أسلوباً هاماً من أساليب الإحصاء التطبيقي عند دراسة العديد من الظواهر فهو مقياس رياضي لمتوسط العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات التوضيحية في العلاقة، ويبني هذا المقياس علي عدة فروض أساسية ينبغي توافرها لإيجاد مقدرات نموذج الانحدار الخطي المتعدد وأبرزها افتراض عدم وجود علاقة خطية تامة بين المتغيرات المستقلة، ويترتب علي إسقاط هذا الفرض حدوث مشكلة الازدواج الخطي *Multicollinearity*، فعندما يرتبط اثنان أو أكثر من المتغيرات التفسيرية بعلاقـــة خطية قوية جداً بدرجة أكبر من ارتباطهم بالمتغير التابع يصبح من الصعب فصل أثر كل متغير تفسيري عن المتغير التابع، وتحدث هذه المشكلة حينما تكون قيمة أحد المتغيرات التفسيرية متساوية لكافة المشاهدات مع متغير تفسيري أو أكثر، وما قد يترتب علي ذلك من آثار سلبية علي معالم الدالة وقدرتها علي تفسير التغيرات التي تطرأ علي المتغير التابع وتشويه هذه المعالم عند تقديرها بطريقة المربعات الصغرى العادية (*OLS*)، فعند صياغة العلاقات المختلفة بين المتغيرات الاقتصادية في صورة دالة عندئذ يمكن قياس شكل واتجاه هذه العلاقات إلا أن وجود الازدواج الخطي بين المتغيرات المستقلة في هذه الدوال قد يتسبب في إعطاء صورة غير واقعية لطبيعة العلاقة بين المتغيرات المستقلة والمتغير التابع في المعادلة بما لا يتفق والنظريات العلمية، وغالباً ما يلجأ الباحثون إلي معالجة هذه المشكلة إما بتكبير حجم العينة أو حذف متغير أو أكثر من المتغيرات التفسيرية التي ترتبط خطياً مع بقية المتغيرات في النموذج، أو بالتحويل المعياري للمتغيرات التفسيرية أو استخدام أساليب تقدير تعتمد علي المعلومات الكمية الأولية.

مشكلة الدراسة:

يترتب علي إغفال الباحثين لمشكلة الازدواج الخطي في نموذج الانحدار المتعدد حدوث تقديرات مضللة لمعالم الدالة ولها أخطاء معيارية كبيرة، وكذلك مدي مساهمة المتغيرات المستقلة في تفسير المتغير التابع، فضلاً عن إضعاف قدرة النموذج الإحصائي علي التنبؤ في المستقبل بالمتغير التابع ومن ثم درجة الثقة في القيم المقدرة.

الهدف من الدراسة:

يهدف البحث إلي دراسة طبيعة مشكلة الازدواج الخطي (*Multicollinearity*)، أسبابها، آثارها، ومن ثم التعرف علي طرق الكشف عن الازدواج الخطي، وبيان المؤشرات الاقتصادية بالنموذج الخطي المتعدد ومعالجته دون حذف أي من المتغيرات التفسيرية، وإيجاد مقدرات إحصائية ذات كفاءة عالية واختبارها إحصائياً بحيث لا تعاني من مشكلة الازدواج الخطي، والاستفادة من الأساليب الإحصائية المطبقة لمعالجة آثار هذه المشكلة ومن ثم استخدام تلك المقدرات كمؤشرات لتحديد شكل العلاقة بين المتغيرات التفسيرية والمتغير التابع.

الطريقة البحثية ومصادر البيانات:

تعتمد الدراسة بصفة أساسية علي أساليب التحليل الكمي مثل مصفوفة الارتباط وأسلوب تحليل الانحدار الخطى المتعدد وطريقة انحدار ريدج لمعالجة مشكلة الازدواج الخطي، واستعانت الدراسة بالبيانات الرسمية المنشورة خلال الفترة (١٩٩٥– ٢٠١٥) التي تصدر عن الجهاز المركزي للتعبئة العامة والإحصاء، وزارة الزراعة واستصلاح الأراضي، الإدارة المركزية للاقتصاد الزراعي إلي جانب بعض الدراسات والبحوث العلمية وثيقة الصلة بالموضوع.

مشكلة الازدواج الخطي (*Multicollinearity*):

تتمثل مشكلة الازدواج الخطى في عدم استقلالية المتغيرات المستقلة وهو ما يعني وجود ارتباط بين المتغيرات المستقلة في نموذج الانحدار المراد تقديره، يعني وجود ارتباط خطي قوي بين بعض أو كل المتغيرات المستقلة في نموذج الانحدار، فإذا كان معامل الارتباط بينهما يساوي الصفر فلا يوجد

هناك أي مشاكل تتعلق بتقديرات معاملات الانحدار، ومن ثم يمكن القول أن مشكلة الازدواج الخطي تؤثر على دقة تقديرات معاملات الانحدار، وتكون العلاقة خطية بين متجهات المتغيرات المستقلة $X_1, X_2, X_3, \ldots\ldots, X_n$ إذا تحقق الشرط التالي:

$$C_1X_1 + C_2X_2 + C_3X_3 = 0$$

حيث أن الثوابت $C_1, C_2, C_3, \ldots\ldots C_n$ لا تساوي جميعها الصفر، وفي أغلب حالات الانحدار الخطي فإنه من غير المحتمل أن تكون العلاقة تامة، فإذا كان:

$$C_1X_1 + C_2X_2 + C_3X_3 \approx 0$$

أي مساوياً للصفر تقريباً، فإننا نقول أن متجهات المتغيرات $X_1, X_2, X_3 \ldots\ldots X_n$ مرتبة خطياً.

أسباب الازدواج الخطي:

تتمثل أهم أسباب الازدواج الخطي في اتجاه المتغيرات الاقتصادية معاً للتغير مع مرور الزمن، استخدام متغيرات مستقلة ذات فترات إبطاء في المعادلة المراد تقديرها، استخدام متغيرات مستقلة ذات علاقة صريحة في المعادلة المراد تقديرها، كما يؤدي صغر حجم العينة بحيث يصبح عدد المشاهدات قريباً من عدد المتغيرات التفسيرية إلي ظهور مشكلة الازدواج الخطي.

الكشف عن الازدواج الخطي:

هناك عدة اختبارات لاكتشاف الازدواج الخطي من أبرزها مصفوفة الارتباط البسيط بين المتغيرات موضع الدراسة ومعامل تضخم التباين والدليل الشرطي وحدود التسامح.

١- مصفوفة الارتباط البسيط (*Simple Correlation Matrix*):

إيجاد مصفوفة معامل الارتباط البسيط (*SCM*) بين المتغيرات المستقلة في المعادلة المراد تقديرها، حيث يشير معامل الارتباط القوي لوجود ارتباط خطي بين المتغيرين المستقلين المعنيين.

٢- معامل تضخم التباين (*Variance Inflation Factor*):

يكمن اكتشاف الازدواج الخطى من خلال حساب ما يعرف بقيمة معامل تضخم التباين (*VIF*) التي تكون أكبر من ١٠ كمؤشر لاكتشاف الازدواج الخطى حيث أن وجود الازدواج الخطي لا يعتبر مشكلة وإنما المشكلة تتمثل في درجة الازدواج الخطي، فإذا كانت قيمة (*VIF*) أقل من ١٠ يعني وجود ازدواج خطي ولكنه لا يمثل مشكلة وإذا زادت قيمته عن ١٠ يعني وجود ازدواج خطي وبالتالي يمثل هنا مشكلة أب أن مشكلة الازدواج الخطي تكمن في درجة هذا الازدواج.

٣- الدليل الشرطي (*Condition Index*):

عبارة عن الجذر التربيعي لحاصل قسمة أكبر جذر كامن (*Eigenvalue*) علي كل جذر كامن مقابل للأبعاد $D_J, J = 1,2,\ldots\ldots 8$ ويحسب هذا الدليل لأكبر جذر كامن وفق العلاقة:

$$K \equiv \frac{Maximum\, eigenvalue}{Minimum\, eigenvalue}$$

$$CI \equiv \sqrt{\frac{Maximum\, eigenvalue}{Minimum\, eigenvalue}} = \sqrt{K}$$

لإيجاد الدليل الشرطي (*CI*) لابد أولاً من احتساب الجذور الكامنة (*Eigenvalues*) حيث توضح كمية الاختلافات الكلية بين المتغيرات، فعندما تكون الجذور المميزة مساوية للصفر فإنه يدل على الازدواج الخطي التام، أما إذا كانت قريبة من الصفر فهذا مؤشر على وجود ازدواج خطي عالي، أما إذا كانت مساوية إلى الواحد فتعتبر الحالة الأمثل في عدم وجود مشكلة الازدواج الخطي ومن ثم يمكن إيجاد العدد الشرطي (K).

يؤخذ سوياً كل من الدليل الشرطي (*Condition Index*) ونسب التباين (*Variance Proportion*) حيث يستفاد منهما في بيان درجة التعدد الخطي والمتغيرات المرتبطة مع بعضها البعض، فإذا كانت قيمة الدليل الشرطي في حدود القيمة (٥- ١٠) وأن اثنين أو أكثر من نسب التباين أقل من ٠.٥ فهذا يدل علي أن الارتباط ضعيف، أما إذا كانت القيمة تقع بين (١٠- ٣٠) $10 \le CI \le 30$ فهذا يعني أن هناك ازدواج خطي من المعتدل إلي العالي، أما إذا تجاوزت قيمة *CI* (٣٠) فإن هذا مؤشر علي وجود ازدواج خطي قوي وعلي حجم المشكلة وخطورتها.

٤- حدود التسامح (*Tolerance*):

يعتبر هذا المقياس أحد المقاييس الدالة علي وجود أو عدم وجود مشكلة الازدواج الخطي، فإذا كانت قيمة *Tolerance* أكبر من ٠.١ دل ذلك علي عدم وجود مشكلة الازدواج الخطي، أما إذا كانت قيمة *Tolerance* أقل من ٠.١ دل ذلك علي وجود مشكلة الازدواج الخطي.

معالجة الازدواج الخطي:

اهتم الباحثون بإيجاد الحلول المناسبة لمعالجة مشكلة الازدواج الخطي سواء كان بزيادة حجم العينة، أو بإحلال متغيرات ذات فترات إبطاء محل المتغيرات المستقلة الأخرى في نماذج فترات الإبطاء الموزعة، أو حذف متغير أو أكثر من المتغيرات المستقلة الأكثر ارتباطاً مع بعضها البعض، أو استخدام طرق تقدير أخري ولكنها متحيزة بخلاف طريقة المربعات الصغرى العادية غير المتحيزة والتي يمكن أن تكون أكثر دقة ومن أهمها طريقة انحدار ريدج.

الأساس النظري لطريقة انحدار ريدج:

اقترحها لأول مرة *Horal* عام ١٩٦٢ بعد ذلك طورت بواسطة *Kennard & Horal* في عام ١٩٧٠ والتي تتضمن إضافة الثابت K إلي عناصر المصفوفة $(X'X)$ قبل أخذ معكوسها.

أولاً: تقدير معلمات انحدار ريدج:

يمكن تقدير معلمات انحدار ريدج من شكلين اثنين كما يلي:

١- تقدير معلمات انحدار ريدج في الشكل المعياري:

يتم تقدير معلمات الانحدار المعيارية وفقا لأسلوب انحدار ريدج كالأتي:

$$\hat{\beta}_{(K)} = (X'X + KI)^{-1} X'Y_i \quad 0 \le K \le 1 \ldots\ldots\ldots(1)$$

$\hat{\beta}_{(K)}$ = شعاع معاملات انحدار ريدج المعيارية المقدرة (لا يتضمن المقدار الثابت أو مقدار المقطع).

K = كمية ثابتة تمثل مقدار التحيز وقيمتها محصورة بين الصفر والواحد الصحيح.

$X'X$ = مصفوفة معاملات الارتباط البسيط ما بين المتغيرات المستقلة وبعضها البعض.

I = مصفوفة الوحدة من الرتبة $P \times P$

$X'Y$ = مصفوفة معاملات البسيط ما بين المتغيرات المستقلة والمتغير التابع.

٢- تقدير معلمات انحدار ريدج في الشكل الأصلي:

كما هو الحال بالنسبة لطريقة المربعات الصغرى عند تحويل المقدرات من شكل المصفوفة إلي الشكل الأصلي سيتم التعامل مع طريقة انحدار ريدج بنفس الشكل مع استثناء مقدار ثابت المعادلة $\hat{\beta}_i(K)$ والذي يتم إيجاده بطريقة مستقلة من خلال الشكل الأصلي باستخدام العلاقة التالية:

$$\hat{\beta}_i(K) = \frac{S_Y}{S_i} \hat{\beta}^*_{i,(k)} \ldots\ldots\ldots(2)$$

$$\hat{\beta}_0(K) = \overline{Y} - \hat{\beta}^*_{i,(k)} \overline{X}_1 - \hat{\beta}^*_{2,(K)} \overline{X}_2 \ldots\ldots\ldots \hat{\beta}^*_{P,(K)} \overline{X}_P \ldots\ldots\ldots(3)$$

كما يمكن تقدير الثابت من المعادلة التالية:

$$\hat{\beta}_0(K) = \hat{\beta}^{*r}_{I,(k)} Z_i \qquad Z_i = \frac{\overline{X}^r_i}{\sum \left(X^r_{ij} - \overline{X}^r_i\right)^2} \text{.....(4)}$$

ثانياً: تحديد القيمة المناسبة لمعلمة انحدار ريدج في ظل مشكلة الازدواج الخطي:

يترتب غالباً في ظل مشكلة الازدواج الخطي كبر حجم الأخطاء المعيارية للمعلمات المقدرة نتيجة كبر قيم العناصر القطرية للمصفوفة $(X'X)^{-1}$ ذلك لأن:

$$VAR(b) = S^2 (X'X)^{-1} \text{.....(5)}$$

يتسبب ذلك في عدم معنوية التقديرات إضافة إلي عدم استقرار المعلمات المقدرة واختلاف إشارتها عند حدوث أي تغيير في حجم العينة ومن الطرق الأساسية الواسعة الاستخدام في الكشف عن وجود مشكلة الازدواج الخطي حساب عوامل تضخم التباين $(VIF's)$ *Variance Inflation Factors* حيث تقيس عوامل تضخم التباين مدي تضخم تباينات معاملات الانحدار المقدرة في ظل وجود الازدواج الخطي ويتم حساب عامل تضخم التباين $VIS's$ لكل متغير مستقل علي حدة كالأتي:

$$VIF = \frac{1}{1 - R_i^2} \qquad i = 1,2,3 \text{.....} n \text{.....(6)}$$

R_i^2 = معامل التحديد لنموذج انحدار المتغير المستقل i علي باقي المتغيرات المستقلة حيث نوجد انحدار كل من (X_i) علي بقية المتغيرات الأخرى وذلك علي النحو التالي:

$$X_i = f(X_2, X_3 \text{.....} X_n) \text{.....(7)}$$

هكذا في كل مرة يتم حساب قيمة معامل التحديد R_i^2 في إيجاد قيمة VIF فإذا كانت القيمة أكبر من ١٠ دل ذلك علي وجود ازدواج خطي، ويقابل هذا الحد معامل تحديد قدره ٠.٩٠ لنموذج انحدار المتغير المستقل i علي بقية المتغيرات المستقلة وتعتبر قيمة VIF أكبر من الواحد الصحيح، وتستخدم قيم عوامل تضخم التباين $(VIF's)$ لقياس مدي بعد مقدرات المربعات الصغرى عن انحرافات معلمات الانحدار المقدرة الحقيقة الصيغة التالية:

$$E\left[\sum_{i=1}^{p} (\hat{\beta}_i - \beta_i)^2\right] = \sigma^2 \sum_{i=1}^{p} VIF_i = \sigma^2 P \text{.....(8)}$$

من ثم يمكن حساب النسبة التالية:

$$\frac{\sigma^2 \sum_{i=1}^{P} VIF_i}{\sigma^2 P} = \frac{\sum_{i=1}^{P} VIF_i}{P} \text{.....(9)}$$

تشير هذه النسبة إلي متوسط قيم $VIF's$ لمعاملات الانحدار المقدرة وإذا كانت المتغيرات المستقلة متعامدة أي لا يوجد بينهما ارتباط فإن هذه النسبة تساوي الواحد الصحيح ولذلك نجد أنه كلما زادت قيمة $VIF's$ عن الواحد الصحيح دل ذلك علي وجود ارتباط خطي بين المتغيرات المستقلة، وتستخدم بعض البرامج الجاهزة معكوس VIF للكشف عن وجود الازدواج الخطي وتحديد دخول أي متغير للنموذج من عدمه ويعرف هذا المقياس ب *Tolerance*.

$$Tolerance = \frac{1}{VIF_i} = 1 - R^2 \text{.....(10)}$$

تعكس قيمة الثابت K مقدار التحيز في المقدرات ويلاحظ انه عندما تكون قيمة الثابت K مساوية للصفر فانه يمكن الحصول علي مقدرات المربعات الصغرى:

$$\beta_i = (X'X)^{-1} X'Y \ldots\ldots\ldots\ldots (11)$$

عندما تكون قيمة الثابت K أكبر من الصفر نحصل علي مقدرات متحيزة إلا أنها أكثر استقراراً من المربعات الصغرى ويعاب علي طريقة انحدار ريدج صعوبة تحديد K التي تعطي أفضل نموذج.

إنتاج اللحوم الحمراء والعوامل المؤثرة عليها:

تعتبر اللحوم الحمراء من أهم مصادر البروتينات الحيوانية للإنسان والتي تمثل مشكلة حقيقية في مصر لانخفاض الكمية المنتجة عن الكمية المستهلكة، وما قد يترتب علي ذلك من انخفاض نسبة الاكتفاء الذاتي إلي نحو ٧٨% وتراجع نصيب الفرد منها فضلاً عن استيراد كميات كبيرة منها بما يؤدي إلي استنزاف النقد الأجنبي، الأمر الذي يتطلب دراسة أهم العوامل المؤثرة عليها وتقدير الدالة بين هذه العوامل والكمية المنتجة من اللحوم الحمراء بشكل دقيق حتى يمكن التنبؤ بالكمية المنتجة بدرجة ثقة عالية، ويوضح جدول (١) إجمالي إنتاج اللحوم الحمراء في مصر وأهم العوامل المنطقية المؤثرة عليها خلال الفترة (١٩٩٥– ٢٠١٥) والتي تشمل عدد الوحدات الحيوانية وكمية الأعلاف الخضراء وكمية الأعلاف الجافة وكمية الأعلاف المركزة والكمية المستهلكة من اللحوم الحمراء وسعر اللحوم الحمراء وسعر الدواجن بالجنيه/ كجم.

جدول (١): إجمالي إنتاج اللحوم الحمراء في مصر وأهم العوامل المؤثرة عليها خلال الفترة (١٩٩٥– ٢٠١٥).

السنوات	إجمالي الكمية المنتجة من اللحوم بالألف طن	عدد الوحدات الحيوانية بالألف رأس	كمية الأعلاف الخضراء بالألف طن	كمية الأعلاف الجافة بالألف طن	كمية الأعلاف المركزة بالألف طن	الكمية المستهلكة من اللحوم حمراء بالألف طن	سعر اللحوم الحمراء بالجنيه/ كجم	سعر الدواجن بالجنيه/ كجم
١٩٩٥	٦٠٣	٧٥٠٨	٥٤٠٦٨	١٢٧٧٧	٣٥١٠	٧٧٣	٣٤.٤٠	١٧.٦٣
١٩٩٦	٦٣٧	٧٦٠٨	٥٥٦٦٨	١٣٧٢١	٣٨٤٠	٩٥٧	٣٣.٣٣	١٦.٢٨
١٩٩٧	٦٤٥	٧٧٣٨	٥٣٨٤٤	١٣٩٧٢	٣٩٤٠	٧٥٠	٣٦.١٤	١٥.٨٤
١٩٩٨	٦٧٢	٧٩٢٣	٥٦٢٩٣	١٣٩١٥	٣٦٧٨	٨٠٤	٣٥.٤١	١٥.٤٣
١٩٩٩	٦٨٩	٨٣٥٢	٥٦٣٨٩	١٤٨٣٧	٣٩٨٧	٨٧٢	٣٥.١٠	١٥.٨٢
٢٠٠٠	٧٠٢	٨٥٤٦	٥٦٧٢٢	١٥٤١٣	٤٠٩٨	٩٣٤	٣٤.٥٣	١٥.٩٢
٢٠٠١	٦٩٦	٩٠٣٠	٥٤٠٨٩	١٥٩٨٨	٤٢٨٥	٧٩٣	٣٤.٣٦	١٥.٨٤
٢٠٠٢	٨١٩	٩٥٨٥	٥٦٩٩٣	١٦٤٩٣	٤٥٥٨	٩٦٠	٣٤.٣١	١٥.٦٣
٢٠٠٣	٨٣٧	٩٨١١	٥٨١٩٩	١٦٠١٢	٤٦٩٣	١٠١٩	٣٤.٨٤	١٥.٠٠
٢٠٠٤	٨١٥	١٠٠٤٨	٦١٢٨٤	١٥٥٨٢	٤٧٨٥	٩٦٠	٢٨.٣٠	١٣.٢٢
٢٠٠٥	٨٥٣	١٠٢٣٧	٦١٦٠٨	١٦٠٨٧	٤٨٠٥	١١٣٣	٣٨.٠٢	١٣.٩٨
٢٠٠٦	٨٧٧	١٠٤٥٢	٦٢٢٣٩	١٦٠٩٩	٤٩١٠	١٣١٢	٣٨.٦١	١٣.١٧
٢٠٠٧	٩١٥	١٠٩٣٠	٦٨١٤٢	١٦٥٤٢	٥٠١٦	١٣٨٢	٣٩.٦٢	١٢.١٤
٢٠٠٨	٩٣٢	١١٠٣٧	٦٧٩٩٧	١٦٥١٣	٥٣٩١	١١٧٦	٤١.٠٣	١٣.٩٦
٢٠٠٩	٩٨٠	١٠٢٧٥	٦٨٢٤٣	١٦٩٦٠	٤٨١٨	١١٩٦	٤٢.٥٧	١٢.٦٣
٢٠١٠	٧٩١	١٠٤٣٠	٥٩١٨٧	١٨١٣٩	٣٨٤٥	١٠٥٢	٤٥.٢٥	١٢.١٥
٢٠١١	٧٨٧	١٠١١٩	٦٠٧٨٧	١٨٣٨٧	٣٩٥٤	١٠٣٣	٥٠.٠٠	١٤.١٩
٢٠١٢	٧٨٨	١١٣٠٦	٥١٠٠٥	١٨٥٦٦	٤١٠٢	١٠٥٢	٥٠.٦٤	١٦.٣٩
٢٠١٣	٧٨٠	١٠٨٠٧	٤٧٥٨٢	١٧٢٨٩	٤٤٤٧	١٣٩٧	٦٤.٤٥	٢٣.٥٧
٢٠١٤	٧٤٢	١٠٨٦٣	٤٤٩٥٥	١٥٣٣٣	٤٦٠٦	٨٧٨	٦٢.٠٨	٢٤.٨٨
٢٠١٥	٧٩٣	١٠٤٥٤	٤٠٠٧١	٢٠٣٦٠	٤٩٩٠	١٠٠٨	٦٥.١٧	٢٦.٦٠

المصدر: وزارة الزراعة واستصلاح الأراضي، قطاع الشئون الاقتصادية، نشرة الميزان الغذائي، أعداد مختلفة.

Y^*= إنتاج اللحوم الحمراء من مذبوحات كل من الأبقار والجاموس والأغنام والماعز والجمال وقدرت علي أساس الأبقار =١، الجاموس = ١.٢٥، الأغنام = ٠.١، الماعز = ٠.٧، الإبل = ٠.٧٥ وحدة حيوانية.

تحليل الانحدار المتعدد استخدام طريقة (*OLS*):

بإجراء تحليل الانحدار الخطي المتعدد بطريقة المربعات الصغرى العادية (*OLS*) لانحدار المتغيرات التفسيرية موضع الدراسة علي كمية إنتاج اللحوم في مصر خلال الفترة (١٩٩٥- ٢٠١٥) تبين أن المعادلة كانت علي النحو التالي:

$$Y_i = \underset{(-1.347^{-})}{-424.76} + \underset{(0.384)^{-}}{0.007X_1} + \underset{(1.61)^{-}}{0.007X_2} + \underset{(2.47)^{-}}{0.017X_3} + \underset{(2.23)^{*}}{0.094X_4} + \underset{(0.37)^{-}}{0.022X_5} + \underset{(0.47)^{-}}{1.250X_6} - \underset{(-0.006)^{-}}{0.060X_7}$$

$$R = 0.97 \qquad R^2 = 0.948 \qquad F = (33.621)^{**} \qquad Sigma = 28.868$$

$$VIF(X_1) = 13.947 \quad VIF(X_2) = 21.011 \quad VIF(X_3) = 3.701 \quad VIF(X_4) = 10.535$$

$$VIF(X_5) = 10.170 \quad VIF(X_6) = 19.861 \quad VIF(X_7) = 36.843$$

حيث:

$\hat{Y}_i$ = القيمة التقديرية لإنتاج اللحوم الحمراء بالألف طن.

X_1 = عدد الوحدات الحيوانية بالألف رأس. X_2 = كمية الأعلاف الخضراء بالألف طن.

X_3 = كمية الأعلاف الجافة بالألف طن. X_4 = كمية الأعلاف المركزة بالألف طن.

X_5 = الكمية المستهلكة من اللحوم الحمراء بالألف طن.

X_6 = سعر اللحوم الحمراء بالجنيه/ كجم. X_7 = سعر الدواجن بالجنيه/ كجم

تشير الدالة السابقة إلي وجود ارتباط قوي موجب بين المتغيرات المستقلة موضع الدراسة ومتوسط إنتاج اللحوم الحمراء حيث بلغت قيمة معامل الارتباط نحو ٠.٩٧ تقريباً كما تبين أن المتغيرات المستقلة موضع الدراسة تساهم بنحو ٩٥% تقريبا من إجمالي العوامل الكمية المؤثرة علي إنتاج اللحوم الحمراء.

يتضح أيضاً معنوية الدالة حيث بلغت قيمة *F* المقدرة نحو ٣٣.٦٢١ كما ثبتت المعنوية الإحصائية لكمية الأعلاف المركزة بالألف طن حيث تبين وجود علاقة طردية بينها وبين الكمية المنتجة من اللحوم الحمراء وأنه بزيادة كمية الأعلاف المركزة بنحو ألف طن يؤدي إلي زيادة كمية إنتاج اللحوم بنحو ٠.٠٩٤ ألف طن في حين لم تثبت المعنوية الإحصائية لباقي المتغيرات المستقلة، كما بلغت قيمة *Sigma* نحو ٢٨.٨٦ وهي تعبر عن الجذر التربيعي لمتوسط مربعات الخطأ والمقدر بنحو (٨٣٣.٣٦٧) في جدول تحليل التباين.

تشير نتائج تحليل التباين والموضحة بالجدول (٢) إلي وجود فروق معنوية في إجمالي الكمية المنتجة من اللحوم الحمراء والمتغيرات المستقلة المؤثرة عليها موضع الدراسة حيث بلغت قيمة *F* المحسوبة نحو ٣٣.٦٢١ وهي معنوية عند مستوي معنوية ٠.٠٠١.

جدول (٢): تحليل التباين لطريقة (*OLS*).

المصدر	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	المعنوية
الانحدار	٧	١٩٦١٢٨.٥١١	٢٨٠١٨.٣٥٩	٣٣.٦٢١	**
الخطأ	١٣	١٠٨٣٣.٧٧٤	٨٣٣.٣٦٧		
المجموع الكلي	٢٠	٢٠٦٩٦٢.٢٨٦			

المصدر: حسبت من بيانات جدول (١).

** معنوية عند مستوي ١%

مؤشرات الكشف عن الازدواج الخطي:

١- مصفوفة الارتباط البسيط:

يوضح جدول (٣) مصفوفة الارتباط البسيط بين مختلف متغيرات الدراسة خلال الفترة (١٩٩٥-٢٠١٥) ومنها يتبين وجود ارتباط معنوي موجب بين Y_i وكل المتغيرات موضع الدراسة باستثناء المتغيرين X_6 و X_7 وأن الارتباط بين Y_i والمتغير X_7 كان سالباً كما يتبين منطقية العلاقات الاقتصادية بين العامل التابع Y_i والمتغيرات المستقلة موضع الدراسة X's، ويتبين وجود ارتباط معنوي موجب بين X_1 وكل المتغيرات المستقلة باستثناء المتغيرين X_2 و X_7 ، وبالنسبة للمتغير X_2 يتبين وجود ارتباط معنوي سالب مع المتغيرين X_6 و X_7 وارتباط غير معنوي سالب مع X_3 وموجب مع كل من X_4 و X_5 ، أما المتغير X_3 فيلاحظ وجود ارتباط معنوي موجب مع كل من X_5 و X_6 وغير معنوي مع X_4 و X_7 ، كما يوجد ارتباط معنوي موجب مع X_5 وغير معنوي مع X_6 وارتباط غير معنوي سالب مع X_7 ، أما X_5 فيتضح وجود ارتباط غير معنوي موجب مع X_6 وغير معنوي سالب مع X_7 وأخيراً فإنه يوجد ارتباط معنوي موجب بين X_7 X_6 .

مما سبق يلاحظ وجود ارتباط قوي بين كل من X_1 وكل من (X_3 و X_5) يقدر بنحو ٠.٧٧ ، ٠.٧٢ علي الترتيب، ووجود ارتباط قوي بين X_2 و X_7 يقدر بنحو -٠.٩٠ تقريباً وارتباط قوي بين X_6 و X_7 يقدر بنحو ٠.٧٤ أي أنه يتوقع وجود مشكلة الازدواج الخطي بين المتغيرات X_1 و X_2 و X_6 و X_7.

يلاحظ أن معامل الارتباط بين (X_1 ، X_3) والذي يقدر بنحو ٠.٧٧ أعلي من معامل الارتباط بين (X_3 ، Y) والذي يقدر بنحو ٠.٥٦، وأن معامل الارتباط بين (X_1 ، X_6) يقدر بنحو ٠.٥٧ وهو أعلي من معامل الارتباط بين (X_6 ، Y) والذي يقدر بنحو ٠.١٦

يلاحظ أيضاً أن معامل الارتباط بين (X_2 ، X_6) كان سالباً ويقدر بنحو -٠٥٩ وهو أعلي من معامل الارتباط بين كل من (X_2 ، Y) و (X_6 ، Y) والذي يقدر بنحو ٠.٥٦ ، ٠.١٦ علي الترتيب، كما كان معامل الارتباط بين (X_7 ، X_2) سالباً ويقدر بنحو -٠٨٩ وهو يزيد عن معامل الارتباط بين كل من (X_2 ، Y) و (X_7 ، Y) والذي يقدر بنحو ٠.٥٦ ، -٠.٣١ علي الترتيب.

يقدر معامل الارتباط بين (X_3 ، X_6) بنحو ٠.٦٤ وهو يزيد عن معامل الارتباط بين كل من (X_3 ، Y) و (X_6 ، Y) والمقدر بنحو ٠.٥٦ ، ٠.١٦ علي الترتيب، كما يقدر معامل الارتباط بين (X_5 ، X_6) بنحو ٠.٣٤ وهو يزيد عن معامل الارتباط بين (X_6 ، Y) والمقدر بنحو ٠.١٦ في حين أن معامل الارتباط بين (X_6 ، X_7) يقدر بنحو ٠.٧٤ وهو أعلي أيضاً من معامل الارتباط بين كل من (X_6 ، Y) و (X_7 ، Y) والمقدر بنحو ٠.١٦ ، -٠.٣١ علي الترتيب.

جدول (٣) مصفوفة الارتباط البسيط

المتغيرات	Y_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Y_i	١							
X_1	**٠.٧٩٦٥	١						
X_2	**٠.٥٦١٦	٠.١٠٥٧	١					
X_3	**٠.٥٦٠٥	**٠.٧٦٥٢	٠.١٣٩٧-	١				
X_4	**٠.٨٥٠٧	**٠.٦٩٨٣	٠.٤٠٥٧	٠.٣٢٩١	١			
X_5	**٠.٧٧٠٠	**٠.٧١٧٢	٠.٣٧٨٧	**٠.٤٨٨٣	**٠.٦٤٢٦	١		
X_6	٠.١٥٩٣	**٠.٥٧٤٧	**٠.٥٨٦٥-	**٠.٦٣٩٨	٠.٠٩٧٠	٠.٣٣٣٧	١	
X_7	٠.٣١٣٣-	٠.٠٧٨٨	**٠.٨٩٧٢-	٠.١٩٣١	٠.١٠٨٠-	٠.١٣٣٥-	**٠.٧٤٤٧	١

المصدر: حسبت من بيانات جدول (١).

٢- معامل تضخم التباين (*VIF*):

تبين قيم معامل تضخم التباين (*VIF*) الموضحة أسفل معادلة الانحدار الخطي المتعدد السابقة أن قيم (*VIF*) لكل المتغيرات التفسيرية موضع الدراسة كانت أكبر من ١٠ باستثناء المتغير X_3 والتي تقدر قيمة (*VIF*) له بنحو ٣.٧٠١ في حين يلاحظ ارتفاع قيمة (*VIF*) للمتغيرات X_2 ، X_6 ، X_7 بدرجة ملحوظة مما يدل علي وجود ازدواج خطي عالي بين هذه المتغيرات، بينما يوجد ازدواج خطي ضعيف للمتغير X_1 وكذلك وجود ازدواج خطي معتدل لبقية المتغيرات التفسيرية الأخرى وهي X_4 ، X_5.

٣- الدليل الشرطي (*CI*):

يوضح جدول (٤) تقدير قيم الدليل الشرطي(*CI*) ومنه يتبين أن قيمة(*CI*) للمتغير $X_1 = 9.05$ وهي بذلك محصورة بين (٥- ١٠) مما يدل علي وجود ازدواج خطي ضعيف، وبالنسبة للمتغيرات X_2 ، X_3 فقد تراوحت قيمة (*CI*) بين (١٠- ٣٠) حيث تقدر بنحو ١٦.٨٢، ٢٧.٩٢ علي الترتيب وهو ما يشير إلي وجود ازدواج خطي معتدل إلي عالي، في حين أن قيمة (*CI*) للمتغيرات X_4 ، X_5 ، X_6 ، X_7 تشير إلي وجود ازدواج خطي بدرجة كبيرة لهذه المتغيرات حيث تقدر قيمة (*CI*) لهذه المتغيرات بنحو ٣٥.٣٢، ٤٨.٧٩، ٨٥.٠٥، ٢٠٨.٨١ علي الترتيب.

تشير قيم الجذر الكامن (*Eigenvalues*) الموضحة بجدول (٤) إلي أن قيمة هذه الجذور اقتربت من الصفر بدرجة كبيرة لكل من المتغيرات التفسيرية X_4 ، X_5 ، X_6 ، X_7 مما يشير إلي وجود ازدواج خطي معتدل إلي عالي لهذه المتغيرات دون غيرها من المتغيرات موضع الدراسة، في حين أن بقية المتغيرات التفسيرية كانت قيم الجذر الكامن المقدرة لها بعيدة عن الصفر خاصة المتغير X_1 مما يشير إلي أن هذه المتغيرات لا تتضمن مشكلة الازدواج الخطي.

جدول (٤) مؤشرات الكشف عن الازدواج الخطي

المتغيرات	*CI*	*VIF*	*Tolerance*	*Eigenvalue*
X_1				
X_2	٩.٠٥	١٣.٩٥	٠.٠٧٢	٧.٨٥٥
X_3	١٦.٨٢	٢١.٠١	٠.٠٤٨	٠.٠٩٦
X_4	٢٧.٩٢	٣.٧٠	٠.٢٧٠	٠.٠٢٨
X_5	٣٥.٣٢	١٠.٥٤	٠.٠٩٥	٠.٠١٠
X_6	٤٨.٧٩	٣.١٧	٠.٣١٥	٠.٠٠٦
X_7	٨٥.٠٥	١٩.٨٦	٠.٠٥٠	٠.٠٠٣
	٢٠٨.٨١	٣٦.٨٤	٠.٢٧	٠.٠٠١

المصدر: حسبت من بيانات جدول (١).

٤- حدود التسامح (*Tolerance*):

يشير جدول (٤) إلي تقدير حدود التسامح (*Tolerance*) في قيم المتغيرات التي تؤثر علي إنتاج اللحوم الحمراء تبين أن هناك أكثر من متغير تظهر فيها مشكلة الازدواج الخطي وهي كل X_2 ، X_3 ، X_4 ، X_5 ، X_6 ، X_7 حيث يقدر قيمته بنحو ٠.٠٩٦، ٠.٠٢٨، ٠.٠١٠، ٠.٠٠٦، ٠.٠٠٣، ٠.٠٠١ علي الترتيب وكل هذه القيم اقل من ٠.١ وهو دليل علي وجود مشكلة الازدواج الخطي.

تطبيق أسلوب *Ridge Regression* لمعالجة مشكلة الازدواج الخطي:

بتطبيق أسلوب *Ridge Regression* عند تحليل البيانات ومعالجة مشكلة الازدواج الخطي كانت المعادلة علي النحو التالي:

$$Y_i = -354.15 + \underset{(0.77)^-}{0.0098X_1} + \underset{(2.48)^{**}}{0.0057X_2} + \underset{(2.44)^{**}}{0.0156X_3} + \underset{(2.83)^{**}}{0.0911X_4} + \underset{(0.58)^-}{0.0346X_5} + \underset{(0.72)^-}{1.0749X_6} - \underset{(-0.24)^-}{1.0070X_7}$$

$$K = 0.02 \qquad R = 0.968 \qquad R^2 = 0.938 \qquad F = (27.88)^{**} \qquad Sigma = 31.53$$

$$VIF(X_1) = 5.060 \qquad VIF(X_2) = 5.486 \qquad VIF(X_3) = 2.663$$

$$VIF(X_4) = 2.819 \quad VIF(X_5) = 2.592 \quad VIF(X_6) = 5.161 \qquad VIF(X_7) = 5.785$$

تشير المعادلة إلي أن زيادة كمية الأعلاف الخضراء بنحو ألف طن سوف يؤدي إلي زيادة الكمية المنتجة من اللحوم الحمراء سنوياً بنحو ٠.٠٠٦ ألف طن وتعتبر هذه الزيادة معنوية إحصائياً، في حين تبين أن زيادة كمية الأعلاف الجافة بنحو ألف طن يؤدي إلي زيادة الكمية المنتجة من اللحوم الحمراء سنوياً بنحو ٠.٠١٦ ألف طن وتعتبر هذه الزيادة معنوية إحصائياً، في حين أن زيادة كمية الأعلاف المركزة بنحو ألف طن يؤدي إلي زيادة إنتاج اللحوم الحمراء سنوياً بنحو ٠.٠٩١ ألف طن وتعتبر هذه الزيادة معنوية إحصائياً، ويتبين عدم معنوية بقية معاملات الانحدار للمتغيرات التفسيرية الأخرى وهي X_1، X_5، X_6، X_7.

تبين أيضاً معنوية المعادلة عند مستوى معنوية ١% وتشير قيمة معامل التحديد R^2 إلي أن هذه العوامل تفسر نحو ٩٤% من إجمالي التغيرات في الكمية المنتجة من اللحوم الحمراء خلال فترة الدراسة، وأن القيمة المثلي التي عندها يتم إزالة الازدواج الخطي هي عند $K = 0.02$ ، كما بلغت قيمة *Sigma* نحو ٣١.٥٣ وهي تعبر عن الجذر التربيعي لمتوسط مربعات الخطأ والمقدر بنحو (٩٩٤.٣٧) في جدول تحليل التباين

تشير نتائج تحليل التباين لطريقة انحدار ريدج والموضحة بالجدول (٥) إلي وجود فروق معنوية في إجمالي الكمية المنتجة من اللحوم الحمراء والمتغيرات المستقلة المؤثرة عليها موضع الدراسة حيث بلغت قيمة F المحسوبة نحو ٢٧.٨٧٦ وهي معنوية عند مستوي معنوية ٠.٠٠١.

جدول (٥) تحليل التباين لطريقة (*RR*)

المصدر	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	المعنوية
الثابت	١٢٧٣٤٣٢٠	١	١٢٧٣٤٣٢٠	٢٧.٨٧٦	**
النموذج	١٩٤٠٣٥.٥	٧	٢٧٧١٩.٣٥		
الخطأ	١٢٩٢٦.٨٣	١٣	٩٩٤.٣٧		
المجموع الكلي	٢٠٦٩٦٢.٣	٢٠	١٠٣٤٨.١١		

المصدر: حسبت من بيانات جدول (١).

** معنوية عند مستوي ١%

معاملات انحدار *Ridge* القياسية:

تشير معاملات الانحدار القياسية إلي أن زيادة الانحراف القياسي لكل قيمة من قيم المتغيرات التفسيرية بوحدة واحدة سوف تؤدي إلي زيادة أو انخفاض الانحراف القياسي للمتغير التابع بما يعادل قيمة معامل الانحدار القياسي لكل متغير مستقل، ويوضح جدول (٦) معاملات انحدار *Ridge* القياسية للمتغيرات المستقلة موضع الدراسة ويلاحظ أنه بزيادة قيمة K فإن قيم انحدار المتغيرات X_1 ، X_5 تتزايد تدريجياً في حين تتناقص قيم كل من المتغيرات X_2 ، X_3 ، X_4 ، X_6 ، X_7 تدريجياً، بمعني أن وجود الازدواج الخطي بين المتغيرات التفسيرية في المعادلة قد تسبب في انحراف قيم معاملات الانحدار لهذه المتغيرات عن قيمتها الحقيقية المفسرة للتغير في المتغير التابع وهو إجمالي الكمية المنتجة من اللحوم الحمراء.

جدول (٦) معاملات انحدار *Ridge* القياسية

K	X_1	X_2	X_3	X_4	X_5	X_6	X_7
٠.٠٠٠٠٠١	٠.٠٩١٠	٠.٤٦٧٢	٠.٣٠١٠	٠.٤٥٨٧	٠.٠٤١٨	٠.١٣١٧	-٠.٠٠٢٤
٠.٠٠٠٠١٠	٠.٠٩١٠	٠.٤٦٧١	٠.٣٠٠٩	٠.٤٥٨٧	٠.٠٤١٩	٠.١٣١٧	-٠.٠٠٢٤
٠.٠٠٠١٠	٠.٠٩١١	٠.٤٦٦٧	٠.٣٠٠٨	٠.٤٥٨٧	٠.٠٤٢٠	٠.١٣١٧	-٠.٠٠٢٨
٠.٠٠٠٤٠٠	٠.٠٩١٤	٠.٤٦٥٣	٠.٣٠٠٣	٠.٤٥٨٧	٠.٠٤٢٤	٠.١٣١٦	-٠.٠٠٣٩
٠.٠٠٤٠٠٠	٠.٠٩٦٢	٠.٤٥٠٢	٠.٢٩٥٠	٠.٤٥٧٦	٠.٠٤٧٠	٠.١٢٩٦	-٠.٠١٤٦
٠.٠٠٧٠٠٠	٠.١٠٠٦	٠.٤٣٩٧	٠.٢٩١٠	٠.٤٥٥٩	٠.٠٥٠٧	٠.١٢٦٩	-٠.٠٢١٣
٠.٠٠٨٠٠٠	٠.١٠٢٢	٠.٤٣٦٥	٠.٢٨٩٨	٠.٤٥٥٢	٠.٠٥١٩	٠.١٢٥٩	-٠.٠٢٣٢
٠.٠٢٠٠٠٠	٠.١١٩٧	٠.٤٠٤٩	٠.٢٧٦٨	٠.٤٤٦١	٠.٠٦٤٧	٠.١١٣٢	-٠.٠٣٩٦

المصدر: حسبت من بيانات جدول (١).

معامل تضخم التباين (*VIF*):

يوضح جدول (٧) معامل تضخم التباين (*VIF*) للمتغيرات المستقلة موضع الدراسة المؤثرة علي إجمالي إنتاج اللحوم الحمراء في مصر خلال الفترة (١٩٩٥- ٢٠١٥) عند كل قيمة من قيم ثابت التحيز K حيث يلاحظ أنه بزيادة قيمة K لنحو ٠.٠٠٠٤ فإن قيمة (*VIF*) للمتغير X_1 انخفضت عن ١٠ بينما زيادة قيمة K لنحو ٠.٠٠٨ انخفضت قيمة (*VIF*) للمتغير X_2 لأقل من ١٠ وبالنسبة للمتغير X_3 كانت قيمة (*VIF*) له أقل من ١٠ لكل قيم K المحددة، وعندما بلغت قيمة K نحو ٠.٠٠٠٤ انخفضت قيمة (*VIF*) لأقل من ١٠ للمتغير X_4 وبالنسبة للمتغير X_5 كانت قيمة (*VIF*) له أقل من ١٠ لكل قيم K المحددة، أما المتغير X_6 فقد انخفضت قيمة (*VIF*) له لأقل من ١٠ عندما بلغت قيمة K نحو ٠.٠٠٠٧.

أخيراً فإن قيمة (*VIF*) انخفضت لأقل من ١٠ عندما بلغت قيمة K للمتغير X_7 نحو ٠.٠٢ أي أن الازدواج الخطي بين المتغيرات المستقلة موضع الدراسة تم إزالته عندما بلغت قيمة K نحو ٠.٠٢ حيث بلغت قيمة (*VIF*) لكل المتغيرات أقل من ١٠ وهو ما يعكس خلو نموذج الانحدار الخطي المتعدد من مشكلة الازدواج الخطي ومن ثم واقعية قيم معاملات الانحدار الخطي المقدرة وكذلك واقعية نسب مساهمة المتغيرات التفسيرية في تفسير قيمة التغير في المتغير التابع.

جدول (٧) معامل تضخم التباين (*VIF*)

K	X_1	X_2	X_3	X_4	X_5	X_6	X_7
٠.٠٠٠٠٠١	١٣.٩٤٦٤	٢١.٠١٠٠	٣.٧٠٠٦	١٠.٥٣٤١	٣.١٧٠٢	١٩.٨٥٩٨	٣٦.٨٣٩٥
٠.٠٠٠٠١٠	١٣.٩٣١٧	٢٠٩٨٢٨	٣.٦٩٩٥	١٠.٥٢٠٠	٣.١٦٩٨	١٩.٨٣٣٧	٣٦.٧٨١٧
٠.٠٠٠١٠٠	١٣.٧٩٤٧	٢٠.٧٢٨٢	٣.٦٨٩٢	١٠.٣٨٨٤	٣.١٦٦٦	١٩.٥٨٩٧	٣٦.٢٤٠٨
٠.٠٠٠٤٠٠	١٣.٣٥٨٥	١٩.٩١٨٨	٣.٦٥٦١	٩.٩٧٠٣	٣.١٥٥٩	١٨.٨١٤١	٣٤.٥٢٣٦
٠.٠٠٤٠٠٠	٩.٨٠٧٠	١٣.٤٢٦٥	٣.٣٥٠٥	٦.٦٥٠٣	٣.٠٣٣٤	١٢.٦١٣٧	٢٠.٩٤٨٢
٠.٠٠٧٠٠٠	٨.١٥٥٨	١٠.٥١٠٦	٣.١٧٠٦	٥.١٩٣٢	٢.٩٣٩١	٩.٨٤٩١	١٥.٠٥٧٧
٠.٠٠٨٠٠٠	٧.٧٤٤٣	٩.٨٠١٥	٣.١١٩٢	٤.٨٤٤٦	٢.٩٠٩٠	٩.١٨٠١	١٣.٦٦١١
٠.٠٢٠٠٠٠	٥.٠٦٠٤	٥.٤٨٥٦	٢.٦٦٣١	٢.٨١٩٠	٢.٥٩١٦	٥.١٦٠٥	٥.٧٨٥٠

المصدر: حسبت من بيانات جدول (١).

تحليل قيم *K*:

يوضح جدول (٨) نتائج تحليل نموذج الانحدار الخطي عند قيم مختلفة من ثابت التحيز K حيث يتبين أنه بزيادة قيمة K يلاحظ تراجع قيمة معامل التحديد أي أن العلاقة بينهما علاقة عكسية فعدما بلغت قيمة K نحو ٠.٠٢ أي عندما تم معالجة مشكلة الازدواج الخطي في نموذج الانحدار المتعدد للمتغيرات المستقلة المؤثرة علي إجمالي إنتاج اللحوم الحمراء خلال الفترة (١٩٩٥- ٢٠١٥) انخفضت قيمة R^2 لتأخذ قيمتها الحقيقية عند ٠.٠٩٣.

تعبر قيم *Sigma* عن الجذر التربيعي لمتوسط مربعات الخطأ ويبين جدول (٨) أن قيمتها تتزايد تدريجياً مع زيادة ثابت التحيز K فكلما زادت قيمة K تزايدت قيمة *Sigma* أي أن العلاقة بينهما علاقة طردية فعندما بلغت قيمة K نحو ٠.٠٠٠٠٠١ بلغت قيمة *Sigma* نحو ٢٨.٨٦٨٢ وعندما أخذت قيمة K في التزايد تدريجياً إلي أن بلغت نحو ٠.٠٢ تزايدت قيم *Sigma* تدريجياً أيضاً إلي أن بلغت نحو ٣١.٥٣٣٧.

أما قيم $B`B$ فهي تعبر عن مجموع مربعات معاملات الانحدار القياسية للمتغيرات المستقلة موضع الدراسة حيث تفترض طريقة انحدار ريدج أن هذه القيمة كبيرة وتحاول الحد منها إلي أن تثبت وعندها تتحدد القيمة المناسبة لثابت التحيز $K$ ومن جدول (٨) يلاحظ أن قيمة $B`B$ ثبتت عندما بلغت قيمة K نحو ٠.٠٢ فعندما بلغت قيمة K نحو ٠.٠٠٠٠٠١ بلغت قيمة $B`B$ نحو ٠.٥٤٦٦ وعند تزايد قيمة $K$ تدريجياً لتبلغ نحو ٠.٠٢ فإن قيمة $B`B$ ثبتت عند ٠.٤٧٢٤ وبصفة عامة فكلما زادت قيمة K كلما انخفضت قيمة $B`B$ إلي أن تثبت أي أن العلاقة بينهما علاقة عكسية.

أيضاً يوضح جدول (٨) قيم متوسط معامل تضخم التباين *AveVIF* حيث يلاحظ أنه عندما بلغت قيمة K نحو ٠.٠٠٠٠٠١ تحدد قيمة هذا المتوسط بنحو ١٥.٥٨٠١ وهذه القيمة أعلي من القيمة ١٠ وهي القيمة المحددة لاختبار وجود مشكلة الازدواج الخطي في البيانات موضع التحليل، وعندما أخذ ثابت التحيز K في الزيادة تدريجياً أخذت قيمة *AveVIF* في الانخفاض تدريجياً أيضاً إلي أن بلغت أقل من ١٠ عندما بلغت قيمة K نحو ٠.٠٠٤ ومن واقع بيانات الجدول (٧) يتبين أن قيمة *VIF* لكل من المتغيرات X_2 ، X_6 ، X_7 كانت أكبر من ١٠ إلا أن مشكلة الازدواج الخطي في البيانات انتهت تماماً عندما بلغت قيمة K نحو ٠.٠٢ ويلاحظ أنه كلما زادت قيم K كلما انخفضت قيم *AveVIF* أي أن العلاقة بينهما علاقة عكسية، ومما سبق يتبين أن هذا المؤشر وإن كان لا يعبر عن إزالة مشكلة ازدواج الخطي من البيانات تماماً إلا أنه يوضح تراجع حدة هذه المشكلة في البيانات التي يتم تحليلها.

جدول (٨) تحليل قيم K

K	R^2	*Sigma*	$B'B$	*AveVIF*	*MaxVIF*
٠.٠٠٠٠٠٠١	٠.٩٤٧٧	٢٨.٨٦٨٢	٠.٥٤٦٦	١٥.٥٨٠١	٣٦.٨٣٩٥
٠.٠٠٠٠١٠	٠.٩٤٧٦	٢٨.٨٦٩٦	٠.٥٤٦٥	١٥.٥٥٩٩	٣٦.٧٨١٧
٠.٠٠٠١٠٠	٠.٩٤٧٦	٢٨.٨٨٣٢	٠.٥٤٦١	١٥.٣٧١١	٣٦.٢٤٠٨
٠.٠٠٠٤٠٠	٠.٩٤٧٤	٢٨.٩٢٨٢	٠.٥٤٤٥	١٤.٧٧١٠	٣٤.٥٢٣٦
٠.٠٠٤٠٠٠	٠.٩٤٥٥	٢٩.٤٥٤٣	٠.٥٢٧٦	٩.٩٧٥٧	٢٠.٩٤٨٢
٠.٠٠٧٠٠٠	٠.٩٤٣٩	٢٩.٨٧٣٩	٠.٥١٥١	٧.٨٣٩٤	١٥.٠٥٧٧
٠.٠٠٨٠٠٠	٠.٩٤٣٤	٣٠.٠١٠٣	٠.٥١١٢	٧.٣٢٢٨	١٣.٦٦١١
٠.٠٢٠٠٠٠	٠.٩٣٧٥	٣١.٥٣٣٧	٠.٤٧٢٤	٤.٢٢٣٦	٥.٧٨٥٠

المصدر: حسبت من بيانات جدول (١).

أخيراً يعبر المؤشر *MaxVIF* في جدول (٨) عن أعلي قيمة لمعامل تضخم التباين للمتغيرات التفسيرية موضع الدراسة حيث يتبين أنه عندما بلغت قيمة K نحو ٠.٠٠٠٠٠٠١ بلغت قيمة *MaxVIF* نحو ٣٦.٨٣٩٥ وفي ظل تزايد قيمة ثابت التحيز K تدريجياً إلي أن بلغت نحو ٠.٠٢ أخذت قيم *MaxVIF* في الانخفاض تدريجياً إلي أن بلغت نحو ٥.٧٨٥٠ للمتغير X_7 أي أنه كلما زادت قيمة ثابت التحيز K كلما انخفضت قيمة *MaxVIF* أي أن العلاقة بينهما علاقة عكسية، ويوضح هذا المؤشر أعلي قيمة لمعامل تضخم التباين للمتغيرات التفسيرية موضع الدراسة ومن الجدول (٨) يتبين أنه عندما بلغت قيمة K نحو ٠.٠٠٨ بلغت قيمة *MaxVIF* نحو ١٣.٦٦١ وكانت للمتغير X_7 في نموذج التحليل وهي

أكبر من القيمة ١٠ مما يدل علي وجود مشكلة الازدواج الخطي في البيانات وقد تم معالجة هذه المشكلة تماماً عندما بلغت قيمة ثابت التحيز K نحو ٠.٠٢

آثار الازدواج الخطي في نموذج الانحدار المتعدد:

بمقارنة نتائج تحليل نموذج الانحدار الخطي المتعدد للمتغيرات المستقلة المؤثرة علي إجمالي إنتاج اللحوم في مصر خلال الفترة (١٩٩٥- ٢٠١٥) بطريقة (RR) بعد معالجة الازدواج الخطي بمثيلتها الذي تم تحليلها بطريقة (OLS) والتي تتضمن مشكلة الازدواج الخطي ويمكن إيجاز أهم آثار الازدواج الخطي التي يمكن ملاحظتها عند مقارنة الطريقتين فيما يلي:

١- ترتب علي إزالة مشكلة الازدواج الخطي بين المتغيرات المستقلة أن تحولت المعنويات الإحصائية لبعض المتغيرات المستقلة من غير معنوية إلي معنوية كما في المتغيرات X_2 ، X_3 أي أنها في الأصل كانت معنوية إلا أن الازدواج الخطي تسبب في تحويلها إلي غير معنوية.

٢- بصفة عامة يمكن أن تتحول معنوية بعض المتغيرات التفسيرية من خلال تقدير قيمة t في نموذج التحليل من معنوية إلي غير معنوية أو العكس.

٣- تغير قيم كل معاملات انحدار المتغيرات التفسيرية بعد معالجة مشكلة الازدواج الخطي عند استخدام طريقة (RR) بشكل ملحوظ بالمقارنة بمثيلتها قبل المعالجة عند استخدام طريقة (OLS).

٤- يمكن أن نستنتج أن مشكلة الازدواج الخطي قد تتسبب أيضاً في تغيير إشارات بعض معاملات الانحدار للمتغيرات التفسيرية في النموذج الإحصائي بما يخالف المنطق أو النظرية الاقتصادية أو طبيعة العلاقة بين المتغيرات التفسيرية والمتغير التابع.

٥- يمكن أن نستنتج أن وجود مشكلة الازدواج الخطي بين المتغيرات التفسيرية قد يتسبب في تغيير معنوية المعادلة من خلال تقدير قيمة F لتتحول من معنوية إلي غير معنوية أو العكس.

٦- يلاحظ انخفاض معامل الارتباط المتعدد (R) من ٠.٩٧ قبل معالجة مشكلة الازدواج الخطي إلي نحو ٠.٩٦٨ بعد معالجة هذه المشكلة بمعني انخفاض قوة الارتباط بين المتغيرات المستقلة والمتغير التابع بعد المعالجة.

٧- تسبب الازدواج الخطي في ارتفاع معامل التحديد (R^2) إلي نحو ٠.٩٤٨ ليأخذ قيمة أعلي من قيمته الحقيقية عند إزالة الازدواج الخطي بين المتغيرات التفسيرية وهي ٠.٩٣٨ مما أعطي قدرة أعلي في تفسير مساهمة المتغيرات المستقلة في إحداث التغير في المتغير التابع.

٨- ارتفعت قيمة الجذر التربيعي لمتوسط مربعات الخطأ ($Sigma$) من ٢٨.٧٧ قبل المعالجة إلي نحو ٣١.٥٣ بعد المعالجة.

٩- انخفاض قيمة معامل تضخم التباين (VIF) لأقل من ١٠ بعد معالجة مشكلة الازدواج الخطي لكل المتغيرات التفسيرية في النموذج القياسي النهائي.

١٠- انخفاض درجة الثقة في النتائج المتحصل عليها من التحليل الإحصائي وانخفاض القدرة علي التنبؤ باستخدام النموذج القياسي لتقدير العلاقة بين المتغيرات المستقلة والمتغير التابع في ظل مشكلة معالجة الازدواج الخطي.

بناء علي ما سبق توصي الدراسة بضرورة اهتمام الباحثين بتطبيق الفروض الأساسية لتحليل الانحدار المتعدد وتجنب إسقاط أي فرض منها وإتباع الاختبارات اللازمة للكشف عن المشاكل المتوقعة الناجمة عن إسقاط أي من هذه الفروض مع تحديد أسبابها بدقة، كما يتبين أهمية وفاعلية طريقة انحدار *Ridge* في معالجة مشكلة الازدواج الخطي بين المتغيرات التفسيرية لتقدير نموذج قياسي للعوامل المؤثرة علي إجمالي إنتاج اللحوم الحمراء في مصر.

المراجع:

١- أحمد كامل السيد أحمد (٢٠٠٩)، نماذج إحصائية مقترحة لتحليل دوال الاستهلاك باستخدام بيانات السلاسل الزمنية وأبحاث الدخل والإنفاق والاستهلاك في مصر، رسالة دكتوراه، قسم الإحصاء، كلية التجارة، جامعة عين شمس.

٢- دومينك سلفاتور (١٩٨٢)، ملخصات شوم ومسائل في الإحصاء والاقتصاد القياسي، ترجمة سعدية حافظ منتصر، دار ماكجروهيل للنشر القاهرة.

٣- ساوس الشيخ (٢٠١٤)، معالجة مشكلة الازدواج الخطي باستخدام انحدار الحرف (دراسة تطبيقية علي دالة الإنفاق الاستهلاكي في الجزائر خلال الفترة ١٩٧٠- ٢٠١١)، مجلة الحقيقة، جامعة أدرار، العدد (٢٩)، يونيه، الجزائر.

٤- عبد القادر محمد عبد القادر عطية (٢٠٠٥)، الحديث في الاقتصاد القياسي بين النظرية والتطبيق، الدار الجامعية، الإسكندرية.

٥- عبد المهدي رضا الجراح (٢٠٠٤)، مقارنة طريقتي المربعات الصغرى والمكونات الرئيسية في تحليل الانحدار باستخدام أسلوب المحاكاة، رسالة ماجستير، قسم الرياضيات، كلية الأدب، جامعة آل البيت، الأردن.

٦- مزاحم محمد يحيي، محمود حمدون عبد الله (٢٠٠٧)، تشخيص التعدد الخطي واستخدام انحدار الحرف في اختيار متغيرات دالة الاستثمار الزراعي في العراق في الفترة (١٩٨٠-٢٠٠٠)، مجلة تكريت للعلوم الإدارية والاقتصادية، جامعة تكريت كلية الإدارة والاقتصاد، مجلد (٣)، العدد (٨)، العراق.

٧- وزارة الزراعة واستصلاح الأراضي، قطاع الشئون. الاقتصادية، نشرة الميزان الغذائي، أعداد مختلفة.

٨- وزارة الزراعة واستصلاح الأراضي، قطاع الشئون الاقتصادية، نشرة الإنتاج الحيواني، أعداد مختلفة.

9- Bager, A.; M. Roman; M. Al-Gelidh and B. Mohammed (2017), Addressing Multicollinearity in Regression Models: A Ridge Regression Application, The Bucharest University of Economic Studies, Munich Personal RePEc Archive (MPRA), 11th International Conference of Applied Statistics, Brasov, 2nd of June

10- Berk, K. N. (1977), Tolerance and condition in Regression computations, Journal of American statistical ssocations, Vol.72.

11- Dorugade , A. V. (2014), On Comparison of Some Ridge Parameters in Ridge Regression, Sri Lankan journal of Applied Statistics Vol. (15-1).

12- E. Hoerl and R. W. Kennard (2000), Ridge Regression: Biased Estimation for No orthogonal Problems, Technometrics, Vol. 42, No. 1, Special 40th Anniversary Issue.

13- El-Dereny M. and N. I. Rashwan (2011), Solving Multicollinearity Problem Using Ridge Regression Models, Int. J. Contemp. Math. Sciences, Vol. (6), No. (12) Algeria.

14- EL-Habil A. M. and Kh. I.A. Almghari (2011), Remedy of multicollinearity using Ridge Regression, Department of Applied Statistics, Journal of Al Azhar University-Gaza (Natural Sciences), No. , 13 : 119-134.

15- Gorgees H. M. and B. A. Ali (2013), Employing Ridge Regression Procedure to Remedy the Multicollinearity Problem, Ibn Al- Haitham Jour. For Pure & Appl. Sci. , University of Baghdad, Vol. 26 (1). Iraq.

16- Gunst, R. F. and, R. L. Mason (1977), Biased estimation in regression: an evaluation using mean squared error. J. Amer. Statist. Assoc. 72.
17- Montgomery, D. C. G. C. Runger, (2002), Applied Statistics and Probability for Engineers, John Wiley & Sons, Inc. USA.
18- Patrick C. C. (2013), The Effects of Multicollinearity in Multilevel Models, Doctor of Philosophy, Wright State University.
19- Ridge Regression, NCSS statistical software, Chapter 335.
20- Tiwari M. and A. Sharma (2017), Predictive Efficiency of Ridge Regression Estimator, Yugoslav Journal of Operations Research, Number 2.

الملخص

يترتب علي إسقاط افتراض عدم وجود علاقة خطية تامة بين المتغيرات المستقلة في نموذج الانحدار الخطي المتعدد حدوث مشكلة الازدواج الخطي، عندئذ تظهر بعض الآثار غير المرغوبة والتي ينشأ عنها تقديرات غير دقيقة لمعالم المعادلة ومن ثم إضعاف قدرة النموذج الإحصائي علي التنبؤ في المستقبل بالمتغير التابع، وتستهدف الدراسة بيان طبيعة وأسباب مشكلة الازدواج الخطي وتحديد طرق الكشف عنه ومعالجته باستخدام طريقة انحدار ريدج كوسيلة فعالة في تخفيف آثار الازدواج الخطي، فضلاً عن مقارنة نتائجها مع نتائج طريقة المربعات الصغرى العادية لبيان أهم الآثار الناجمة عن هذه المشكلة، وقد اعتمدت الدراسة في إبراز هذه المشكلة في الجانب التطبيقي علي بيانات العوامل المؤثرة علي إجمالي إنتاج اللحوم الحمراء في مصر خلال الفترة (١٩٩٥– ٢٠١٥)، حيث تبين أن أهم الآثار الناجمة عن مشكلة الازدواج الخطي هي أنه قد يتسبب في تغيير إشارات ومعنويات ومقدرات معاملات الانحدار للمتغيرات التفسيرية وزيادة قيمة معامل الارتباط المتعدد ومعامل التحديد بقيمة أكبر من قيمتهم الحقيقية، وبالتالي انخفاض درجة الثقة في النتائج المتحصل عليها من التحليل الإحصائي وانخفاض القدرة علي التنبؤ بالعلاقة بين المتغيرات المستقلة والمتغير التابع، وتوصي الدراسة بضرورة اهتمام الباحثين بالكشف عن الازدواج الخطي وتحديد أسبابه ومعالجته بطريقة انحدار ريدج.

REMEDY OF EFFECTS OF MULTICOLLINEARITY IN MULTIPLE LINEAR REGRESSION MODEL

Gad, M. A. and H. F. M. Hussein
Cent. Lab. for Design and Stat. Analysis Res., ARC

ABSTRACT

Assuming that there is no complete linear relationship between the independent variables in the multiple linear regression models leads to the multicollinearity problem. Some undesirable effects, according to this problem, resulting imprecise estimators of the equation, and lower ability of the statistical model to predict the dependent variable in the future. The study aims at investigating the nature and causes of the multicollinearity problem, methods of detection and remedy of it using the Ridge regression method as an effective way of mitigating the effects of multicollinearity.

The study found that the main effect of the multicollinearity problem is that it may change the signs, significance and estimators of the explanatory variables. Thus increasing the value of the coefficient of multiple correlation and the coefficient of determination at a value greater than their true value. Consequently the low confidence in the results obtained from the statistical analysis and the low predictability of the relationship between independent and dependent variables. The study recommended detecting multicollinearity and identify its causes, and remedy of it using Ridge regression approach.