

معالجة البيانات غير التامة وتقديرها بطريقة انحدار المركبات الرئيسية

اسوان محمد طيب النعيمي

قسم نظم المعلومات الادارية / كلية الادارة والاقتصاد / جامعة الموصل

Mechanism of Missing Data and Estimating them by Principal Component Regression

Aswan Mohamed Taib Al-Nueimy

Department of Management Information System
College of Management and Economic / University of Mosul

Abstract

The mechanism of missing data requires knowing the reason behind missing them because different easy in which data is missed give rise to different specifications of the available sample. The data on which this study depends has been missed completely at random (MCAR) and changed into complete one. It can also be possible to find the magnitude of unbias after changing the data into complete one. Then observing the problem of multi linear variance which can be detected by Variance Inflation Factor (VIF) gauge. The estimation of data by the principal component method follows that. Finally the study applies this data on the variants of oil products and the consumption of the refined products in certain regions in the world (1000 barrels a day from 1970 to 2004).

المستخلص

ان معالجة البيانات غير التامة (المفقودة) يتطلب معرفة سبب الفقدان ، وان اختلاف الآلية التي فقدت بها البيانات يؤدي الى اختلاف مواصفات العينة المتاحة ، والبيانات المعتمدة في الدراسة فقدت تماماً وبشكل عشوائي (MCAR) ، وتم تحويلها الى بيانات تامة . ثم ملاحظة وجود مشكلة التعدد الخطي وكشف عنها بمقياس معامل تضخيم التباين (VIF) ثم تقديرها بطريقة انحدار المركبات الرئيسية وطبق ذلك على متغيرات إنتاج النفط واستهلاك المنتجات المكررة في أقاليم معينة من العالم (ألف برميل يومياً) للفترة من (1970-2004) .

مقدمة

ان فكرة انجاز بحث حول البيانات غير التامة جاءت من ظروف بلدنا العزيز وما تعرض له من ويلات الحروب حيث أدى ذلك الى فقدان الكثير من البيانات المهمة وفي جميع نواحي الحياة الاقتصادية والطبيعية والصحية والعلمية الصرفة ... الخ وتحويل هذه البيانات غير التامة الى بيانات تامة . كما ان أسباب الفقدان مختلفة ، منها ما يكون خارج عن ارادة المعنيين او تكون بإرادة المعنيين أي يكون مخطط لذلك بسبب الكلفة او المخاطرة او بسبب عدم توفير الإمكانيات للمعاينة ، كما ان اسباب الفقدان تؤدي الى اختلاف الآلية التي فقدت بها البيانات وان اختلاف هذه الآلية يؤدي الى اختلاف مواصفات العينة المتاحة للدراسة . وعند إيجاد المشاهدات المفقودة للمتغيرات التي تعاني من الفقدان فإنه لا بد من ظهور مشكلة تعدد العلاقة الخطية لخرق إحدى فروض التحليل ، لذلك وجب معالجتها بطريقة انحدار المركبات الرئيسية بعد الكشف عنها بمقياس تضخم تباين المعاملات (VIF) .

نبذة تاريخية

في حالة ظهور مشكلة فقدان قسم من البيانات للعينة غالباً ما يلجأ سابقاً الى حذف المشاهدات التي تتضمن قيماً مفقودة من العينة المدروسة والاعتماد على العينة المتبقية في التحليل ، وقد استخدم هذا الأسلوب في عام (1932) من قبل Yates الذي يعد اول من استخدم هذا الأسلوب لمعالجة البيانات غير التامة.

وفي عام (1939) جاء Bartlett بأسلوب حسابي يكافئ الاسلوب الذي جاء به Yates معتمداً على حساب التباين والتباين المشترك واستخدام نتائجه لحساب معالم نموذج الانحدار في حالة كون بعض قيم المتغير المستجيب Y^{s} مفقودة وقد تم تطبيق الاسلوب نفسه في عام (1952) من قبل Tacher ، في عام (1966) استخدم Affif & Elashoff نفس الاسلوب السابق ولكن في حالة كون المتغير المستجيب Y^{s} يتعرض لفقدان جزء من قيمه. واستمر البحث في هذا الموضوع لعام (1967) وعام (1969) لنفس العالم أعلاه واقتصر على نماذج الانحدار الخطي البسيط .

وفي عام (1986) قام Glynn & Laird باستخدام أسلوب التحليل للبيانات التي يعتمد فقدانها على القيم المفقودة نفسها أي (Not MAR) وحصل على نتائج معقولة قياساً بطرائق اخرى . وقدمت عدة تقييمات لهذه الطريقة أبرزها ما جاء به عام (1987) (Litter & Rubin). كما استخدم (الطائي ، 1998) البيانات غير التامة وتحليلها احصائياً في نماذج الانحدار المتعدد.

آلية البيانات غير التامة (المفقودة) Missing a Mechanisms

وباختلاف البيانات غير التامة تختلف الطرائق الإحصائية لتحليل هذه البيانات وايضاً تختلف الآلية التي تؤدي الى فقدان البيانات وان فهم هذه الآلية وتحديد طبيعتها يساعد كثيراً في اختيار الطريقة المناسبة للتحليل بل يعد المدخل لتشخيص الطريقة التي تقترب نتائجها من الامثلية للبيانات المدروسة . وتكون علاقة فقدان البيانات لمتغير معين بقيم المتغير نفسه او لقيم المتغيرات الأخرى وكما يلي :

- أ- ان فقدان X_j يكون مستقلاً عن قيم المتغيرات الأخرى وعن القيم المفقودة نفسها .
- ب- ان فقدان قيم X_j يعتمد على القيمة نفسها .
- ج- اعتماد القيم المفقودة X_j عن قيم المتغيرات الأخرى في العينة.

اذا كان سبب الفقدان مستقلاً عن القيمة المفقودة وعن قيم المتغيرات الأخرى نفسها في العينة عندها يمكن القول ان البيانات تفقد تماماً بشكل عشوائي (MCAR) Miss Completely at Random . اما اذا كان سبب الفقدان له علاقة بقيم المتغيرات الأخرى فقط ومستقل عن القيمة المفقودة ففي مثل هذه يكون فقدان البيانات بشكل عشوائي Missing at Random (MAR) ، واحياناً يكون سبب الفقدان ناتج عن القيمة المفقودة نفسها ومستقل عن قيم المتغيرات الأخرى . فالبيانات هنا لا تفقد بشكل عشوائي (not MAR) عند تحليل هذا النوع من البيانات ويجب اخذ آلية الفقدان بنظر الاعتبار . اما في حالة (MCAR) و (MAR) يمكن ان يهمل التوزيع اليه . (الطائي، 1998).

ويمكن توضيح كل ما سبق بالمخطط التالي :



طريقة المربعات الصغرى بالتعويض عن البيانات غير التامة

Lest Square (LS) on Imputed Data
(Little & Rubin , 1987)

يوجد العديد من الطرائق في إيجاد البيانات غير التامة التي فقدت تماماً بشكل عشوائي (MCAR) او التي فقدت بشكل عشوائي (MAR) او بشكل غير عشوائي (not MAR) ، لكن منها ما تبتعد عن استخدام المتغيرات التي تتضمن قيماً مفقودة ومنها ما تستخدم البيانات المتاحة وفي بعض التطبيقات تتخلى عن جزءاً منها كما في حساب التباينات المشتركة ومعامل الارتباط . اما طريقة المربعات الصغرى بالتعويض عن البيانات غير التامة بأسلوبها الأول التعويض بالمتوسط الشرطي Conditional mean Imputation والاخر اسلوب التعويض بالمتوسط غير الشرطي Unconditional mean Imputation اذ تعتبر هذه الطريقة بأسلوبها بانها عامة ومرنة في معالجة مشكلة البيانات غير التامة حيث تعتمد هذه الطريقة على اشغال مواقع القيم المفقودة بقيم تقديرية تختلف بأسلوب تقديرها ولكن الهدف هو إكمال العينة واستخدامها في التحليل النهائي ، اذ تعتبر هذه الطريقة ايجابية من جهة لانها توفر بيانات بديلة عن البيانات المفقودة بشكل حسابي ولكنها تعتبر سلبية لانها تستخدم الطرائق التقليدية في تحليل بيانات العينة المكملة التي تضم بيانات حقيقية (مشاهدة) وبيانات تقديرية (محسوبة) ، والأخيرة قد تكون عرضة للتحيز عند تقديرها وهذا يرجع الى الأسلوب المعتمد في التقدير ومدى امكانية الحصول على تقديرات كفاءة للبيانات المفقودة وسوف نستخدم أسلوب التعويض غير الشرطي للبيانات الغير التامة التي فقدت تماماً بشكل عشوائي في هذا البحث.

التعويض بالمتوسط غير الشرطي Unconditional mean Imputation

يتم التعويض عن القيم المفقودة لاي متغير تحت الدراسة بمتوسط قيم المشاهدة له

$$\tilde{X}_j = \sum_{n_j} X_{obs} / n_j \dots\dots\dots(1)$$

n_j : يمثل عدد القيم المشاهدة فعلاً للمتغير X_j

وفي حالة إحلال \tilde{X}_j محل القيم المفقودة X_{mis} لـ X_j فان المتوسط العام للمتغير \bar{X}_j يكون مساوياً لـ \tilde{X}_j وتباين X_j للعينة المكملة هو حيث

$$(n_j - 1)S_{jj} / (n - 1) \dots\dots\dots(2)$$

S_{jj} : يمثل التباين المقدر من القيم المتاحة حسب الصيغة

$$\tilde{S}_{jj}^2 = \sum_{i=1}^{n_j} (X_{ij} - \tilde{X}_j)^2 / n_j \dots\dots\dots(3)$$

$$\tilde{S}_{kk}^2 = \sum_{i=1}^{n_k} (X_{ik} - \tilde{X}_k)^2 / n_k \dots\dots(4)$$

$$\tilde{X}_j = \sum_{n_j} X_{ij} / n_j \dots\dots\dots (5)$$

$$\tilde{X}_k = \sum_{n_k} X_{ik} / n_k \dots\dots\dots(6)$$

وتحت فرض ان البيانات غير التامة بشكل عشوائي تماماً (MCAR) فان S_{jj} هو تقدير للتباين الحقيقي كذلك فان تباين العينة متحيز بالمعامل $(n_j-1)(n-1)$ لان التعويض عن القيم المفقودة تم باستخدام مركز التوزيع المتمثل بالوسط الحسابي وهذا ينسحب على التباين المشترك للمتغيرين X_k , X_j اللذان يتضمنان بيانات حقيقية مشاهدة وبيانات تقديرية محسوبة فيكون

$$[(n_{jk} - 1)/(n - 1)]\tilde{S}_{jk} \dots\dots\dots (7)$$

حيث \tilde{S}_{jk} محسوبة وفق الصيغة

$$\tilde{S}_{jk} = \sum_{i=1}^{n_{jk}} (X_{ij} - \tilde{X}_j)(X_{ik} - \tilde{X}_k) / (n_{jk} - 1) \dots\dots (8)$$

وهو تقدير متسق للتباين المشترك المحسوب على وفق هذه الطريقة ومتحيزاً بالمعامل

$$(n_{jk} - 1)/(n - 1) \dots\dots\dots (9)$$

وهذا بالإضافة عن إمكانية ان تكون مصفوفة التباين والتباين المشترك (Positive semi definite) .

انحدار المركبات الرئيسية Principal Component Regression

(الشكري ، 2005) ، (المشهداني ، 1994)

في حالة مصادفة معادلة الانحدار الخطي المتعدد مشكلة تعدد العلاقة الخطية لوجود ارتباطات بين المتغيرات التنبؤية إذ أن التقديرات التي تنتج لدينا تكون متأثرة بالعلاقات بين المتغيرات وليس نتيجة العلاقة بين المتغير المستجيب والمتغيرات التوضيحية ، وذلك يعني خرق احد فروض التحليل التي تنص على أن أعمدة وصفوف المتغيرات التوضيحية X يجب ان تكون مستقلة خطياً مع بعضها البعض إذ توجد العديد من الطرائق المتحيزة للتخلص من مشكلة تعدد العلاقة الخطية بعد الكشف عنها بإحدى طرق الكشف كمقياس تضخم تباينات المعاملات (VIF) التي ستذكر لاحقاً . ومن ثم استخدام طريقة المركبات الرئيسية المتحيزة في التحليل .

طريقة انحدار المركبات الرئيسية

تعتمد طريقة تحليل المركبات الرئيسية على أسلوب تحويل المتغيرات التوضيحية الأصلية الى متغيرات جديدة غير مرتبطة . هذه المتغيرات الجديدة تسمى بالمركبات الرئيسية، حيث ان كل مكون رئيسي هو عبارة عن تركيبة خطية في المتغيرات المستقلة الأصلية . ويتم تحويل المتغيرات المستقلة الى المكونات الرئيسية بالشكل الآتي :

$$Y = \beta_0 \underline{I} + XVV'\underline{\beta} + \underline{U} \dots\dots\dots (10)$$

حيث V عبارة عن مصفوفة المتجهات المميزة لمصفوفة الارتباط بين المتغيرات التوضيحية فاذا عوضنا عن XV بكمية ثابتة PC مصفوفة ذات بعد $(n \times p)$ أعمدها عبارة عن معاملات انحدار النموذج المحور الذي يأخذ الشكل الآتي :

$$Y = \beta_0 \underline{I} + PC\underline{\alpha} + \underline{U} \dots\dots\dots (11)$$

كما ويعرف مقدار التحيز للمكونات الرئيسية كالاتي :

$$b(\beta_{PC}^0)E(\hat{\beta}_{PC}) - \beta = VV'\underline{\beta} \dots\dots\dots (12)$$

Variance Inflation Factors (VIF) مقياس تضخم تباينات المعاملات

(سعيد ، 1996)

يعد هذا المقياس سبباً كافياً لإهمال المتغير X_j من التحليل او استخدام طريقة اخرى كبديل عن المربعات الصغرى في التقدير .

وقد اكد بعض الباحثين ان مقدار المقياس $[a_{jj} = VIF \geq 10]$ فانه دليل على خرق فروض التحليل لوجود علاقة بين المتغيرات التوضيحية، ويعتمد هذا المقياس على فحص عناصر القطر الرئيسي للمصفوفة $(X'X)^{-1}$ حيث ان

$$a_{jj} = (1 - R_j^2) \dots\dots\dots (13)$$

وان R_j^2 عبارة عن معامل التحديد لانحدار X_j عن بقية المتغيرات التوضيحية ، توضح المعادلة (13) بانه لوجود تداخل خطي بين المتغيرات التوضيحية فان R_j^2 ستكون قريبة او مساوية للواحد وبذلك تكون a_{jj} كبيرة الحجم . اما اذا كانت X_j مستقلة عن بقية المتغيرات التوضيحية الأخرى $R_j^2 = 0$ وبذلك تكون $a_{jj}=1$.

جمع البيانات

أخذت البيانات من (2004) OPEC Annual Statistical Bulletin والمعتمدة في أطروحة دكتوراه* . والبيانات تمثل إنتاج النفط واستهلاك المنتجات المكررة في اقاليم معينة من العالم (الف برميل يومياً) للفترة من (1970-2004) لـ 35 مشاهدة ولسبعة متغيرات هي :

X_1 إنتاج أوروبا وأمريكا من النفط

X_2 استهلاك أوروبا وأمريكا من المنتجات المكررة

X_3 العجز الأوروبي والأمريكي بين إنتاج النفط واستهلاك المنتجات المكررة

X_4 إنتاج آسيا والهادي من النفط

X_5 إنتاج آسيا والهادي من المنتجات المكررة

X_6 عجز آسيا والهادي بين إنتاج النفط واستهلاك المنتجات المكررة

Y تمثل عدد سنوات الدراسة للفترة من (1970-2004)

وآلية فقدان البيانات غير التامة بشكل عشوائي تماماً (MCAR).

التحليل الإحصائي

1- طريقة المربعات الصغرى بالتعويض عن البيانات غير التامة

أخذت احد المتغيرات وليكن X_4 الذي يعاني من نقص في مشاهداته . إذ تم إيجاد المشاهدة المفقودة بالتعويض بالمتوسط غير الشرطي حسب الصيغة (1) والتباين حسب الصيغة (3) . وكانت قيمة المتوسط والتباين على التوالي عندما احد المشاهدات المفقودة للمتغير X_4 أي $n=34$ هي 5402.08 ، اما التباين 3224733 وعند التعويض عن القيمة المفقودة بالمتوسط 5402.08 فان المتوسط المقدر هو 5402.08 والتباين المقدر 3129888 أي عندما $n=35$.

وللمقارنة مع المتوسط والتباين للمتغير X_4 قبل فقدان احد قيمه أي عندما $n=35$ الحقيقية قبل فقدان فان المتوسط يساوي 5341.96 والتباين يساوي 3256378 من ذلك يمكن ان نستنتج ان المتوسط والتباين المقدر قريب من المتوسط والتباين الحقيقي .

* اطروحة دكتوراه مقدمة من عمار محمد سلو احمد ، كلية الادارة والاقتصاد ، جامعة الموصل (2006) الموسومة " السياسة الانتاجية والسعرية للملكة العربية السعودية " .

2- الكشف عن مشكلة تعدد العلاقة الخطية بمقياس معامل تضخم التباين (VIF)

من ما تقدم وبعد ايجاد القيمة المفقودة لاي من المتغيرات لابد من ان يكون هنالك وجود مشكلة تعدد علاقة خطية بين المتغيرات وهذا يعني خرق لإحدى فروض التحليل ويمكن ملاحظة ذلك من خلال معامل تضخم التباينات (VIF) بين المتغيرات التوضيحية بالنسبة لبيانات إنتاج النفط واستهلاك المنتجات المكررة في أقاليم معينة من العالم . ان هذا المقياس يعتمد اذ كان (VIF) اكبر من (10) فانه دليل على ان المتغير X_j يعاني من وجود مشكلة تعدد العلاقة الخطية ، اذ يلاحظ VIF للمتغيرات التوضيحية لبيانات إنتاج النفط كما في الجدول رقم (1) .

جدول رقم (1)

يمثل مقياس تضخم التباين VIF لمتغيرات إنتاج النفط

X_j	VIF
X_1	3.4
X_2	4.3
X_3	1.4
X_4	7.5
X_5	72.3*
X_6	49.6*

$$J=1,2,3,4,5,6$$

* تعني ان هذه المتغيرات تعاني من تضخم تباين معاملاتها

وبما ان قيم تضخم تباين المعاملات VIF في الجدول رقم (1) للمتغيرات التوضيحية لكل من X_6, X_5 اكبر من (10) فان هذا يعني ان هذه المتغيرات تعاني من مشكلة تعدد العلاقة الخطية بين المتغيرات التوضيحية أي وجود علاقة او ارتباط بين المتغيرات التوضيحية .

3- طريقة المركبات الرئيسية

بعد تحليل مصفوفة الارتباط البسيط تمكنا من تمييز مكونين رئيسيين فسر ان نسبة (84%) من اجمالي التباين الكلي . كما تم تمييز المتغيرات المؤثرة في كل مكون رئيس من خلال مصفوفة المتجهات المميزة كما يلي : عكس المكون الرئيسي الأول أهمية كبيرة بتفسير ما نسبته (57.8%) من التباين الكلي واشتمل هذا المكون على متغيرين ذي تأثير معنوي وهو المتغير (X_5) والمتغير (X_6) كما ظهر المتغيران (X_1) و (X_3) ذات تأثير معنوي في المكون الرئيس الثاني والذي فسر ما نسبته (26.5%) من التباين الكلي فضلاً عن ذلك لوحظ

عدم أهمية المتغيرين (X2) و (X4) في كلا المكونين الرئيسيين ويمكن توضيح ما سبق في الجدول رقم (2) الجذور المميزة للمتغيرات . اذ يوضح هذا الجدول أي الجذور المميزة الأكبر من الواحد التي على أساسها تم اختبارهم ثم استخدام المكونات الرئيسية .

جدول رقم (2)

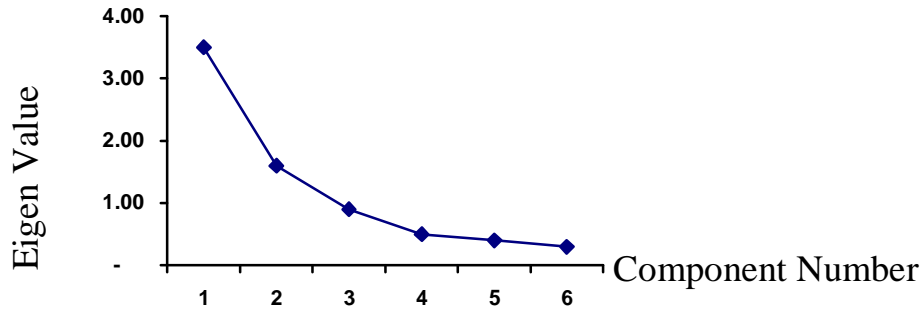
الجذور المميزة للمتغيرات

قيمة الجذر المميز	الجذور المميزة
3.4654**	λ_1
1.5896**	λ_2
0.6137	λ_3
0.1933	λ_4
0.1298	λ_5
0.0081	λ_6

** يمثل الجذور المميزة والتي قيمتها أكبر من الواحد .

شكل (1)

عدد الجذور المميزة التي اختيرت



يوضح الشكل رقم (1) الجذور المميزة والتي قيمتها أكبر من الواحد التي اعتمدت في طريقة المركبات الرئيسية .

وبعد اعتماد طريقة المركبات الرئيسية تم الكشف مرة أخرى على مشكلة التعدد الخطي بطريقة (VIF) فلو حظ أن المتغيرات التي كانت تعاني من تضخم في تباين معاملاتها وهي X_5 ، X_6 ، أصبح معامل تضخم التباين على التوالي للمتغيرات (1.0,1.0) وهذا يعني زوال مشكلة تعدد العلاقة الخطية للمتغيرات قيد الدراسة لاعتمادنا طريقة المركبات الرئيسية للبيانات التي كانت غير تامة ومعالجتها وتحويلها الى بيانات تامة وتقديرها بطريقة المركبات الرئيسية.

الاستنتاجات

يمكن توضيح أهم الاستنتاجات بما يلي :

- 1- ان حل مشكلة البيانات المفقودة تغني عن البحث وبذل المزيد من الوقت والجهد والتكلفة في إيجاد طرق او بيانات أخرى لحل مشكلة فقدان في البيانات قيد الدراسة او البحث .
- 2- ان التعرف على آلية فقدان للبيانات يعتبر المفتاح لتشخيص أسلوب التحليل المعتمد للظاهرة . كما وان الاستلال حول البيانات غير التامة قد يكون توزيعاً مجهولاً او توزيعاً طبيعياً للمتعدد المتغيرات خاصة اذا كان حجم العينة كبيراً .
- 3- كما وان البيانات الغير تامة بعد معالجتها بالأسلوب المناسب لنوع الفقدان يمكن ان تظهر مشكلة التعدد الخطي لخرق إحدى فروض التحليل كما ويمكن الكشف عنها ومعالجتها بإحدى طرق التقدير المتحيزة كما ورد في بحثنا وتقديرها بطريقة انحدار المركبات الرئيسية.
- 4- يمكن تقدير مقدار التحيز للبيانات غير التامة بعد معالجتها مهما كانت آلية الفقدان واعتمادها كبيانات تامة في كافة التطبيقات الإحصائية.

المصادر

أ- المصادر العربية

- البلداوي ، عبد الحميد عبد المجيد (2004) : " الأساليب الإحصائية التطبيقية " ، دار الشرق الأوسط ، التوزيع الأردن .
- الشكرجي ، ذنون يونس ذنون (2005) : "استخدام مصفوفتي Q-Mode – R-Mode في التحليل العاملي" ، رسالة ماجستير ، كلية الادارة والاقتصاد ، جامعة الموصل .
- الطائي ، خالد ضاري عباس (1998) : "التحليل الإحصائي للبيانات غير التامة في نماذج الانحدار المتعدد" ، أطروحة دكتوراه ، كلية الإدارة والاقتصاد ، جامعة الموصل .
- المشهداني ، ايمان محمد عبد الله (1994) : "استخدام المركبات الرئيسية في تشخيص ومعالجة مشكلة التعدد الخطي مع تطبيق عملي لبعض الظواهر الاقتصادية" ، رسالة ماجستير ، كلية الإدارة والاقتصاد ، جامعة بغداد .

- سعيد ، هيفاء عبد الجواد (1996) : "طرق التعرف على تعدد العلاقة الخطية وكيفية معالجتها بطرائق التقدير المتحيزة" ، رسالة ماجستير ، كلية الإدارة والاقتصاد ، جامعة الموصل.

- سلو ، عمار محمد (2006) : "السياسة الإنتاجية والسعرية للملكة العربية السعودية في مجال النفط الخام" ، رسالة دكتوراه ، كلية الإدارة والاقتصاد ، جامعة الموصل .

- كاظم ، أموري هادي (1998) : "طرق القياس الاقتصادي" ، جامعة بغداد ، مطبعة التعليم العالي ، الطبعة الأولى.

- كمال ، غفران إسماعيل (1990) : "دراسة تقويمية لطرق التقديرات المتحيزة لمعامل الانحدار المتعدد عند مخالفة بعض الفرضيات باستخدام أسلوب المحاكاة" ، رسالة ماجستير في الإحصاء ، كلية الإدارة والاقتصاد ، جامعة بغداد. -

ب- المصادر الأجنبية

- Gourieroux, G. and Mont Fort (1981) “ on the problem of Missing Data an Linear Models” Review of Economic Studies, XL VIII, P: 579-586.
- Little, and Rubin, (1987) “Statistical Analysis with Missing Data”, New York: John-Wiley, P: 88,125-134.
- Little R.J.A. (1988) “Robust Estimation of the Mean Covariance Matrix from Data with Missing Values” Applied Statistics, P: 37,33-29.
- Theil, H. (1971) “Principal of Econometrics” John Wiley & Sons. Inc. P: 90-94