مقارنة بعض طرائق تعويض البيانات المفقودة لنماذج لامعلميه وشبه معلميه مع تطبيق

أ.د. مناف يوسف حمود   مياسة محمد كاطع

جامعة بغداد–قسم الاحصاء

munaf_yousif@yahoo.com

المستخلص

يهدف هذه البحث الى مقارنة طريقتين تقديريتين لانموذج الانحدار الخطي الجزئي شبه المعلمي في حالة وجود بيانات مفقودة في متغير الاستجابة بفرض ان المتغيرات التوضيحية (التفسيرية) تكون تامة المشاهدة ، وهاتان الطريقتان هما طرائق شبه معلميه ولامعلميه ممثلة بالمقدر اللبي Nadaraya-Watson مع مقدر أنموذج لامعلمي لبي متعدد. واقتصرت المقارنة لحالتين من حالات التعويض وهما حالتي التعويض الثابت Deterministic (Imputation) والعشوائي (Random Imputation).

وقد تم استعمال اسلوب المحاكاة لغرض مقارنة هذه المقدرات مع وضع نماذج افتراضية وحجوم عينات وتباينات ونسب فقدان مختلفة وقد اثبتت نتائج المحاكاة أن أفضل طريقة في التعويض عن البيانات المفقودة هي طريقة التعويض شبه المعلمي الثابت (Deterministic) في جميع الحالات المفترضة وهذا ما تم اثباته ايضا في نتائج التطبيق العملي والمتمثلة بدراسة العوامل المؤثرة على الناتج المحلي الاجمالي في العراق والتي اشارت نتائجة بوضوح الى افضلية طريقة التعويض شبه المعلمي الثابت.

كلمات مفتاحية:الانحدار شبه المعلمي، الانحدار الخطي الجزئي، بيانات المفقودة،طريقة التعويض العشوائية ،طريقة التعويض الثابتة.

A Comparison some of Imputation Missing Data Methods for Nonparametric and Semiparametric Models with Application

Munaf Yousif Hmood **Mayasa Mohammed**
University of Baghdad, Department of Statistics

Abstract :
This study aims to compare two estimation methods for the semiparametric partial linear model when the response variable has missing data (non response). These methods are depends on Nadaraya-Watson. The mechanism of imputation the missing data are (Deterministic imputation and Random imputation).

Simulation technique is used with different sample sizes, models, variances and missing ratios. The results proved that the deterministic semiparametric imputation method is the best in all cases that proposed. The same result is proved in real data that represent the Gross Domestic Product (GDP) in Iraq, that is the result for real case is match with the simulation results.

Keywords: partial linear regression model, missing data, deterministic imputation method, random imputation method.

1− المقدمة

ان اغلب طرائق التحليل الاحصائي تفترض توفر بيانات تامة المشاهدة للعينات المدروسة وقد بنيت على هذا الاساس، ولكن في كثير من الاحيان وخصوصاً في المجالات التطبيقية تتعرض البيانات الى فقدان او عدم مشاهدة وتختلف الاسباب المؤدية الى ذلك فمنه ما يكون مخطط له ويقع تحت تصميم العينه بسبب الكلفة او المخاطرة او بسبب عدم توفر الامكانيات للمعاينة ومنها ما يكون خارج ارادة المعنيين مثل عطل اجهزة التسجيل او بسبب الكوارث و الحروب .[5]

ونظرا لذلك فقد حظيت مشكلة البيانات المفقودة باهتمام كثير من الباحثين وقد بدأ الاهتمام بهذه المشكلة في تصاميم التجارب (Experimental Designs) منذ بداية القرن العشرين الا ان الاهتمام بهذه المشكلة في تحليل الانحدار لم يظهر الا في منتصف القرن العشرين ويكمن السبب في ذلك لما تمتاز به تصاميم التجارب من وجود خاصية التعامد مما يجعل احلال قيم تقديرية عوضاً عن البيانات المفقودة امراً ممكنا في حين تنعدم هذه الخاصية في تحليل الانحدار مما يضيف للمشكلة ابعاداً اخرى ومزيدا من التعقيد.[2]

وقد يقود الفقدان الى تحيز في البيانات و من ثم تأثيره على نوعية البيانات ومعرفة أداءها ودراستها وبصورة عامة يمكن تصنيف طرائق التعامل مع البيانات المفقودة الى نوعين:

- حالة الحذف (Case Deletion) اي حذف المشاهدات او البيانات المقابلة للقيم المفقودة (اي دون التعامل مع البيانات المفقودة او محاولة تعويضها) .

وتعرف حالة الحذف أيضاً بــــ (List wise deletion) (LD) أي التعامل مع البيانات دون معالجتها وتحليل الحالة كاملة وهي الطريقة الاكثر شيوعاً والتي بموجبها تحذف حالات البيانات المفقودة وتواصل تحليل فقط المتبقي منها، وعلى الرغم من أن حالة الحذف هذه تنتج أنحساراً في حجم العينة المدروسة للتحليل الا أنها لها فوائداً مهمة عند افتراض ان البيانات مفقودة تماما وبشكل عشوائي (Missing Completely at Random) (MCAR) فأن هذا سيؤدي الى تقديرات معلميه غير متحيزة، وتعد هذه الطريقة مناسبة في حال كانت البيانات المفقودة قليلة، اما عندما تكون البيانات المفقودة كبيرة أو عندما يكون توزيع هذه البيانات غيرعشوائي فيمكن لهذه الطريقة أن تسبب تحيزا كبيرا ونتائج خاطئة نتيجة فقدان الكثير من المصادر والمعلومات.

- تعويض البيانات المفقودة (Missing data imputation) .

الحالة الثانية تتمثل بطريقة تعويض (Imputation) البيانات المفقودة، اي عملية استبدال البيانات المفقودة لمجموعة بيانات بقيم أخرى، ولهذه الطريقة فوائد عدة منها أن معالجة هذه البيانات المفقودة لاتعتمد على طريقة محددة دائما وهذا ما سيسمح للباحثين باختيار طريقة احتساب اكثر ملائمة لتطبيقاتهم.

ومع هذا فقد أكد الباحث (Dempster) عام 1983 أن طرائق احتساب البيانات المفقودة هي عامة ومرنة للتعامل مع هذا النوع من المشاكل ولكنها لاتخلو من المخاطر اذ لابد من أخذ الحذر عند توظيف طرائق الاحتساب لانها ممكن ان تولد تحيزاً جوهرياً بين البيانات الحقيقية أو الفعلية وتلك البيانات المحتسبة.وتوجد عدة تقنيات لادارة البيانات ذات البيانات المفقودة ولكن لا توجد واحدة أفضل من الاخرى بشكل مطلق،فالمواقف المختلفة تتطلب حلولا مختلفة.

اما الباحث (Allison) فقد اشار عام (2001) ان "الحل الجيد (أو الوحيد) لمشكلة البيانات المفقودة هو عدم وجود أي نقص أو فقدان في البيانات" فضلا عن أن كفاءة طرائق معالجة البيانات المفقودة ستعتمد بشكل جوهري على اليات الفقدان.[12]

ونتيجة لذلك قام الباحثان (Little & Rubin) بتصنيف اليات الفقدان الى ثلاثة أنواع وكما تم الاشارة اليها في المصادر [14,12,11,2] :

1- الفقدان العشوائي التام (Missing Completely at Random)

هذا النوع من الفقدان يعود الى ان الفقدان يكون مستقلا عن البيانات المفقودة نفسها وكذلك مستقلا عن قيم المتغيرات الاخرى في العينة، عندها يمكن القول أن البيانات تفقد تماما و بشكل عشوائي (MCAR) .

2- الفقدان العشوائي (Missing at Random)

هذا النوع من الفقدان له علاقة بقيم المتغيرات الاخرى فقط في حين يكون مستقلا عن بياناته المفقودة ،ففي مثل هذه الحالة يكون الفقدان بشكل عشوائي (MAR).

3- الفقدان بشكل غير عشوائي(Not Missing at Random)

يتولد سبب هذا الفقدان نتيجة البيانات المفقودة ذاتها أي ان فقدان البيانات سيكون بشكل متعمد و غير عشوائي .

اما الانموذج شبه المعلمي والذي يعد أول ظهور لمصطلحه (Semiparametric) عام 1980 من قبل الباحثين (Santner,Cail,&Brown) في مجال الاحصاءات الحيوية (Biometric) و كذلك عام 1981 من قبل الباحث (Whitehead) في كتابه في مجال الرياضيات وأطلق عليه مصطلح (Partially parametric) اي الانموذج المعلمي جزئيا، كذلك في نفس العام تم استعمال هذا المصطلح من قبل الباحثين (Finnas & Hoen) في مجال التحليل الديموغرافي (Demography).

ثم توالت بعد ذلك البحوث والدراسات في هذا المجال وقد شهدت نظريات التقدير والاختبار للنماذج شبه المعلميه (semiparametric) تطوراً سريعاً منذ عام 1993.

اما دراسة نماذج الانحدار شبه المعلمي بوجود قيم مفقودة فقد تم دراسته لاول مرة من قبل الباحثين (Zhao ,Rotnitzky &Robins) عام 1994 اذ درسوا النماذج شبه المعلميه بوجود قيم مفقودة عشوائياً . [16]

بصورة عامة يكون لنماذج الانحدار نوعان من الفقدان في البيانات، يتمثل الاول بأن يكون الفقدان في المتغيرات التوضيحية (التفسيرية) بينما الثاني يتمثل بالفقدان في متغير الاستجابة Y وهو ما سيتم التطرق اليه في هذا البحث.

2 - هدف البحث

يهدف هذه البحث الى تقدير أنموذج الانحدار الخطي الجزئي شبه المعلمي بوجود بيانات مفقودة في متغير الاستجابة وذلك بأستعمال طرائق شبه معلميه متمثلة بمقدر Speckman مع

مقدر أنموذج لامعلمي لبي متعدد المتغيرات وبأفتراض ان المتغيرات التوضيحية (التفسيرية) تكون تامة المشاهدة.

ومن ثم مقارنة تلك المقدرات للانموذجين ولحالتين من التعويض عن البيانات المفقودة هما المقدرات التعويضية (الثابتة Deterministic Imputation والعشوائية Random Imputation).

3- أنموذج الانحدار شبه المعلمي:

يعد انموذج الانحدار الخطي الجزئي Partial Linear Regression Model أحد نماذج الانحدار شبه المعلميه [6] ، ويرمزله بـــPLM وهو من نماذج الانحدار التي تعتمد على متغيرات خطية Linear وأخرى غير خطية لامعلميه Nonparametric وغالبا ما تكون متغيرات هذا الانموذج مستمرة Continuous Explanatory Variables .

وتوثر هذه المتغيرات الخطية واللاخطية في متغير الاستجابة Y ، وهو تعميم لتقنيات الانحدار الخطي القياسي ويعد حالة خاصة من النماذج التجميعية Additive Models لذلك يكون أفضل من النماذج اللامعلميه لانه يتجنب مشكلة تعدد الابعاد The curse of dimensionalityالتي تحدث في النماذج اللامعلميه عند زيادة عدد المتغيرات ومن ناحية أخرى هو أكثر مرونة من النماذج المعلمية الخطية لانها تقلل من الافتراضات الخطية المفروضة على هذه النماذج . [7]

ويمكن تمثيل أنموذج الانحدار شبه المعلمي بالصيغة الاتية:

$$\underline{Y} = X\underline{\beta} + g(t) + \varepsilon \qquad \dots (1)$$

اذ يشير $\underline{Y}$: الى متجة المتغير المعتمد أو متغير الاستجابة من الدرجة $n\times 1$.

X: يمثل مصفوفة المتغيرات التوضيحية المعلميه من الدرجة $n\times p$.

$\underline{\beta}$: متجة المعالم المجهولة من درجة $p\times 1$

$X\underline{\beta}$: الجزء المعلمي للأنموذج المدروس.

$\underline{t}$: متغير مستمر ويمثل عادة المتغير اللامعلمي في البيانات و من الدرجة $n\times 1$.

g(t): دالة تمهيدية غير معرفة مجهولة من الدرجة $n\times 1$.

$\underline{\varepsilon}$: متجة الاخطاء العشوائية من درجة $n\times 1$ وتكون مستقلة بمتوسط صفر وتباين σ^2.

للانموذج المذكور انفاً عدة تسميات اذ يسمى بأنموذج الانحدار شبه المعلمي Semiparametric Regression Model [5][6]، اوالانموذج الخطي الجزئي

Partially Linear Model [10] ،وسبب تسميته خطي كونه يتضمن جزئين جزء معلمي خطي فضلا عن جزء لامعلمي لاخطي وتربط هذه الاجزاء علاقة تجميعية. [9]

وقد تم استعمال هذا الانموذج على نطاق واسع في مجالات مختلفة كالدراسات الاقتصادية ، المالية والبيئية (كدرجات الحرارة ، تلوث الهواء وسرعة الرياح) وغيرها من التطبيقات الاخرى.

ولكي يتم أبقاء التنبؤات سليمة للأنموذج المذكور انفا يتم العمل على تقدير كل من متجة المعلمات $\underline{\beta}$ فضلا عن الدالة التمهيدية g(t) علما أن هذا النوع من النماذج يتم تقديرها ليس بشكل منفرد أي لكل جزء على حدة وأنما يتم بطريقة تكرارية بحيث يتم تقدير الجزء اللامعلمي أولا بعد وضع أفتراضات أولية حول قيمة المعالم المجهولة $\underline{\beta}$ باستعمال احدى الطرائق المعلميه (كطريقة المربعات الصغرى الاعتيادية OLS) ومن ثم يتم تقدير متجه المعلمات المجهولة $\underline{\beta}$ بعد تقدير الجزء اللامعلمي منه وأن المقدر الناتج وفق هذه الطريقة يدعى بالمقدر شبه المعلمي.[10]

4- أنموذج الانحدار الخطي الجزئي (PLM) في حالة البيانات غير التامة [11][12]

لأنموذج الانحدار الخطي الجزئي في المعادلة (1) وباعتبار وجود بيانات غير تامة في قيم متغير الاستجابة Y ولعينة حجمها n مع كون كلاً من قيم المتغيرات T، X هي مشاهدات تامة عندها يمكن وضع متغير مؤشر (Index variable) للتعبير عن وجود ام عدم وجود فقدان للبيانات .

$$\delta_i = \begin{cases} 0 & \text{if} \quad Y \quad \text{is missing} \\ 1 & \text{o.w} \end{cases}$$

وتحت ألية فقدان (MAR) وعلى فرض أن

$$r = \sum_{i=1}^{n} \delta_i$$

وأن

$$m = n - r$$

اذ يشير r : الى مؤشر مجموع حالات الاستجابة للمتغير Y.

n : يمثل مشاهدات العينة .

m : يمثل مجموع حالات عدم الاستجابة أو الفقدان للمتغير Y .

وعليه يمكن وضع رموز لحالات الاستجابة وعدم الاستجابة ب Sr وSm على التوالي .

مع وضع قيم أولية معرفة للمتجة β كي يتم تقدير الجزء اللامعلمي وفق المعادلة الاتية

$$Y_i - X_i{}'\beta = g(T_i) + \varepsilon_i \qquad i = 1,2,\dots.r \qquad \dots(2)$$

للـــدالة $g(t)$ يكون $\hat{g}(t)$ Kernel عندها يكون مقدر اللب

$$\hat{g}(T_i) = \frac{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right)\left(Y_j - X_j\beta\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad \dots(3)$$

في المعادلة (3) وباستعمال المقدر $\hat{g}(t)$ Gaussian kernel تشير الى دالة $K(.)$ وبفرض أن عوضاً عن $g(t)$ في المعادلة (2) نحصل على:

$$Y_i - X_i{}'\beta \approx \frac{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right)\left(Y_j - X_j\beta\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad i \in s_r \qquad \dots(4)$$

وبأستعمال التحويلات الاتية نحصل على :

$$Z_i \approx U_i{}'\beta \quad , \quad i \in s_r \qquad \dots(5)$$

$$Z_i = Y_i - \frac{\sum_{j=1}^{n} \delta_j Y_j K\left(\frac{T_i - T_j}{h}\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad \dots(6)$$

$$U_i = X_i - \frac{\sum_{j=1}^{n} \delta_j X_j K\left(\frac{T_i - T_j}{h}\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \quad , i \in s_r \qquad \dots(7)$$

وحسب نظرية أنموذج الانحدار الخطي فأنه بالامكان تقدير قيمة المعالم $\underline{\beta}$ كما يأتي:

$$\hat{\beta}_n = \left(\sum_{i=1}^{n} \delta_i U_i U_i{}'\right)^{-1} \left(\sum_{i=1}^{n} \delta_i U_i Z_i\right) \qquad \dots(8)$$

وبتعويض هذه المعادلة في المعادلة (3) نحصل على المقدر شبه المعلمي:

$$\hat{g}_n(T_i) = \frac{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right)\left(Y_j - X_j\hat{\beta}_n\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad \dots(9)$$

5. مقدر اللب لأنموذج الانحدار اللامعلمي المتعدد في حالة البيانات غير التامة [15][12][11]

في هذا البحث تم أدخال مقدر لامعلمي لدالة الانحدار متعددة المتغيرات وأن الهدف من هذا المقدر يستند على أفتراض أن المتغيرات في الأنموذج تسلك سلوكاً لامعلمياً ولاخطياً مع متغير الاستجابة مما يتطلب تقدير تلك العلاقة بالاعتماد على مقدر لامعلمي وأن صيغة هذا النوع من المقدرات تكون بالاعتماد على أنموذج الانحدار الاتي :

$$Y = g(X,T) + \varepsilon \qquad \dots(10)$$

اذ تشير الدالة g(.,.) الى دالة انحدار لامعلميه بأعتبار ان المتغيرات X,T تشير الى متغيرات لامعلميه وأن المقدر لتلك الدالة يكون وفق الصيغة الاتية:

$$\hat{g}_{\mathrm{n}}(\mathrm{x},\mathrm{t}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Y_i} \mathrm{K}_1\left(\frac{\mathrm{x}-\mathrm{X_i}}{\mathrm{h_x}}\right) \mathrm{K}_2\left(\frac{\mathrm{t}-\mathrm{T_i}}{\mathrm{h_T}}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{K}_1\left(\frac{\mathrm{x}-\mathrm{X_i}}{\mathrm{h_x}}\right) \mathrm{K}_2\left(\frac{\mathrm{t}-\mathrm{T_i}}{\mathrm{h_T}}\right) + \mathrm{n}^{-2}} \qquad \dots(11)$$

علما ان هذا المقدر يكون بأفتراض وجود متغيرين وبالامكان تعميم هذا النوع من المقدرات بتعميم تلك الصيغة وفق عدة متغيرات توضيحية مستعملة، ويشير كلاً من $K_1(.)$ و$K_2(.)$ الى دوال لبية قد تكون متساوية او مختلفة أي قد تكون$K_1(.)$=$K_2(.)$ أو قد لا تكون متساوية $K_1(.) \neq K_2(.)$.

في هذا البحث تم استعمال ثلاث متغيرات توضيحية، لذلك ستكون الصيغة للمقدرفي حالة البيانات التامة كما يأتي:

$$\hat{g}_{\mathrm{n}}(\mathrm{x},\mathrm{t}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Y_i} \mathrm{K}_1\left(\frac{\mathrm{x}_1-\mathrm{X_{i1}}}{\mathrm{h_{x1}}}\right) \mathrm{K}_2\left(\frac{\mathrm{x}_2-\mathrm{X_{i2}}}{\mathrm{h_{x2}}}\right) K_3\left(\frac{\mathrm{t}-\mathrm{T_i}}{\mathrm{h_T}}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{K}_1\left(\frac{\mathrm{x}_1-\mathrm{X_{i1}}}{\mathrm{h_{x1}}}\right) \mathrm{K}_2\left(\frac{\mathrm{x}_2-\mathrm{X_{i2}}}{\mathrm{h_{x2}}}\right) \mathrm{K}_3\left(\frac{\mathrm{t}-\mathrm{T_i}}{\mathrm{h_T}}\right) + \mathrm{n}^{-2}} \qquad \dots(12)$$

اما في حالة البيانات غير التامة تكون صيغة المقدر كما يأتي:

$$\hat{g}_{\mathrm{n}}(\mathrm{x},\mathrm{t}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \delta_i \mathrm{Y_i} \mathrm{K}_1\left(\frac{\mathrm{x}_1-\mathrm{X_{i1}}}{\mathrm{h_{x1}}}\right) \mathrm{K}_2\left(\frac{\mathrm{x}_2-\mathrm{X_{i2}}}{\mathrm{h_{x2}}}\right) K_3\left(\frac{\mathrm{t}-\mathrm{T_i}}{\mathrm{h_T}}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \delta_\mathrm{i} \mathrm{K}_1\left(\frac{\mathrm{x}_1-\mathrm{X_{i1}}}{\mathrm{h_{x1}}}\right) \mathrm{K}_2\left(\frac{\mathrm{x}_2-\mathrm{X_{i2}}}{\mathrm{h_{x2}}}\right) \mathrm{K}_3\left(\frac{\mathrm{t}-\mathrm{T_i}}{\mathrm{h_T}}\right) + \mathrm{n}^{-2}} \qquad \dots(13)$$

6. التعويض عن البيانات المفقودة في النماذج شبه المعلميه [12,11]

للتعويض عن البيانات المفقودة يوجد اسلوبان متبعان في التعويض وهما الاسلوب المحدد (او الثابت Deterministic) والاسلوب العشوائي (Random) ويمكن تلخيص هذين الاسلوبين وفق الاتي:

نفرض أن Y_i^R, Y_i^D, $i \in s_r$ ، يشيران الى طرائق التعويض عن البيانات المفقودة التي تعتمد على الانحدار شبه المعلمي الثابت (Deterministic) و العشوائي (Random) على التوالي اذ أن:

$$Y_i^D = X_i{}'\hat{\beta}_n + \hat{g}_n(T_i) \qquad , i \in s_m$$

$$Y_i^R = X_i{}'\hat{\beta}_n + \hat{g}_n(T_i) + e_i^* \equiv Y_i^D + e_i^* \qquad \dots (14)$$

6-1 طريقة التعويض المحدد (الثابت) شبه المعلمي

Deterministic Semiparametric Imputation Method

تستند هذه الطريقة على تقدير دالة الانحدار شبه المعلمي أي تقدير كل من الدالة $X'\underline{\beta}$ والدالة اللامعلمية $g(T)$ وعليه يكون لدينا الاسلوب الاتي في التقدير :

$$Y_i^D = X_i'\hat{\beta}_n + \hat{g}_n(T_i) \qquad \dots (15)$$

وبعد يتم التعويض عن البيانات المفقودة وفق الصيغة (15)من المعادلة الحصول على قيمة Y_i^D الاتية:

$$Y_{D,i} = \delta_i Y_i + (1 - \delta_i) Y_i^D \qquad \dots (16)$$

6-2 طريقة التعويض العشوائي شبه المعلمي

Random Semiparametric Imputation Method

وتستند هذه الطريقة على تقدير دالة الانحدار شبه المعلمي اي تقدير كلاً من $X'\underline{\hat{\beta}}_n$ و $\hat{g}(T_i)$ ومن ثم تقدير الانموذج شبه المعلمي وفق الصيغة الاتية:

$$Y_i^R = X_i{}'\hat{\beta}_n + \hat{g}_n(T_i) + e_i^* \qquad \dots (17)$$

$$\cong Y_i^D + e_i^* \qquad \dots (18)$$

$$e_i^* = Y_i - X'_i\hat{\beta}_n - \hat{g}_n(T_i) \qquad \dots (19)$$

وبعد الحصول على قيمة Y_i^R في المعادلة (17) يتم التعويض عن البيانات المفقودة وفق الصيغة التالية:

$$Y_{R,i} = \delta_i Y_i + (1 - \delta_i) Y_i^R \quad , i = 1,2 \dots . n \qquad \dots (20)$$

7. التعويض عن البيانات المفقودة للأنموذج اللامعلمي المتعدد

7-1 طريقة التعويض المحدد (الثابت) اللامعلمي

Deterministic Nonparametric Imputation Method

وتستند هذه الطريقة على تقدير دالة الانحدار اللامعلمي $\hat{g}_n(X_{i1}, X_{i2}, T_i)$ وعليه يكون التقدير للدالة كالأتي :

$$Y_i^D = \hat{g}_n(X_{i1}, X_{i2}, T_i) \qquad \dots (21)$$

وبعد الحصول على قيمة Y_i^D من المعادلة (21) يتم التعويض عن البيانات المفقودة وفق الصيغة الاتية:

$$Y_{D,i} = \delta_i Y_i + (1 - \delta_i) Y_i^D \qquad \dots (22)$$

7-2 طريقة التعويض العشوائي اللامعلمي

Random nonparametric imputation method

وتستند هذه الطريقة على تقدير دالة الانحدار اللامعلمي $\hat{g}_n(X_{i1}, X_{i2}, T_i)$ وحسب الصيغة الاتية :

$$Y_i^R = \hat{g}_n(X_{i1}, X_{i2}, T_i) + e_i^* \qquad \dots (23)$$

$$\cong Y_i^D + e_i^* \qquad \dots (24)$$

اذ ان:

$$e_i^* = Y_i - \hat{g}_n(X_{i1}, X_{i2}, T_i) \qquad \dots (25)$$

وبعد الحصول على قيمة Y_i^R في المعادلة (23) يتم التعويض عن البيانات المفقودة وفق الصيغة:

$$Y_{R,i} = \delta_i Y_i + (1 - \delta_i) Y_i^R \quad , i = 1,2 \dots . n \qquad \dots (26)$$

8. الجانب التجريبي

تم تنفيذ تجارب المحاكاة باستعمال ثلاث حجوم للعينات (n=40,60,100) وبتكرار مقداره 500 لكل تجربة وكما يأتي:

1- تتوزع المتغيرات التوضيحية Xi توزيعاً طبيعياً بوسط (0) وتباين (1) اي ان $X \sim N(0,1)$

2- يتوزع المتغير التوضيحي ti يتوزع توزيعاً طبيعياً بوسط (0) وتباين (1) اي ان $t \sim N(0,1)$

3- الاخطاء العشوائية تتوزع توزيعاً طبيعياً بوسط صفر وتباين σ^2 اي ان $\varepsilon \sim N(0,\sigma^2)$
وقد تم أفتراض ثلاث قيم لتباين الخطأ هي (0.5 ،1، 1.5) .

4- المتغير المعتمد Yi يتم توليده من خلال النماذج المستعملة في تجارب المحاكاة وذلك باستعمال دوال الانحدار المستعملة بدلالة المتغيرات التوضيحية في الفقرة (1 و2) مضافا اليها الخطأ .

9. النماذج المستعملة في المحاكاة

ان التنوع الكبير في النماذج ناتج من تنوع الظواهر التي تمثلها وقد تم اختيار بعض النماذج التي تناسب الطرائق المستعملة في هذه الرسالة من بحوث منشورة فعلا :

$1- \; g(t) = 3.2t^2 - 1$ [11] [12] [15]

$2 - g(t) = 0.5\sin(2\pi t_i)$ [7]

$3 - g(t) = t_i - 3t_i^2 + 3t_i^3$ [13]

أما الدالة اللبية Kernel المستعملة في هذا البحث هي دالة كثافة طبيعية Gaussian Kernel.

$$K(.) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$

$$P(x,t) = P(\delta = 1|X = x) = 0.8 + 0.2(|x-1+|t-0.5|), if|x-1| + |t-0.5| \le 1 \;,\; and = 0.95\,, elsewhere.$$

ومن الجدير بالاشارة الى وجود كثير من الدراسات والبحوث حول ايجاد المقدر الافضل للمعلمة التمهيدية الا ان في هذا البحث سوف يتم التطرق الى أحدى طرائق تقدير المعلمة التمهيدية والتي تدعى بطريقة قاعدة الابهام Rule of thumb Method.

اما لحالة البيانات المفقودة وحسب الية الفقدان (MAR) فتم توليدها وفق الصيغة الاتية [15] [12] [11].

وبنسب فقدان 10% و20% و30% .

علما ان تفسير الرموز في الجداول كانت :

PLM	Partial linear regression model	أنموذج الانحدار الخطي الجزئي
N.W	Nadaraya-Watson Kernel Estimator	مقدر ناداريا–واتسن اللبي
Y_i^D	Deterministic imputation method	طريقة التعويض الثابت
Y_i^R	Random imputation method	طريقة التعويض العشوائي
Non	Non parametric Kernel Estimator	مقدر لامعلمي لبي

جدول (1)

معدل متوسط مربعات الخطا للانموذج الاول في حالة التعويض عن البيانات المفقودة

	Method	PL-N.W Y_i^R			PL- N.W Y_i^D			Non-N.W Y_i^R			Non-N.W Y_i^D		
نسبة الفقدان	σ^2 / n	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
10%	40	0.2724	0.1387	0.2274	0.0692	0.0769	0.0761	1.0895	0.5552	0.9096	0.1332	0.3860	0.2205
	60	0.1104	0.1834	0.1663	0.0094	0.0125	0.0118	0.4418	0.7338	0.6651	0.1619	0.0099	0.1083
	100	0.0872	0.0483	0.0786	0.0023	0.0063	0.0048	0.3489	0.1930	0.3143	0.0114	0.0075	0.0081
20%	40	0.1921	0.1502	0.2619	0.2313	0.2122	0.2519	0.7682	0.6008	1.0475	0.8338	0.2377	0.2831
	60	0.0777	0.1086	0.0616	0.0320	0.0298	0.0358	0.3108	0.4346	0.2467	0.0309	0.1201	0.0546
	100	0.0862	0.0343	0.0957	0.0097	0.0144	0.0202	0.3449	0.1373	0.3830	0.0115	0.0119	0.0313
30%	40	0.1512	0.1848	0.1152	0.2959	0.3115	0.2849	0.6047	0.7391	0.4608	1.2281	1.6294	0.4174
	60	0.1606	0.1091	0.1168	0.0423	0.0376	0.0387	0.6424	0.4365	0.3743	0.3752	0.2255	0.3244
	100	0.0709	0.0964	0.0776	0.0217	0.0369	0.0301	0.2837	0.3856	0.3103	0.0934	0.0604	0.0507

توضح النتائج في جدول (1) لكل من طرائق التعويض للطرائق شبه المعلميه (PLM) واللامعلميه ولجميع حجوم العينات والتباينات المستعملة للأنموذج الاول:

1- أظهرت النتائج أن أفضل طريقة في التعويض عن البيانات المفقودة عند جميع نسب الفقدان هي طريقة التعويض شبه المعلمي الثابت (Deterministic) لحجوم العينات والتباينات المستعملة، ماعدا عند نسبة فقدان 10% وتباين $\sigma^2 = 1$ وحجم عينة (n=60) كذلك عند نسبة فقدان 20% وتباين $\sigma^2 = 0.5$ وحجم عينة (n=60), وعند تباين $\sigma^2 = 1$ وحجم عينة 100 اظهرت النتائج ان أفضل طريقة هي طريقة التعويض اللامعلمي الثابت (Deterministic)، اما في حالة نسبة الفقدان 30% وحجم عينة (n=40) ولجميع التباينات اظهرت نتائج ان طريقة التعويض شبه المعلمي العشوائي (Random) هي الافضل.

2- أشارت النتائج أن أقل الطرائق أداءاً هي طريقة في التعويض عن البيانات المفقودة هي طريقة التعويض اللامعلمي العشوائي(Random)، ماعدا عند نسبة الفقدان 20% وحجم

عينة (n=40) وتباين $\sigma^2 = 0.5$ ،و حالة نسبة الفقدان 30% وحجم عينة (n=40) وتباين $(\sigma^2 = 0.5, 1)$ إذ أظهرت النتائج أن اضعف الطرائق أداءاً هي طريقة التعويض الثابت (Deterministic) للأنموذج اللامعلمي.

3- اشارت النتائج بصورة عامة ان قيم (MASE) تتذبذب عند استعمال طريقة التعويض العشوائي (Random) للأنموذجين شبه المعلمي (PLM) واللامعلمي عند زيادة نسب الفقدان، ماعدا عند حجم عينة (n=40 ,100) وتباين $\sigma^2 = 0.5$ في الأنموذج شبه المعلمي (PLM) اذ تقل بزيادة حجم العينة وكذلك عند حجم عينة (n=40) وتباين $\sigma^2 = 0.5$ في الأنموذج اللامعلمي.

4- كذلك اشارت النتائج أن قيم (MASE) تزداد بزيادة نسب الفقدان عند استعمال طريقة التعويض الثابتة (Deterministic) في الأنموذجين شبه المعلمي (PLM) واللامعلمي ولجميع حجوم العينات وتباينات الخطأ المستعملة، ماعدا عند حجم عينة (n=40) وتباين $\sigma^2 = 1$ في الأنموذج اللامعلمي نلاحظ تذبذب في قيم (MASE) .

5- تقل قيمة (MASE) عند زيادة حجوم العينات ولجميع النماذج وقيم التباينات ونسب الفقدان المستعملة ولكلا طرائق التعويض للأنموذج شبه المعلمي (PLM) واللامعلمي، ماعدا عند استعمال طريقة التعويض شبه المعلمي العشوائي لكلا الأنموذجين وعند نسبة فقدان 10% وتباين $\sigma^2 = 0.5$ نلاحظ تذبذب في قيم (MASE)، وكذلك عند نسبة فقدان 10% وتباين $\sigma^2 = 0.5$ للأنموذج اللامعلمي.

6- تتذبذب قيم (MASE) عند زيادة تباين الخطأ ولجميع الطرائق والنماذج وحجوم العينات المستعملة .

جدول (2)

يبين معدل متوسط مربعات الخطا للانموذج الثاني في حالة التعويض عن البيانات المفقودة

نسبة الفقدان	Method	PL-N.W Y_i^R			PL- N.W Y_i^D			Non-N.W Y_i^R			Non-N.W Y_i^D		
	σ^2 / n	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
10%	40	0.2172	0.1824	0.1386	0.0173	0.0154	0.0164	0.8688	0.7304	0.5544	0.0183	0.0177	0.0343
	60	0.2220	0.1784	0.1699	0.0105	0.0068	0.0121	0.8881	0.7135	0.6797	0.0103	0.0113	0.0104
	100	0.1136	0.0989	0.1263	0.0087	0.0079	0.0078	0.4544	0.3957	0.5050	0.0011	0.0046	0.0065
20%	40	0.1893	0.1636	0.1286	0.0332	0.0331	0.0348	0.7572	0.6543	0.5145	0.0382	0.2404	0.8617
	60	0.1550	0.1144	0.1432	0.0231	0.0239	0.0226	0.6202	0.4576	0.5727	0.0312	0.0205	0.8067
	100	0.1256	0.0680	0.0663	0.0064	0.0085	0.0090	0.5025	0.2721	0.2654	0.0379	0.0015	0.1236
30%	40	0.2197	0.1528	0.1554	0.0438	0.0393	0.0391	0.8789	0.6111	0.6216	0.0497	0.4055	0.0637
	60	0.1346	0.1298	0.1027	0.0406	0.0296	0.0306	0.5385	0.5193	0.4108	0.0185	0.1175	0.0409
	100	0.0635	0.0629	0.0337	0.0228	0.0117	0.0095	0.2542	0.2519	0.1348	0.0024	0.0422	0.0267

توضح النتائج في جدول (2) لكل من طرائق التعويض للطرائق شبه المعلميه (PLM) واللامعلميه ولجميع حجوم العينات والتباينات المستعملة للأنموذج الثاني:

1- أظهرت النتائج أن أفضل طريقة في التعويض عن البيانات المفقودة عند جميع نسب الفقدان هي طريقة التعويض شبه المعلمي الثابت (Deterministic) لحجوم العينات والتباينات المستعملة، ما عدا عند (نسبة فقدان 10% ، تباين $\sigma^2 = 0.5,1.5$ وحجم عينة 60 ، تباين$\sigma^2 = 0.5,1$ ، وحجم عينة 100، و عند نسبة فقدان 20% وتباين$1= \sigma^2$ وحجوم عينات n=60,100)، اذ اظهرت النتائج ان أفضل طريقة هي طريقة التعويض اللامعلمي الثابت (Deterministic) ،كذلك في حالة نسبة الفقدان 30% وحجم عينة (n=60,100) و تباين $1= \sigma^2$.

2- أشارت النتائج أن أن أقل الطرائق أداءاً في التعويض عن البيانات المفقودة هي طريقة التعويض اللامعلمي العشوائي (Random) لجميع نسب الفقدان والتباينات وحجوم العينات المستعملة ،ما عدا عند نسبة فقدان 20% ، تباين $1.5 =\sigma^2$ وحجم عينة (n=40 , 60)، اذ اشارت النتائج الى ان طريقة التعويض اللامعلمي الثابت هي الاقل اداءاً.

3- اشارت النتائج بصورة عامة ان قيم (MASE) تتذبذب عند استعمال طريقة التعويض العشوائي (Random) للأنموذجين شبه المعلمي (PLM) واللامعلمي عند زيادة نسب الفقدان.

4- كذلك اوضحت النتائج أن قيم (MASE) تزداد بزيادة نسب الفقدان عند استعمال طريقة التعويض الثابتة (Deterministic) في الأنموذج شبه المعلمي (PLM) ولجميع حجوم العينات وتباينات الخطأ المستعملة، ماعدا عند حجم عينة n=100 وتباين $\sigma^2 = 0.5$. في حين يلاحظ تناقص قيم (MASE) عند زيادة حجم العينة للأنموذج شبه المعلمي (PLM) ، اما في الأنموذج اللامعلمي فيلاحظ تذبذب قيم (MASE) ايضا عند استعمال طريقة التعويض الثابتة (Deterministic).

5- تتناقص قيم (MASE) بزيادة حجم العينة عند استعمال طريقة التعويض العشوائي (Random) للانموذجين شبه المعلمي واللامعلمي، ماعدا عند نسبة فقدان 10%,20% وتباين خطأ $\sigma^2 = 1.5$ فيلاحظ وجود تذبذب في قيمتها، كذلك تتناقص قيم (MASE) عند زيادة حجم العينة في حالة التعويض بطريقة التعويض الثابتة (Deterministic) لكلا الانموذجين شبه المعلمي واللامعلمي ولجميع قيم تباين الخطأ ونسب الفقدان المستعملة.

6- عند زيادة تباين الخطأ يلاحظ تناقص قيم (MASE) عند استعمال طريقة التعويض العشوائي (Random) لكلا الأنموذجين شبه المعلمي واللامعلمي، عدا عند حجم عينة n=100 ونسبة فقدان 10% كذلك عند حجم عينة n=60 ونسبة فقدان 20% اذ يلاحظ تذبذب في قيمتها،اما في حالة استعمال طريقة التعويض الثابتة في الأنموذجين فيلاحظ تذبذب في قيمة (MASE) .

جدول (3)

معدل متوسط مربعات الخطا للأنموذج الثالث في حالة التعويض عن البيانات المفقودة

	Method	PL-N.W Y_i^R			PL- N.W Y_i^D			Non-N.W Y_i^R			Non-N.W Y_i^D		
نسبة الفقدان	σ^2 / n	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
10%	40	0.2207	0.1870	0.2078	0.0120	0.0072	0.0164	0.8831	0.7481	0.8311	0.0397	0.0129	0.0573
	60	0.1188	0.1163	0.1168	0.0032	0.0039	0.0041	0.4752	0.4653	0.4673	0.0087	0.0076	0.0077
	100	0.0891	0.1226	0.0776	0.0018	0.0013	0.0019	0.3564	0.4906	0.3105	0.0022	0.0065	0.0013
20%	40	0.1560	0.1401	0.2475	0.0493	0.0206	0.0407	0.6241	0.5603	0.9902	0.1086	0.0394	0.0619
	60	0.1263	0.1049	0.1355	0.0046	0.0061	0.0063	0.5052	0.4198	0.5422	0.0058	0.0039	0.0032
	100	0.0597	0.0817	0.0691	0.0039	0.0041	0.0041	0.2388	0.3270	0.2764	0.0021	0.0130	0.0017
30%	40	0.1766	0.1016	0.1647	0.0532	0.0458	0.0396	0.7063	0.4064	0.6588	0.1018	0.1351	0.0243
	60	0.1111	0.0820	0.0751	0.0107	0.0102	0.0104	0.4444	0.3281	0.3003	0.0306	0.0297	0.0116
	100	0.0516	0.0742	0.0349	0.0047	0.0046	0.0059	0.2065	0.2966	0.1399	0.0030	0.0062	0.0108

توضح النتائج في جدول (3) لكل من طرائق التعويض للطرائق شبه المعلميه (PLM) واللامعلميه ولجميع حجوم العينات والتباينات المستعملة للأنموذج الثالث:

1- أظهرت النتائج أن أفضل طريقة في التعويض عن البيانات المفقودة هي طريقة التعويض شبه المعلمي الثابت (Deterministic) ولجميع النماذج وحجوم العينات التباينات ونسب الفقدان المستعملة،ماعدا الحالات التي تكون فيها طريقة التعويض اللامعلمي الثابت (Deterministic) هي الافضل وهذه الحالات هي (نسبة فقدان 10% ،حجم عينة n=100 و تباين $\sigma^2 = 1.5$) ، (نسبة فقدان 20% ،حجم عينة n=60 وتباينات $\sigma^2 = 1,1.5$) و(حجم عينة (n=100) وتباينات $\sigma^2 = 0.5,1.5$) ، (نسبة فقدان 30% ،حجم عينة (n=100) وتباين $\sigma^2 = 0.5$).

2- أشارت النتائج أن أن اضعف الطرائق أداءاً في التعويض عن البيانات المفقودة هي طريقة التعويض اللامعلمي العشوائي(Random) ولجميع النماذج وحجوم العينات والتباينات ونسب الفقدان المستعملة.

3- بصورة عامة اشارت النتائج ان قيم (MASE) تتذبذب عند استعمال طريقة التعويض العشوائي (Random) للأنموذجين شبه المعلمي (PLM) و اللامعلمي وذلك عند زيادة نسب الفقدان.

4- كذلك اشارت النتائج أن قيم (MASE) تزداد بزيادة نسب الفقدان عند استعمال طريقة التعويض الثابتة (Deterministic) في الأنموذج شبه المعلمي (PLM) ولجميع حجوم العينات وتباينات الخطأ المستعملة، ماعدا عند حجم عينة (n=40) وتباين $\sigma^2 = 1.5$، اذ يلاحظ تذبذب قيم (MASE) عند زيادة حجم العينة للانموذج شبه المعلمي (PLM)، اما للأنموذج اللامعلمي فيلاحظ وجود تذبذب في قيم (MASE) في عدة حالات مع ازدياد قيمتها في حالات اخرى عند استعمال طريقة التعويض الثابتة (Deterministic).

5- تقل قيم (MASE) عند زيادة حجم العينة وعند استعمال طريقة التعويض العشوائي (Random) للأنموذجين شبه المعلمي (PLM) واللامعلمي، ماعدا عند نسبة فقدان 10% وتباين خطأ $\sigma^2 = 1$ فيلاحظ تذبذب في قيمتها في الأنموذج اللامعلمي، كذلك تقل قيم (MASE) عند زيادة حجم العينة في حالة التعويض بطريقة الثابتة (Deterministic) في كلا الأنموذجين شبه المعلمي واللامعلمي ولجميع قيم تباين الخطأ ونسب الفقدان المستعملة عدا عند نسبة فقدان 20% و تباين خطأ $\sigma^2 = 1$ فيلاحظ وجود تذبذب في قيمتها في الأنموذج اللامعلمي .

6- بصورة عامة يلاحظ تذبذب قيم (MASE) عند زيادة تباين الخطأ ولجميع الطرائق والنماذج وحجوم العينات المستعملة.

10.الجانب التطبيقي

في المبحث السابق والمتمثل بالجانب التجريبي تمت مقارنة تلك الطرائق وفق اسلوب المحاكاة لكن في هذا المبحث تم تطبيق الطريقتين الافضل على بيانات حقيقية تمثل الناتج المحلي في العراق (GDP) و العوامل المؤثرة عليه (عرض النقد M1 والانفاق الحكومي GE) وتم اختيار هذه العوامل او المتغيرات وفق النظريات الكنزية وفريدمان [1][3][4] في الاقتصاد لما لهذه العوامل من تأثير مهم في الناتج المحلي الاجمالي.

في هذا المبحث تم اعتماد بيانات الخاصة في الجهاز المركزي للاحصاء قسم الحسابات القومية للعراق من عام 1971 ولغاية 2010 وقد تم اعتماد هذه البيانات بالاسعار الثابتة لسنة 1988م كسنة اساس مقاسة بالدينار العراقي وكانت بيانات الانفاق الحكومي GE وبيانات عرض النقد M1 للعراق من سنة 1971–2010م مقاسة بالدينار العراقي ايضاً.

10-1 توصيف الأنموذج و تحليل النتائج

تعد مرحلة التوصيف من اهم مراحل اعداد الأنموذج الاقتصادي القياسي، اذ يتم فيه تحديد العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة الداخلة في الأنموذج في ضوء معطيات النظرية الاقتصادية.

ولغرض بناء أنموذج قياسي يوضح تأثير دور (عرض النقد والانفاق الحكومي والفترات الزمنية) على الناتج المحلي الاجمالي في العراق تم توصيف المتغيرات التوضيحية والمتغير التابع كما يأتي:

$$GDP=\beta 0+\beta 1M1+\beta 2GE+g(t)+\epsilon$$

إن الأنموذج المذكور أنفا يحتاج إلى تقديرات أولية لذلك تم استخراج قيم هذه التقديرات باستعمال طريقة المربعات الصغرى الاعتيادية وكانت كما في الجدول الآتي :

جدول (4)

القيم التقديرية لمعالم الأنموذج بطريقة (OLS)

Parameter	β1	B2
OLS	-1.2049	2.0274

اذ ان المتغيرات التوضيحية هي:

1- عرض النقد M1

اذ يعد عرض النقد المتغير التوضيحي الاول "وهو متغير معلمي" كون العلاقة بين هذا المتغير ومتغير الناتج المحلي الاجمالي هي علاقة خطية. [3][4]

2- الانفاق الحكومي GE

ويعد المتغير التوضيحي الثاني وهو متغير معلمي ايضاً كون العلاقة بين متغيري الانفاق الحكومي والناتج المحلي الاجمالي هي علاقة خطية.[1] ، لذلك اصبح الانموذج المعلمي ملائم لهما.

3- متغير الزمن T

يمكن ادخال متغير الزمن كمتغير مستمر يتبع دالة لامعلمية تقيس العلاقة بين الناتج المحلي الاجمالي والزمن حيث علاقة الناتج المحلي الاجمالي مع الزمن علاقة غير خطية.[8]

4- المتغير المعتمد (الناتج المحلي الاجماليGDP)

حسب النظرية الاقتصادية يكون الناتج المحلي الاجمالي ذو علاقة طردية مع العرض النقدي و الانفاق الحكومي فضلا عن متغير الزمن بوصف تلك المتغيرات متغيرات توضيحية وينتج من هذه الاعتمادية انموذج الانحدار شبه المعلمي.

وقد تم تطبيق هذه البيانات بنسبة فقدان 20% من اجمالي البيانات وهذه البيانات التي تم فقدانها تشير بيانات بعض السنوات وهي :

(1971,1972,1980,1981,1991,1995,2003,2006)

وقد تم اختيار هذه السنوات حسب المشاكل السياسية والاقتصادية التي يعاني منها البلد في هذه الفترة وهي اكثر السنوات ممكن ان تكون غير دقيقة و اشارت سنة 1972م الى سنة تأميم النفط فضلاً عن بداية سنوات الحروب و الازمات الاقتصادية و السياسية التي مرت على العراق والتي أثرت سلبا على اقتصاده.

جدول (5)

يشير إلى قيم متوسط مربعات الخطأ لمقدرات الأنموذج شبه المعلمي

لبيانات الناتج المحلي الاجمالي في العراق في حالة البيانات غير التامة

PL- N.W Y_i^D	Y_i^R PL- N.W
0.208561	0.03900

من خلال دراسة النتائج للبيانات غير التامة للناتج المحلي الاجمالي نلاحظ ان افضل طريقة هي طريقة التعويض شبه المعلمي العشوائي (Random) لأنموذج (PLM) تليها طريقة التعويض شبه المعلمي الثابت (Deterministic) لنفس الأنموذج .

11- الاستنتاجـــــــــات:

11-1 الجانب التجريبي

بناءا على ما تم تحليله من نتائج المحاكاة وفقا لكل أنموذج من نماذج الانحدار اللامعلمي ولجميع الحالات الأخرى تم التوصل إلى الاستنتاجات الآتية :

1- أفضل طريقة في التعويض عن البيانات المفقودة في الانموذج الاول هي طريقة التعويض شبه المعلمي العشوائي تليها طريقة التعويض شبه المعلمي الثابت اما للأنموذجين الثاني والثالث فقد اظهرت النتائج افضلية طريقة التعويض شبه المعلمي الثابت تليها طريقة التعويض شبه المعلمي العشوائي عند حجم عينة 40 وتباين خطأ $\sigma^2 = 1$ ونسبة فقدان 20%.

2- كانت طريقة التعويض اللامعلمي العشوائي اضعف الطرائق اداءً وتمثيلاً للتعويض عن البيانات المفقودة.

3- بصورة عامة تذبذبت قيم (MASE) عند استعمال طريقة التعويض العشوائي للأنموذجين اللامعلمي و شبه المعلمي عند زيادة نسب الفقدان. في حين تزداد قيم (MASE) بزيادة نسب الفقدان عند استعمال طريقة التعويض الثابتة في الأنموذجين شبه المعلمي واللامعلمي ولجميع حجوم العينات وتباينات الخطأ المستعملة .

4- تقل قيمة (MASE) عند زيادة حجوم العينات ولجميع النماذج وقيم التباينات ونسب الفقدان المستعملة ولكلا طرائق التعويض للنماذج اللامعلمية وشبه المعلمية.

5- بصورة عامة تتذبذب قيم (MASE) عند زيادة تباين الخطأ ولجميع الطرائق وحجوم العينات المستعملة.

11-2 الجانب التطبيقي

1- في حالة وجود قيم مفقودة في الانموذج شبه المعلمي يتم استعمال طريقة التعويض شبه المعلمي العشوائي.

2- يلاحظ ان عرض النقد M1 لم يسهم بشكل فاعل في زيادة الناتج المحلي الاجمالي وهذا ناتج عن تشوه هذا الاقتصاد اذ اظهرت النتائج ان تأثيره سلبي في العراق وذلك لان السياسة النقدية في العراق لم تشهد تحسنا كبيراً في دورها المطلوب منها وهذا يعود الى جملة من التغيرات على ادائها ودورها في التأثير على المتغيرات الاقتصادية المهمة،اذ حصلت ظروفا

صعبة على المستوى السياسي و الاقتصادي أسهمت في اضعاف دورها في تحقيق الاستقرار الاقتصادي في العراق فالحروب العسكرية الثلاثة في وقت قياسي دمرت البنى التحتية للاقتصاد العراقي واخرت عملية التنمية الاقتصادية و البشرية لعدة عقود أضافة الى الحصار الاقتصادي في اوائل التسعينات فقد اوقع العراق في مشاكل اقتصادية منها على سبيل الذكر لا الحصر الفساد الاداري والمالي وضياع فرص الاستثمار وظهور المضاربات و الاسواق الموازية وتغير النظام السياسي في نيسان 2003م و الغاء قوانين وتشريعات تحكم على مهنية السياسة النقدية بالتبعية لقرارات الدولة المركزية، فضلا عن عمليات التهجير الطائفي عام 2006-2007م كل هذه العوامل ادت الى ضعف دور السياسة النقدية على الناتج المحلي الاجمالي في العراق.

12- التوصيات

يمكن إجمال التوصيات الآتية على ضوء الاستنتاجات التي توصلنا إليها من خلال هذا البحث:

1- يوصي باستعمال نماذج الانحدار شبه المعلمي عند دراسة المشاكل الاقتصادية والمالية في العراق سواء اكانت البيانات تامة اوغير تامة.

2- دعم البنك المركزي وإعطاؤه دوراً اكبر لممارسة السياسة النقدية على اكمل وجه وذلك من خلال استعمال ادوات السياسة النقدية المختلفة والتي تدعو الحاجة لاستعمالها في معالجة الكثير من القضايا الاقتصادية التي واجهت وتواجهه الدولة .

3- توجيه المزيد من الاستثمارات نحو القطاعات الرئيسية للاقتصاد الوطني والتي تتيح قدرة انتاجية اكبر وبالذات نحو زيادة الانفاق على القطاع الزراعي و الصناعي ،كي يتم الاستغناء عن الاستيرادات لمواجهة الطلب الاستهلاكي المحلي من جهة وزيادة الناتج بما يحاول ان يحقق التوازن بينه وبين الانفاق الحكومي من جهة اخرى.

المصادر

1- الأيدامي،حمدية شاكر مسلم،(2001) "العلاقة السببية بين الناتج المحلي الاجمالي و الانفاق الحكومي في العراق في المدة (1970-1995)" رسالة ماجستير في العلوم الاقتصـــادية،كلية الادارة والاقتصاد،جامعة بغداد.

2- البياع،مهدي محمد،(2000) "مقارنة بعض طرائق تقدير معالم نماذج الانحدار الخطـــي بوجود بيانات غير تامة" رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.

3- الرفيعي،افتخار محمد مناحي، (2007) "السيولة العامة وفاعلية السياسة النقدية في السيطرة عليها مع اشارة تطبيقية للعراق " اطروحة دكتوراة في العلوم الاقتصادية،كلية الادارة والاقتصاد، جامعة بغداد.

4- الشيخاني،عامرعبد الله مجيد،(2008) " تحليل وقياس العلاقة السببية بين عرض النقود والناتج المحلي الاجمالي في بلدان مختارة (السعودية ومصر) للمدة (1990-2006)" رسالة ماجستير في الاقتصاد ،كلية الادارة والاقتصاد،جامعة بغداد.

5- الطائي،خالد ضاري عباس،(1998) "التحليل الأحصائي للبيانات غير التامة في نماذج الانحدار المتعدد" أطروحة دكتوراه في الاحصاء،كلية الادارة والاقتصاد، جامعة بغداد.

6- عيسى، اسيل مسلم،(2011) "مقارنة بعض المقدرات شبة المعلميه لتقديردالة استهلاك الطاقة الكهربائية لمدينة بغداد" رسالة ماجستير في الإحصاء كلية الادارة والاقتصاد، جامعة بغداد.

7-Aydin D. (2011) "Partially Linear Models Based on Smoothing Spline Estimated by Different Selection Methods: A Simulation Study" Department of Statistics, Faculty of Arts and Sciences, Muğla University. www.interstat.statjournals.net/year2011/articles/1105006.

8-Aydin D.(2007) " Estimation of GDP in Turkey by nonparametric regression models" Proceedings of the 6th WSEAS International Conference on Instrumentation, Measurement, Circuits & Systems, Hang Zhou, China, pp.221-225

9-Bhattacharya P.K. and Zhao P.L. (1997) " Semiparametric Inference In Partial Linear Model" The Annals of Statistics, Vol. 25, No. 1, PP 244-262.

10- Chin S. and Keilegom I. (2013) "Estimation in semiparametric models with missing data" Annals of the institute of statistical mathematics, Vol.65, issue.4, pp 785-805

11- Kartiko S.H.,(2012) " Semiparametric Regression with Missing Values" Math Dept, UGM, Yogyakarta, Indonesia. http://mfile.narotama.ac.id/files/Umum/JURNAL%20UGM/SEMIPARAMETRIC.

12-Qin Y., Zhang S. and Zhu X. (2006) " Semiparametric optimization for missing data imputation" Springer.

13-Schimek M.G. (2000)“ Estimation and inference in Partially Linear Models with Smoothing Splines “Journal of Statistical Planning and Inference, PP 525_540.

14-Tsiatis A. (2006) "Semiparametric Theory and Missing Data” Springer, New York.

15-Wang Q., Linton O. and Härdle W. (2004) "Semiparametric Regression Analysis with Missing Response at Random” The institute for fiscal studies Department of Economics.

16-Wellner J., Klaassen J. and Ritov Y. (2005) " Semiparametric models : A Review.