

استخدام معايير الدقة التنبؤية في تحديد الطريقة المثلى في تقدير القيم المفقودة
(بيانات البحوث الزراعية نموذجا)

علي درب كسار الحبالي

استاذ مساعد

جامعة بغداد - كلية الزراعة - قسم الاقتصاد الزراعي

adk_1966@yahoo.com

المستخلص

تمثل مشكلة الفقد في بعض مشاهدات السلسلة الزمنية اهمية خاصة لاسيما في عملية تقدير النماذج الرياضية والاحصائية من ثم التنبؤ بها لان التوصل الى نتائج معينة من خلال نتائج هذا التحليل المبني على تقدير المشاهدات المفقودة بطرائق معينة سيكون له ابلغ الاثر في القرارات المبنية على هذه النتائج لاسيما في البحوث الزراعية، الامر الذي استوجب دراسة طرائق تقدير المشاهدات المفقودة والمتمثلة بطريقة متوسط السلسلة الزمنية وطريقة المتوسط الحسابي للقيم المتجاورة والوسيط للقيم المتجاورة وطريقة الاستكمال الخطي واخيرا طريقة التعويض بالانحدار واختبارها باستخدام معايير الدقة التنبؤية وهي نسبة متوسط القيمة المطلقة للاخطاء (MAPE) (Mean Absolute Percentage Error)، ومتوسط القيمة المطلقة للاخطاء (MAE) (Mean Absolute Error)، ومتوسط مربع الاخطاء (MSE) (Mean Square Error)، فضلا عن استخدام نماذج الانحدار البسيطة ومقارنة بعض الاختبارات الاحصائية المتحصل عليها من هذه النماذج كاختبار F واختبار R^2 واختبار R لتدعيم النتائج المتحصل عليها من اختبارات الدقة التنبؤية للحكم على افضل الطرائق في تقدير المشاهدات المفقودة، وتوصل البحث الى افضلية طريقة التعويض بالانحدار من خلال تفوقها باختبارات الدقة التنبؤية فضلا عن اختبارات نموذج الانحدار الخطي البسيط، وبالتالي يمكن الجزم بتفوق نتائج القيم المقدرة بهذه الطريقة والتي على اساسها يمكن الحكم على القرار الذي سوف يتم تبنيه من قبل صاحب القرار بانه قرار يحمل اخطاء تنبؤية صغيرة بالمقارنة مع نتائج الطرائق الاخرى، ويوصي البحث باستخدامها ضمن شروط معينة، مع الاخذ بنظر الاعتبار ضرورة ملائمة البيانات للطريقة المستخدمة في تقدير المشاهدات المفقودة لغرض الحصول على نتائج يعتد بها، كما يوصي البحث باهمية اجراء بحوث حول الحالات الاخرى التي لم يتطرق اليها البحث لغرض وضع صورة متكاملة لاسيما وان واقع البيانات في العراق يفتقد الدقة وبالتالي تأثر النتائج المتحصل عليها وتكون القرارات المتخذة غير دقيقة ولاسيما في القطاع الزراعي.

الكلمات المفتاحية: المشاهدات المفقودة، التنبؤ، طريقة التعويض بالانحدار، القيم المفقودة.

The Iraqi Journal of Agricultural Sciences - 44(4): 509-517, 2013

Al-Hiyali

USING FORECASTING ACCURACY CRITERIA TO DETERMINE OPTIMUM
METHOD FOR ESTIMATING MISSING VALUES
(AGRICULTURAL RESEARCHES DATA AS A CASE STUDY)

Ali D. K. Al-Hiyali

Assistant Professor

Dept . of Agricultural Economics, College of Agriculture, University of Baghdad

adk_1966@yahoo.com

ABSTRACT

Missing data in time series is considered as an important issue especially in the process of mathematical and statistical model estimation and consequently its forecasting. Arriving at certain results through the analysis based on the estimation of missing values by using different methods would have great effect on decisions based on the results especially in agricultural researchers. This requires studying the methods of missing data estimation represented by time series average method, arithmetic mean for adjacent values, median for adjacent values, linear interpolation and regression imputation and testing them by using forecasting accuracy criteria such as mean absolute percentage error (MAPE), mean absolute error (MAE), and mean square error (MSE) in addition to using simple regression models. Then comparing some statistical tests resulted from these results such as F -test, R^2 -test and R -test to support the resulting results from forecasting accuracy tests to judge for the best methods in estimating missing values. The results of this research showed the suitability of regression imputation method according to its advantage in forecasting accuracy tests in addition to simple linear regression model tests which this research is advising to be used within the condition explained within the research text. It is necessary that data are suitable to get valuable results.

Key words: missing observations, forecasting, preliminary test estimator, missing values, data imputation method, regression imputation.

و 23 و 24)، وتطرق مجموعة اخرى من البحوث والدراسات الى الاختبارات الاحصائية الواجب اجراؤها عندما تنفذ البيانات بعض المشاهدات (1 و 13 و 18).

الاطار النظري

تعددت الطرائق التي حاولت تقدير القيم المفقودة في بيانات السلسلة الزمنية وسيتم تناول خمس طرائق وهي (3):-

1- المتوسط الحسابي لبيانات السلسلة الزمنية
2- المتوسط الحسابي لاقرب النقاط حول القيم المفقودة Mean of nearby points وقد تكون نقطتان او اكثر وتسمى ايضا طريقة المتوسطات المتحركة (Moving averages).

3- الوسيط لاقرب النقاط حول القيم المفقودة Median of nearby points.

4- طريقة الاستكمال الخطي Linear Interpolation والذي يعتمد اخر قيمة قبل القيمة المفقودة واول قيمة بعد القيمة المفقودة في تقدير القيم المفقودة. وفي حال ان القيمة المفقودة كانت في اول السلسلة او اخرها فان هذه الطريقة لاتصلح لتقدير القيم المفقودة.

5- تقدير القيم المفقودة بالقيم التنبؤية Linear trend at point المتحصل عليها من انحدار قيم السلسلة المتوفرة (المتغير التابع على المتغير المستقل) الذي يأخذ قيما تسلسليا من 1 الى n، وتسمى ايضا طريقة التعويض بالانحدار (Regression imputation).

وادناه توضيح مبسط لكل طريقة:

1- الوسط الحسابي لبيانات السلسلة الزمنية: يستخدم الوسط الحسابي في هذه الطريقة لتقدير القيم المفقودة إذ يتم وضع المتوسط الحسابي للسلسلة الزمنية بدلا عن القيم المفقودة، وتجدر الاشارة الى ان استخدام الوسط الحسابي هنا سيكون متأثرا بالخصائص التي يمتاز بها الوسط الحسابي وأهمها هو تأثره بالقيم المتطرفة في البيانات وياخذ الوسط الحسابي الصيغة الاتية للبيانات غير المبوبة:

$$\bar{X} = \frac{\sum X_i}{n}$$

2- الوسيط Median هنا يتم حساب الوسيط لاقرب نقطتين او اكثر المتواجدة حول القيمة المفقودة . والوسيط هو القيمة x بحيث أن $P(X < x) \leq \frac{1}{2}$ و

يؤدي فقدان بعض المشاهدات في السلسلة الزمنية الى التأثير في عملية تقدير النموذج الانحدار ومن ثم التأثير في معاملات النموذج، فضلا عن قدرته التنبؤية، ويعاني كثير من الباحثين من هذه المشكلة والتي تؤدي الى جعل النماذج المقدر غير مؤهلة للاعتماد عليها في اتخاذ القرارات لاسيما اذا ما كانت هذه النماذج تمثل اهمية قصوى في معالجة مشكلة ما زراعية كانت ام صناعية او غير ذلك، وقد تعلق الامر بالجانب الزراعي فان الامر يكتسب اهميته عندما يتطلب تقديم توصية لصانع القرار من خلال المعاملات التي يتم التوصل اليها في النموذج الانحدار. تجدر الاشارة الى ان البيانات الزراعية في العراق تقتد الى الدقة وفقدان بعض البيانات خلال السلسلة الزمنية، فضلا عن وجود اكثر من مصدر للبيانات نفسها الامر الذي يضع الباحث امام مفترق طرق في اختيار اي من هذه البيانات لاغراض بحثه من ثم تقديم التوصية لاصحاب القرار لاسيما في القطاع الزراعي حول المعاملات المقدر لاسيما في نماذج الانحدار، ويكتسب الامر اهمية اخرى عندما يتم استخدام هذه النماذج للتنبؤ بواقع الحال للمستقبل وما يحمله من غموض ومن خلال هذه التنبؤات يتم التخطيط لهذا المستقبل. تتركز مشكلة البحث في الاثار السلبية التي يمكن ان تنشأ عن فقدان بعض المشاهدات في السلسلة الزمنية ولاسيما عند القيام بالتنبؤ بالنماذج المقدر من تلك البيانات وما ينشأ عنها من تشوه في قيم المعاملات المقدر فضلا عن القيم المتنبأ بها الامر الذي يجعل من التوصية باللاخذ بقيم هذه المعاملات في البحوث الزراعية امرا معقدا ومشكوكا فيه. يهدف البحث الى استخدام معايير الدقة التنبؤية في الحكم على الطريقة المتبعة في تقدير القيم المفقودة في بيانات السلسلة الزمنية ولاسيما في البحوث الزراعية او تلك المعتمدة على بيانات تخص القطاع الزراعي وصولا الى الحكم باهمية هذه الطريقة من غيرها. تناولت مجموعة كبيرة من الدراسات والبحوث موضوع البيانات المفقودة وتأثيرها في النماذج المقدر، فمنها من تناول نماذج السلاسل الزمنية غير الخطية وتقدير المشاهدات المفقودة (1)، في حين تناول البعض الاخر ميكانيكية الفقد في البيانات باستخدام طرائق معينة (2) و 4 و 5 و 6 و 7 و 8 و 10 و 14 و 15 و 17 و 19 و 21 و 22

$$y = \hat{B}_0 + \hat{B}_1 x_{i1} + \dots + \hat{B}_k x_{ik}$$

$$i = r+1, \dots, n$$

تجدر الإشارة الى ان هذه الطريقة تحافظ على الحالات مع القيم المفقودة ويصون حجم العينة.

ان الملاحظات التي تم ادراجها اعلاه ينبغي ان تؤخذ بنظر الاعتبار عند استخدامها في عملية التنبؤ ومن ثم الحكم على افضلها من خلال معايير الدقة التنبؤية الاتية (9):

1- نسبة متوسط القيمة المطلقة للاخطاء وصيغته هي

$$MAPE = \frac{\sum |Y_i - \hat{Y}_i|}{n} \times 100 \text{ :-}$$

2- متوسط القيمة المطلقة للاخطاء وصيغته هي :-

$$MAD = \frac{\sum |Y_i - \hat{Y}_i|}{n}$$

3- متوسط مربع الاخطاء وصيغته هي :-

$$MSD = \frac{\sum (Y_i - \hat{Y}_i)^2}{n}$$

اذ ان :

Y_i : القيمة الفعلية للمتغير \hat{Y}_i : قيم المتغير المقدرة من الانموذج وتجدر الإشارة الى ان قاعدة اتخاذ القرار بافضلية احدى الطرائق على غيرها تقاس بناء على اصغر قيمة للمعايير اعلاه.

المواد والطرائق

تم استخدام بيانات اجمالي تكوين راس المال الزراعي في العراق (بالأسعار الثابتة) للمدة 1980-2009، بفقدان مشاهدة واحدة خلال العام 2003 وهي مشاهدة مفقودة اصلا ولم يتم حذفها من البيانات، وتجدر الإشارة الى ان البحث لا يتطرق الى انماط البيانات المفقودة ولا الى آليات فقدانها، اي ان البحث سيتطرق الى فقدان واحدة من المشاهدات وسيتم تعويضها باحدى الطرائق المشار اليها في الاطار النظري، ومن خلال استخدام معايير الدقة التنبؤية سيتم الحكم على افضل طرائق تقدير القيم التنبؤية فضلا عن معايير اخرى كالانحراف القياسي للبيانات بعد تقدير المشاهدات المفقودة. يشير الجدول 1 الى بيانات اجمالي تكوين راس المال الزراعي (بالاسعار الثابتة) في العراق خلال المدة 1980-2009 (18).

، وبلغة اخرى يكون الوسيط القيمة التي $P(X > x) \leq \frac{1}{2}$

يكون نصف القيم x_1, x_2, \dots, x_n أكبر منها ونصفها اصغر منها، وبحسب الوسيط بطريقتين حسب البيانات المتوافرة وكما يأتي:-

البيانات الزوجية: يكون الوسيط هو القيمة التي ترتيبها $\frac{n}{2}$

البيانات الفردية: الوسيط هو الوسط الحسابي للقيمتين اللتين

$$\frac{n}{2}, \frac{n}{2} + 1 \text{ ترتيبهما هما:-}$$

$$Me = \frac{n/2 + (n/2) + 1}{2}$$

3- الاستكمال الخطي: وتعد من ابسط الطرق ويطلق عليها

احيانا مختصر (lerp) وبشكل عام فان طريقة الاستكمال الخطي تاخذ نقطتين ولتكن مثلا (x_a, y_a) و

(x_b, y_b) ، وعليه يكون الاستكمال بالصيغة الاتية (12):

$$y = y_a + (y_b - y_a) \frac{(x - x_a)}{(x_b - x_a)} \text{ at the point } (x, y)$$

ومن مميزات هذه الطريقة انها سهلة وسريعة غير انها ليست دقيقة بالشكل الكافي . كما ان النقصان الذي تعاني منه هذه الطريقة ان الاستكمال قد يكون غير خطي اي متعدد الحدود (Polynomial interpolation) ، ولهذا سيفترض البحث ان العلاقة بين متغيرات السلسلة الزمنية هي علاقة خطية عندما يتم التقدير بهذه الطريقة.

4- تعويض الانحدار (Regression Imputation)،

وتعتمد على تعويض القيم المفقودة بقيم مشتقة من معادلة الانحدار بالاعتماد على متغيرات في مجموعة البيانات والتي لا تحتوي قيما مفقودة وذلك باستخدام معلومات عن متغيرات مساعدة عددها K (x_1, x_2, \dots, x_k) ، وافرض انه تم مشاهدة هذه المتغيرات بشكل كامل - استجابة كاملة. وافرض ان المتغير y شوهد لأول r من القيم ولم يشاهد للقيم الاخيرة $n - r$. ويعتمد تعويض الانحدار على ملائمة انموذج الانحدار بالاعتماد على الحالات الكاملة والتي عددها r (8) :

$$y = B_0 + B_1 x_{i1} + \dots + B_k x_{ik} + \varepsilon_i$$

$$i = 1, \dots, r, r+1, \dots, n$$

اذ ان:-

$$\varepsilon_i \sim (0, \sigma^2) \text{ اخطاء عشوائية مستقلة.}$$

ثم نملاً القيم المفقودة بتنبؤات الانحدار وكما يأتي:

جدول 1. اجمالي تكوين رأس المال الزراعي في العراق خلال المدة (1980-2009) بالاسعار الثابتة (مليون دينار)

السنوات	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	
1980	827.14	1985	611.27	1990	254.00	1995	51.84	2000	138.69	2005	70.00
1981	948.90	1986	454.83	1991	51.77	1996	19.43	2001	149.00	2006	288.00
1982	972.55	1987	362.54	1992	125.67	1997	107.74	2002	228.10	2007	75.50
1983	754.11	1988	437.40	1993	126.45	1998	92.36	2003	-	2008	101.60
1984	694.88	1989	435.55	1994	97.13	1999	102.85	2004	107.10	2009	213.60

*عدم توافر بيانات

طرائق عديدة مثل الاوساط المتحركة (Moving Averages) ، وطرائق التمهيد الاسي المفرد والمزدوج (Single & Double Exp Smoothing) ، وطريقة ونتر (Winters Method) ، وطريقة الانحدار الذاتي والمتوسطات المتحركة (ARIMA) (16) . غير ان هدف البحث هو الحكم على افضلية طرائق تقدير القيم المفقودة من خلال معايير الدقة التنبؤية الامر الذي حدا بالباحث الاكتفاء بهذا خدمة لاغراض البحث. يشير الجدول 4 الى تفوق الطريقة السابعة وهي طريقة التعويض بالانحدار في استخدامها لتقدير القيمة المفقودة من خلال تحقيقها لاقل قيمة للمعايير الثلاثة اذ جاءت بالمرتبة الاولى تليها الطريقة الرابعة ثم الثانية وتليها الخامسة ومن بعدها الثالثة ثم السادسة واتخذت الطريقة الاولى المرتبة الاخيرة، وهذه النتائج تشابهت في المعايير الثلاثة الامر الذي يؤكد تفوق الطريقة الاولى في ذلك، وباستخدام معياري المتوسط الحسابي والانحراف المعياري فقد كان ترتيب الطرائق حسب اقل قيمة لكل من المعيارين كما في الجدول 5.

تم تقدير القيمة المفقودة في عام 2003 بخمس طرائق (وتم اعتبار كل من الطريقة الثالثة والرابعة وهما تابعتان الى طريقة واحدة طريقتين منفصلتين، وكذلك الحال فيما يتعلق بالطريقة الرابعة والخامسة) وذلك للمقارنة بين النتائج وبهذا يصبح مجموع الطرائق سبعة، وهذه الطرائق هي:

- 1- الوسط الحسابي للسلسلة الزمنية
- 2- الوسط الحسابي للقيم المجاورة وبيانات مشاهدتين قبل وبعد المشاهدة المفقودة.
- 3- الوسط الحسابي للقيم المجاورة وبيانات ثلاث مشاهدات قبل وبعد المشاهدة المفقودة.
- 4- الوسيط للقيم المجاورة وبيانات مشاهدتين قبل وبعد المشاهدة المفقودة.
- 5- الوسيط للقيم المجاورة وبيانات ثلاث مشاهدات قبل وبعد المشاهدة المفقودة.
- 6- طريقة الاستكمال الخطي.
- 7- التعويض بالانحدار.

كما ينبغي الاشارة الى استخدام تحليل الاتجاه العام (Trend Analysis) في توليد التنبؤات، علما ان هناك مجموعة

جدول 2. تقدير المشاهدة المفقودة للعام 2003 بمختلف الطرائق

السنوات	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي	الزراعي
2003	---	306.90	138.55	163.48	128.05	143.85	167.60

المصدر: من حساب الباحث بالاعتماد على البرنامج الاحصائي SPSS.

جدول 3. معايير الدقة التنبؤية والوسط الحسابي والانحراف المعياري لطرائق التنبؤ المختلفة

التعويض بالانحدار	طريقة الاستكمال الخطي	طريقة الوسيط للقيم المتجاورة span=3	طريقة الوسيط للقيم المتجاورة Span=2	طريقة المتوسط الحسابي للقيم المتجاورة span=3	طريقة المتوسط الحسابي للقيم المتجاورة	طريقة المتوسط الحسابي للسلسلة	معايير الدقة التنبؤية
Linear trend at point	Linear Interpolation	Median of nearby points Span=3	Median of nearby points Span=2	Mean of nearby points Span=3	Mean of nearby points Span=2	Series Mean	
7	6	5	4	3	2	1	
139.1	140.9	140.5	140.3	140.8	140.5	142.0	MAPE
146.8	149.1	148.4	148.0	149.0	148.3	152.9	MAD
31938.0	32149.9	32045.4	31995.4	32129.2	32026.9	33470.1	MSD
299.5	203.3	301.5	300.9	302.1	301.3	306.9	Mean
286.9	285.2	285.6	285.9	285.3	285.7	284.1	S.Dv

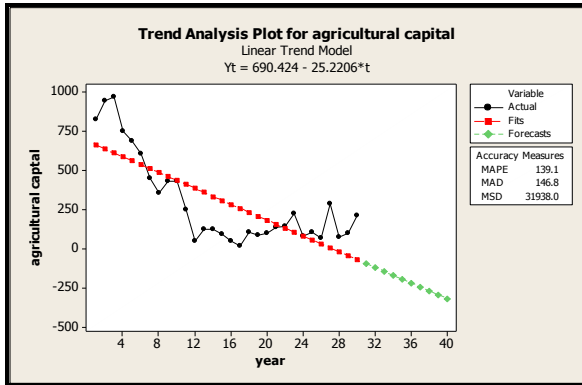
جدول 4. ترتيب الطرائق حسب معايير الدقة التنبؤية

ترتيبها	الطريقة	قيمة المعيار	المعيار	ترتيبها	الطريقة	قيمة المعيار	المعيار	ترتيبها	الطريقة	قيمة المعيار	المعيار
1	7	31938.0	MSD	1	7	146.8	MAD	1	7	139.1	MAPE
2	4	31995.4		2	4	148.0		2	4	140.3	
3	2	32026.9		3	2	148.3		3	2	140.5	
4	5	32045.4		4	5	148.4		4	5	140.5	
5	3	32129.2		5	3	149.0		5	3	140.8	
6	6	32149.9		6	6	149.1		6	6	140.9	
7	1	33470.1		7	1	152.9		7	1	142.0	

جدول 5. ترتيب الطرائق بحسب مقياس الوسط الحسابي (الى اليمين) والانحراف المعياري (الى اليسار)

الانحراف المعياري (Standard Deviation)		الوسط الحسابي (Mean)	
ترتيبها	الطريقة	ترتيبها	الطريقة
1	Series Mean	7	Series Mean
5	Mean of nearby points (span=2)	3	Mean of nearby points (span=2)
3	Mean of nearby points (span=3)	5	Mean of nearby points (span=3)
6	Median of nearby points (span=2)	2	Median of nearby points (span=2)
4	Median of nearby points (span=3)	4	Median of nearby points (span=3)
2	Linear Interpolation	6	Linear Interpolation
7	Linear trend at point (Regression Imputation)	1	Linear trend at point (Regression Imputation)

شكل 1. تحليل الاتجاه العام والقيم المتوقعة بالطريقة السابعة



جدول 6. نتائج التحليل للبيانات بفقدان 3 مشاهدات

الطريقة	مقاييس الدقة التنبؤية		
	MSD	MAD	MAPE
1	29435.1	138.5	125.1
2	29116.3	142.4	138.5
3	-	-	-
4	28757.9	141.2	136.9
5	-	-	-
6	29807.1	144.6	138.9
7	25787.0	125.0	124.8

يشير الجدول 7 الى ترتيب الطرائق السبعة (عند فقدان ثلاث مشاهدات) وحسب معايير الدقة التنبؤية. وفقا لبيانات الجدول 7 فان طريقة التعويض بالانحدار قد سجلت المرتبة الاولى في تحقيقها لأقل قيمة لمعايير الدقة التنبؤية والذي يجعلها الطريقة المفضلة وفقا للفرضيات التي وضعها البحث وجاءت الطرائق الاخرى بالمراتب الاخرى، وسجلت طريقة الاستكمال الخطي المرتبة الاخيرة في المعايير الثلاثة.

وفقا للجدول 5 فان الطريقة السابعة قد حققت اقل قيمة للوسط الحسابي لبيانات السلسلة في حين انها كانت في الترتيب الاخير في تحقيقها لقيمة الانحراف المعياري والسبب يعزى الى ان قيمة المشاهدة المتنبأ بها كانت صغيرة بالمقارنة مع الطرق الاخرى مما زاد من قيم التشتت في البيانات وان كان الفرق بين قيم الانحرافات المعيارية قليل جدا. بيانيا قد تم توقيع البيانات بعد تقدير المشاهدة المفقودة وكانت الرسوم جميعها تشير الى ان الاتجاه العام للبيانات وبمختلف الطرق كان سالبا الامر الذي يتوقع ان تكون البيئات المتنبأ بها تتجه نحو التناقص و هذا طبيعي لان التنبؤ بطريقة الاتجاه العام سيكون متأثرا الى حد كبير بالنزعة الاتجاهية للبيانات وهنا نلاحظ ان اتجاه البيانات كان سالبا ، وهذا واضح من خلال الشكل البياني الاتي (تم اعتماد الشكل البياني للطريقة السابعة- تعويض الانحدار) ان الحكم بأفضلية طريقة دون غيرها بفقدان مشاهدة واحدة قد يكون متحيزا بدرجة ما الى طريقة دون اخرى لاعتبارات الخصائص التي تتميز بها هذه الطرائق، الامر الذي حدا بالبحث ان يقوم بمحاولة حذف بعض المشاهدات وكان عددها ثلاث مشاهدات واجراء نفس التحليل وكانت النتائج ممثلة في الجدول 6. تجدر الاشارة ان فقدان 3 مشاهدات لم يتم فيها استخدام الطريقة الثالثة والخامسة لعد ملائمتها بسبب وقوع المشاهدة المفقودة الاولى في الترتيب الثالث للبيانات اي عام 1982 وهذا يعني وجود مشاهدتين اثنتين فوق مشاهدة المفقودة، وعليه سلاحظ عدم وجود قيم لمقاييس الدقة التنبؤية للطريقتين الثالثة والخامسة وهما لم يؤثر في نتائج التحليل.

جدول 7. ترتيب الطرائق (فقدان 3 مشاهدات) حسب معايير الدقة التنبؤية

المعيار	قيمة المعيار	الطريقة	ترتيبها	المعيار	قيمة المعيار	الطريقة	ترتيبها
MAPE	124.8	7	1	MAD	125.0	7	1
	125.1	1	2		138.5	1	2
	136.9	4	3		141.2	4	3
	138.5	2	4		142.4	2	4
	138.9	6	5		144.6	6	5
MSD	25787.0	7	1	MAPE	29435.1	7	1
	29435.1	4	2		29116.3	4	2
	28757.9	2	3		28757.9	2	3
	29116.3	1	4		29116.3	1	4
	29807.1	6	5		29807.1	6	5

النتائج والمناقشة

من خلال نتائج التحليل الاحصائي بالطرائق المختلفة المشار اليها يمكن القول بتفوق طريقة التعويض بالانحدار في تحقيق اقل القيم لمعايير الدقة التنبؤية والتي تعد مقياسا مهما في الحكم على افضلية طريقة دون غيرها مع الاخذ بنظر الاعتبار ضرورة ان تكون الطرائق جميعها ملائمة للتحليل نفسه والا فلا يمكن الحكم بالأفضلية ان اختلفت ظروف التحليل، ولهذا نجد ان تفوق الطريقة السابعة في تحقيق اقل قيم لمعايير الدقة التنبؤية كان مشروطا بملائمة البيانات لهذه الطريقة، وكذلك الحال عندما يتعلق الامر بطريقة الاستكمال الخطي اذ افترض البحث وجود استكمال خطي وليس متعدد الحدود، وتشابهت النتائج في حالة فقدان اكثر من مشاهدة اذ سجلت الطريقة السابعة افضل النتائج للمعايير الثلاثة. ينبغي عند الحصول على النتائج توظيفها للتأكد من صحة النتائج المتحصل عليها اي استخدام بيانات المتغير قيد البحث (اجمالي تكوين راس المال الزراعي) وعلاقتها بقيمة الانتاج الزراعي في نموذج انحدار بسيط والحصول على معاملات الانحدار ومقارنتها مع بعضها البعض ، ووفقا لمفهوم النظرية الاقتصادية فان العلاقة بين المتغيرين تصاغ حسب العلاقة الآتية:-

$$Y_i = a \pm bX_i$$

اذ ان:-

Y_i = اجمالي تكوين راس المال الزراعي للمدة (1980-2009)
بالاسعار الثابتة (1988=100)
 X_i = قيمة الانتاج الزراعي للمدة (1980-2009)
بالاسعار الثابتة (1988=100)

لقد تم استخدام بيانات اجمالي تكوين راس المال الزراعي بعد تقدير المشاهدة المفقودة بالطرائق السبعة المشار اليها في اعلاه واعتبارها متغيرا تابعا وعلاقته بالمتغير المستقل (قيمة الانتاج الزراعي) وتمثيل العلاقة بأنموذج انحدار بسيط ومن ثم مقارنة معاملات الانحدار المتحصل عليها فضلا عن الاختبارات الاحصائية والمتمثلة باختبار F واختبار R واختبار R^2 ويتم الحكم على طرائق تقدير المشاهدة المفقودة من خلال تحقيقها لاعلى قيم للاختبارات اعلاه، وجميع النتائج ممثلة في (الجدولين 8 و 9). تشير نتائج الجدولين 8 و 9 الى تفوق الطريقة السابعة (التعويض بالانحدار)

(Linear trend at point) في جميع الاختبارات فضلا عن تفوق قيم معاملات الانحدار المقدره في هذه الطريقة، حيث تشير قيمة المقدار الثابت والبالغة (973.242) وقيمة معامل الانحدار (0.174) الى تفوقها على جميع المعاملات في معادلات الانحدار المقدره بالطرائق الاخرى، الامر الذي يؤكد النتائج المتحصل عليها فيما يتعلق بمعايير الدقة التنبؤية والتي اشارت الى تفوق الطريقة السابعة، ومن ثم جاءت النتائج هنا والتي تؤكد تفوق الطريقة السابعة مرة اخرى. فضلا عن قيم معامل الارتباط ، والتحديد المتعدد واحصائية F ، والنتائج وان كانت متقاربة الا انها تبقى مهمة ومعنوية لاسيما عندما يتعلق الامر بعلاقات انحدارية يتم من خلالها حساب مرونة وغيرها. ان تفوق قيم المعايير الاحصائية في الطريقة السابعة تؤكد ان قيمة المشاهدة المفقودة المقدره بهذه الطريقة كانت اكثر واقعية وانها كانت متلائمة مع واقع البيانات وجاءت متوافقة مع التسلسل الزمني للبيانات وبالتالي حققت هذه النتائج، فضلا عن ان التنبؤ سيكون اكثر واقعية استنادا لهذه النتائج حتى وان كان واقع البيانات يشير الى اتجاه متناقص للبيانات وهذا لا يؤثر في نتائج التحليل بقدر ما يؤثر حالة البيانات قيد البحث فهي سواء كانت متجهة اتجاها متزايدا ام متناقصا فان التنبؤ بها لا بد وان يشوبه شيء من التحيز عليه فان اختيار الطريقة المناسبة لتقدير المشاهدات المفقودة سيكون له اثر حسن في القيم المتنبأ بها وبالتالي تعطي انطبعا جيدا لما سيحدث في المستقبل ويكون القرار المبني على النتائج المتحصل عليها قرارا مناسباً ويكون الخطأ التنبؤي صغيرا نسبيا. كما ان معاملات الانحدار المقدره بالطريقة المناسبة يمكن التعويل عليها افضل من غيرها لاسيما عندما يتعلق الامر بحساب قيم مادية كما هو الحال هنا، اذ ان اجمالي تكوين راس المال الزراعي سيمكن حسابه بدقة اكثر من خلال المعاملات المقدره ولا سيما ان الامر متعلق بمبالغ تقدر بملايين الدولارات، ومثل هذه المبالغ ينبغي تقديرها بشكل دقيق بحيث يكون القرار المتعلق بتوظيفها قرارا دقيقا الى حد ما.

جدول 8. التقدير الاحصائي لمعاملات دوال الانحدار حسب طرائق تقدير المشاهدات المفقودة

الطريقة	معادلة الانحدار	R	R ²	F	الطريقة	معادلة الانحدار	R	R ²	F
1	$Y = 937.464 - 0.163X$	0.668	0.447	22.591	5	$Y = 963.768 - 0.171X$	0.698	0.487	26.611
2	$Y = 964.624 - 0.171X$	0.699	0.488	26.736	6	$Y = 959.937 - 0.170X$	0.694	0.482	26.045
3	$Y = 960.602 - 0.170X$	0.695	0.483	26.144	7	$Y = 973.242 - 0.174X$	0.707	0.50	27.959
4	$Y = 966.317 - 0.172X$	0.701	0.491	26.981					

جدول 9. ترتيب الطرائق حسب الاختبارات الاحصائية

احصائية F		معامل التحديد المتعدد R ²		معامل الارتباط R		احصائية F		معامل التحديد المتعدد R ²		معامل الارتباط R	
ترتيبها	الطريقة	ترتيبها	الطريقة	ترتيبها	الطريقة	ترتيبها	الطريقة	ترتيبها	الطريقة	ترتيبها	الطريقة
4	5	3	5	3	5	7	1	7	1	6	1
6	6	5	6	4	6	3	2	6	2	5	2
1	7	1	7	1	7	5	3	4	3	7	3
						2	4	2	4	2	4

في القطاع الزراعي الذي يشهد واقعه تذبذبا واضحا الامر الذي ينبغي مراقبة بياناته بشكل دقيق.

المصادر

1. Abraham, B., and A. Thavaneswaran. 1991. A nonlinear time series model and estimation of missing observations. Ann. Ins. Statist .Math, 43(3):493-504.
2. AlNuamy, A. M. 2009. Mechanism of missing data and estimating them by principal component regression. The second scientific conference for mathematics and statistics .Coll. of computers and mathematics sciences. Univ. of Mosul, (in Arabic) p 312-322.
3. Chien-Pai H., and Y. Li. 2011. Regression analysis with block missing values and variables selection. Pak. J. Stat. oper. Ves. 7(2): 391-400.
4. David, S. C. 2006. Methods for The Estimation of Missing Values in Time Series. Master Thesis, Coll. of Communication, Health and science Edith Cowan University. Perth Western Australia, pp. 37.
5. Deluyker, H., A. S. Azari, R. H. Shumway, and L. Weaver. 1988. Replacement of missing values in time series data by kalman smoothed estimators using STATE-SPACE modeling.

مما تقدم استنتج البحث الى تفوق طريقة تعويض الانحدار (Linear trend at point) على الطرائق الاخرى في جميع المعايير الاحصائية فضلا عن تفوق معاملات الانحدار المقدره بهذه الطريقة وبالتالي يمكن الجزم بتفوق نتائج القيم المقدره بهذه الطريقة والتي على اساسها يمكن الحكم على القرار الذي سوف يتم تبنيه من قبل صاحب القرار بانه قرار مناسب يحمل اخطاء تنبؤية صغيرة بالمقارنة مع نتائج الطرائق الاخرى. كما واستنتج البحث الى اهمية ان تكون الطريقة المتبعة في تقدير المشاهدات المفقودة ملائمة الى حد كبير للبيانات والا فان النتائج ستكون متحيزة الى حد كبير فضلا عن عدم الاعتماد على القرارات المنبثقة منها. مما تقدم يوصي البحث اولا بضرورة قراءة البيانات المستهدفة لغرض تحديد الطريقة الملائمة لها، وفي حالتنا هذه فان طريقة تعويض الانحدار هي اكثر الطرائق ملائمة. كما ويوصي البحث بأهمية اجراء بحوث حول الحالات الاخرى التي لم يتطرق اليها البحث لغرض وضع صورة متكاملة لاسيما وان واقع البيانات في العراق يفترق الدقة وبالتالي تأثر النتائج المتحصل عليها وتكون القرارات المتخذة غير دقيقة ولاسيما

- Tokyo. Mc Grow. Hill. Kogakosha, Ltd. 2nd. pp. 150.
15. Raghunahan, T. E., M. L. James, V. John, and S. Peter. 2001. A multivariate technique for multiply imputing missing values using sequence of regression models. *Statistics Canada*, 27(1): 85-95.
16. Ryan, K. F., and E. David. 1988. Testing for unit roots in economic time- series with missing observations. Published on line: www.google.com.
17. Saunders, J. A., M. H. Nancy, S. Edward, D. Peter, K. P. Enola, and P. Richard. 2006. Imputing missing data: A comparison of methods for social work researchers. *Social work research*. 30 (1):19-31.
18. Shukr, A. S. 2010. An Analysis of Investment in Main Agricultural Machinery and Equipment (Fixed Agricultural Capital) in Iraq for the Period (1980-2009). M.Sc. Thesis, Dept. of Agricultural Economics., Coll. of Agric., Univ. of Baghdad (in Arabic). pp. 62.
19. Shumway, R. H., and D. S. Stoffer. 2010. An approach to time series smoothing and forecasting using the EM Algorithm. Published on line: www.google.com.
20. Simple ad-hoc methods and their shortcoming. Published on line: www.missingdata.org.uk.
21. Wang, Q., and G. E. Dinse. 2011. Linear regression analysis of survival data with missing censoring indicators. *J. Lifetime Data analysis*. 17(2): 256-279.
22. Zhoo, L.P., L. Stuart, and L. Danika. 1996. Regression analysis with missing covariate data using estimating equations. *Biometrics*. 52: 1165-1182.
- Sth international Symposium on veterinary Epidemiology and Economics. p 486-489.
6. Dunsumuir ,W., and P. M. Robinson. 1981. Estimation of time series models in the presence of missing data. *Journal of the American Statistical Association*. 76(375): 560-569
7. Eida, A. A. 2011. Statistical methods for estimating missing data in survey (in Arabic). Published on line www.google.com.
8. Hussein, S. 2007. Demand Forecasting for Wheat in Iraq. M.Sc. Thesis, Dept. of Agricultural Economics, Coll. of Agric., Univ. of Baghdad, (in Arabic). pp. 105.
9. Howell, D. C. 2012 . Treatment of missing data- Published on line missingdatanorm.html.
10. Lacon, F., S. J. Leybourne, and A. Rebrtaylor. 2013. Testing for a break in trend when the order of integration is unknown. *J of econometrics*. 176(1): 30-45.
11. Liu, Y., and S. Brown. 2012. Comparison of five iterative imputation methods for multivariate classification. *J. Chemometrics and Intelligent Laboratory Systems*. 120: 106-115.
12. Michael, P., S. R. Lipsitz, J. G. Ibrahim, and S. Lipshultz. 2002. A weighted estimating equation for linear regression with missing covariate data. *J. Statistics in Medicine*, 21 (16): 2421-2436.
13. Nasir, H. A. 2011. Estimating of lost value of responding variable in the multi regression model. *J. Economic Sciences (in Arabic)*. 8(30): 231-246.
14. Pindyck, R. S., and D. L. Rubinfeld. 1982. *Econometric Models and Economic Forecasts*.